Bent Jesper Christensen
Carsten Kowalczyk *Editors*

Globalization

Strategies and Effects

Springer

Globalization

Bent Jesper Christensen • Carsten Kowalczyk
Editors

Globalization

Strategies and Effects

Springer

Editors
Bent Jesper Christensen
Department of Economics and Business Economics
Aarhus University
Aarhus, Denmark

Carsten Kowalczyk
The Fletcher School
Tufts University
Medford
Massachusetts, USA

ISBN 978-3-662-49500-1 ISBN 978-3-662-49502-5 (eBook)
DOI 10.1007/978-3-662-49502-5

Library of Congress Control Number: 2017932094

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Germany
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

To my parents, Doris and Jens H. Christensen
Bent Jesper Christensen

To my daughter, Sara Kowalczyk
Carsten Kowalczyk

Preface

This volume presents work on globalization by researchers from economics, business, political science, law, and the humanities as the culmination of a decade-long focus on this topic at Aarhus University (AU), Denmark, starting with the inclusion of globalization as a research focus area in the 2006 development contract between AU and the Ministry of Science, Technology and Innovation. The contract provided funding for projects on globalization at AU, with Bent Jesper Christensen (BJC, AU, coeditor) in charge of two large projects, including several contributors to this volume. As an adjunct faculty member at AU, Carsten Kowalczyk, Tufts University and AU (CK, coeditor), participated in the first of these projects, on "Globalization and European Integration" (GEI, Grant No. AU-121722-1344, also including Torben M. Andersen, AU (TMA), a contributor to this volume, and others). The second project grant, "Globalization, Migration and Capital Flows" (Grant No. AU-121722-1345, the project also including Annette S. Hansen, Peter Nannestad, and Martin Paldam (MP), AU researchers and contributors to the volume, and others), supplemented external funding obtained by BJC from the Social Science Research Council for the project "Globalization, Migration and Foreign Aid" (Grant No. SSF-275-05-0319, also with MP among the participants). BJC spent 1 year of the project period at the Department of Economics, Harvard University, and in Cambridge met with CK to make plans for project activities. The GEI grant provided for part-time buyout from teaching at Tufts University for CK to help organize an international conference on globalization in Denmark. After the project period for the original grants, BJC obtained a new external grant from the Social Science Research Council, for a research center on "Cycles, Adjustment, and Policy" (CAP, Grant No. DFF-10-079705, also including contributors to this volume TMA, Dale T. Mortensen, Northwestern University and AU, and Allan Sørensen, AU). CAP started in 2010 and financed a conference on "Globalization: Strategies and Effects," organized by BJC and CK, which took place at Hotel Koldingfjord, Denmark, November 9–12, 2011. At the conference, papers contributed in response to an international open call, invited papers, and papers coming out of CAP and the AU research focus area were presented, and some of the papers were later submitted for consideration for publication in this volume. Also considered were some papers

coming out of CAP and the AU research focus area, by authors who were unable to attend the conference, including Hale Utar, Bielefeld University, a contributor to the volume. In addition, before the conference, in September, 2011, Anne O. Krueger, Johns Hopkins University and Stanford University, had been awarded an Honorary Doctorate at AU, and her paper was considered too. For editing and final publication of the volume, BJC obtained grants from the AU Research Foundation (Grant No. AUFF-EFLS-14-430), the Aarhus Institute of Advanced Studies (AIAS) at AU, and CREATES, funded by the Danish National Research Foundation (Grant No. DNRF-78). We hope that the resulting volume will be of interest to researchers and advanced students of globalization alike and that it will help spur further research on this exciting and important topic.

Many individuals and institutions have helped along the way to this publication—too many for all to be mentioned here. We are grateful to Aarhus University, the AU Research Foundation, the National Research Foundation, and the Social Science Research Council for the grants mentioned above, to Tufts University for allowing the buyout of teaching for CK, and to AIAS and the Department of Economics, Harvard University, for their hospitality and support. Among the people who have helped with administration, organization, the conference, editorial assistance, photography, and many other matters are Juan Carlos P. Alvarez, Peter A. Bache, Ritwik Banerjee, Jeffrey H. Bergstrand, Henning Bunzel, Menzie Chinn, Anna P. Damm, Carl Davidson, Steven J. Davis, Salvatore Dell'Erba, Jan Fidrmuc, Holger Goerg, Ingrid Henriksen, Samuel S. Kortum, Asli Leblebicioglu, Rune V. Lesner, Edith Liu, Kristine Midtgaard, Violeta Moreno-Lax, Jørgen B. Mortensen, Jørgen U. Nielsen, Alexandros X. M. Ntovas, Paola A. B. Quiroga, Valerie Ramey, Bo S. Rasmussen, Helena S. Raulus, Davide Sala, Thomas A. Sampson, Christian Schultz, Niclas M. Skovbjerg, Kirsten Stentoft, Michael Søndergaard, Kenneth L. Sørensen, Claus Vastrup, and Rune M. Vejlin. We thank them all.

Aarhus, Denmark — Bent Jesper Christensen
Medford, MA, USA — Carsten Kowalczyk

About the Conference

Hotel Koldingfjord, Denmark, venue of the Aarhus University conference Globalization: Strategies and Effects, organized by Bent Jesper Christensen and Carsten Kowalczyk. (Photo taken by Niclas Maabari Skovbjerg)

Participants at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark, organized by Bent Jesper Christensen and Carsten Kowalczyk. (Photo taken by Niclas Maabari Skovbjerg)

Contents

Introduction to Globalization: Strategies and Effects

Bent Jesper Christensen and Carsten Kowalczyk

Abstract Globalization implies the continuing expansion and intensification of economic, political, social, cultural and judicial relations across borders. It is furthered by reductions in transportation and communication costs, the rise of new information technologies, such as the internet, and liberalizations in the markets for goods, services, labor, capital, and technology. Although it also occurs within existing legal structures, globalization in many cases involves political decisions about deregulation, free trade, and the integration of markets. It changes the life styles and living conditions for people around the world, presenting new opportunities to some, but risks and threats to others. Individuals, firms, governments, and transnational organizations that are lifted out of the framework of the nation state, like the World Bank, United Nations, the European Union, and multinational firms all face challenges of how to respond to globalization. The present volume provides important information to private and public decision makers who are choosing strategies for production, investment, and public policy in the increasingly globalized society.

1 Introduction

The ongoing process of globalization that the world is experiencing today implies the continuing expansion and intensification of economic, political, social, cultural and judicial relations across borders. Globalization is furthered by reductions in transportation and communication costs, the rise of new information technologies, such as the internet, and liberalizations in the markets for goods, services, labor, capital, and technology. Although it also occurs within existing legal structures, globalization in many cases involves political decisions about deregulation, free trade, and the integration of markets. It changes the life styles and living conditions for people around the world, presenting new opportunities to some, but risks and

B.J. Christensen (✉)
Aarhus University, Aarhus C, Denmark
e-mail: bjchristensen@econ.au.dk

C. Kowalczyk
Tufts University, Medford, MA, USA

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_1

threats to others. Individuals, firms, governments, and transnational organizations that are lifted out of the framework of the nation state, like the World Bank, United Nations, the European Union, and multinational firms all face the challenge of how to respond to globalization.

The present volume brings together researchers from economics, business, political science, law, and the humanities to shed light on the consequences of globalization processes, and to provide information to private and public decision makers who are choosing strategies for production, investment, and public policy in the increasingly globalized world. The volume consists of seven interrelated parts, on search and matching, multinational firms, financial markets and aggregate fluctuations, international integration and migration, policy issues, and development, with a closing part on what we know and what we would like to know about globalization. Here, we briefly introduce each part, and the papers therein.

2 Search, Matching, and Globalization

International trade in goods has been increasing over time, supported by trade agreements such as GATT and WTO,[1] and trade in services is now part of the international market, too. Net gains from globalization should be accessible, through free trade, migration, and better matching of skills and needs. At the same time, there is competition between countries regarding where production should take place, and thus where jobs are created. Outsourcing and foreign direct investment (FDI) allow shifting production abroad. These globalization processes have important consequences for labor markets. Even if workers are not as mobile as products and capital, jobs are mobile, and workers are thereby affected by the international competition. This has implications for wages and work conditions, and is often seen as increasing uncertainty and posing threats to low skilled workers in developed economies.

The volume opens in the first part with a set of papers using search and matching models to explore the consequences of market frictions for globalization, trade, outsourcing, and labor markets. Frictions may come in the form of costly information acquisition, adjustment, and market entry, and may lead to delays in matching of trading partners, inefficiencies, and mismatching, as in Diamond (1982) and Mortensen (1982). It takes time and effort for a person to find a job, and it takes time and effort for a firm to fill a vacancy. Once a match is made, it need not be at the same wage as in other worker-firm matches. Similarly, if production requires that two firms work together, frictions may make it costly for a firm to find a partner

[1]The World Trade Organization (WTO) came into being in 1995, to reduce trade barriers and enforce trade rules, following the original 1947 General Agreement on Tariffs and Trade (GATT), regulating trade among 153 countries.

in production. An important question is how these features are impacted by the opening of the economy, e.g., through trade, or offshoring of production.

In "Product and Labor Market Entry Costs, Underemployment and International Trade," Spiros Bougheas and Raymond Riezman use a matching model in which firms (entrepreneurs) face entry costs, and workers face costs of skill acquisition, to study the implications of the resulting frictions for underemployment and international trade. There is a high-tech and a low-tech sector. Establishing a comparative advantage in the high-tech sector at the national level would according to the Ricardian theory of international trade require technological know-how, and according to the Heckscher-Ohlin model sufficient skill endowments. However, as frictions are impediments to the matching of factors of production, cross-country differences in market rigidities may generate comparative advantages, too, as in research following Krugman (1995). In the model of Bougheas and Riezman, firms must pay the necessary investment cost to enter the high-tech sector, or remain low-tech. Workers must acquire the skills required to work in the high-tech sector in order to be considered there, but even skilled workers may fail to find a match with a high-tech firm, thus generating underemployment. Workers and entrepreneurs have heterogeneous abilities and take these into account when deciding whether to invest in skill acquisition, e.g., education. Bougheas and Riezman demonstrate how countries' trade patterns may depend on their labor market institutions as these may affect the search and matching costs and hence relative autarky prices. They show that strong inefficiencies in the labor market could lead to trade liberalization being welfare reducing. They argue, however, that rather than imposing trade restrictions, the appropriate policy response may be to improve product and labor market institutions, to reduce the frictions causing the mismatching and underemployment in the first place.

In a related paper, "Offshoring, Mismatch, and Labor Market Outcomes," David M. Arseneau and Brendan Epstein similarly study an economy where high-tech firms require high-skilled workers, whereas low-tech firms make do with either high- or low-skilled workers. Under conditions laid out in the paper, search frictions imply that some high-skilled workers may take low-tech jobs, in order to escape faster from unemployment. This is again a problem of mismatch. Other conditions lead to segmentation, i.e., high-skilled workers only accept job offers from high-tech firms. In this framework, the authors investigate the implications of offshoring, i.e., firms hiring foreign workers to do jobs that were previously done domestically. An important case is North-South trade, with Northern high-tech firms going multinational. The concern is that offshoring may reduce domestic (here, Northern) welfare, by reducing wages and increasing unemployment. The authors use a version of their model calibrated to U.S. data and a representative emerging market economy to support the analysis. They show that, starting from segmentation in autarky, trade liberalization and the resulting high-tech offshoring may lead to the mismatching problem, with some high-tech workers now having to accept low-tech jobs. However, they also show that this mismatching actually reduces the welfare loss resulting from shifting some of the production abroad, compared to a situation with segmentation even after the opening of the economy. In this sense,

the mismatching may rather be thought of as productive reallocation. As the authors note, it would be of interest to further study the response of the wage distribution to offshoring by allowing for on-the-job search, a mechanism that finds empirical support in Christensen et al. (2005). From a policy perspective, the paper shows the importance for economic welfare of maintaining a flexible labor market as a country participates in globalization.

In his paper "Good Times and Bad Times, With Endogenous Trade Policy Responses," T. Huw Edwards uses a third search and matching model, again one of North-South trade, to endogenize globalization cycles. The model, which shares some features with that of Grossman and Helpman (2002), explains why prolonged periods of economic growth tend to be associated with increased trade flows and low barriers to trade, while, once a recession hits, there is a tendency for tariff barriers to go up again. Growth is the exogenous driving process. There are two production stages, where upstream firms sell semi-finished goods to downstream firms that complete the manufacture and sell it to consumers. Informational barriers generate frictions, leading to a search process, where upstream and downstream firms must meet and match in order to produce. The North is assumed to have higher skill endowments than the South, implying a comparative advantage in downstream production in the North, and lower labor costs in the South. In an initial phase, trade costs are high, and most goods are produced through matches within the North. When the economy is opened to trade in a phase of growth and globalization, some downstream Northern firms search for upstream Southern partners. Search is time-consuming and requires credit. The author shows that if growth is sufficiently high, so that there is more entry into outsourcing trade, optimal tariffs are low, as the costs of higher tariffs stemming from the loss of searching firms more than outweigh the domestic welfare benefits. On the other hand, when growth falters, e.g., following a credit shock, optimal tariffs go up, and a protectionist regime with lower trade flows may arise. The paper shows the need to pursue a strategy of continued multilateral commitment to trade liberalization, even following crises and recessions.

3 Multinational Firms

In addition to market frictions, there is tremendous heterogeneity among trading partners. Heterogeneity at many different levels, including people, firms, and markets, is important for globalization, and in particular for multinational enterprises (MNEs), which are studied in the second part of the volume.

In his paper "Dynamic Capabilities and the Multinational Enterprise," on the theoretical foundations and business perspectives of the MNE, David J. Teece argues that the role of managers and other sources of cross-firm heterogeneity are insufficiently captured by traditional economic theories of MNEs, with their roots in Coase (1937), or by financial and agency theory, e.g., Jensen and Meckling (1976). Since globalization is less than complete, MNEs must deal with markets and countries characterized by different needs, desires, institutions, regulations,

and policies, with labor that is not fully mobile, and with other frictions and heterogeneity. As a result, each MNE has its own important history, spanning a unique set of markets, technologies, national and regional infrastructures, and so on. While traditional economic theory has focused on the transaction level and efficiency, explaining the rise of FDI and the MNE alternative to contracting as market failures due to transaction costs, with more recent work following Antràs and Helpman (2004) focusing on property rights, and research following Melitz (2003) coupling contracting with a heterogeneous firms approach, Teece argues for the importance of his complementary dynamic capabilities approach. This approach emphasizes the "signature" organizational processes rooted in the distinct histories and cultures of MNEs, the non-routine actions of entrepreneurial managers and creativity of expert staff that generate growth and prosperity, and the design of good business models and strategies. In the resulting framework characterized by heterogeneity, firms should reward and promote awareness of the dynamic capabilities of managers and experts addressing uncertainties and changes in market conditions on an ongoing basis. A case study illustrating how this took place in a large multinational oil firm is offered. The paper helps highlighting the entrepreneurial role of MNEs and their managers, and the complexity of managerial control.

Empirically, heterogeneity at the firm level is documented, e.g., by Bunzel et al. (2017), based on the detailed official register data from Denmark. The other papers in this part of the volume draw on similar firm level registers from Sweden and Denmark to explore MNEs empirically. Thus, Susanna Thede in her paper "Determinants of the Firm's Foreign Internationalization Decision" presents an empirical test of the Antràs and Helpman (2004) global sourcing model of organizational choice in a world of heterogeneous firms and imperfect contracting. She shows that firms select FDI over outsourcing exactly when there is more heterogeneity in the industry, here in terms of variation in total factor productivity (TFP) levels, or better foreign contracting enforcement, here in the form of legislative support of headquarter provision. Thus, her findings support the complementarity of the heterogeneity and transactions explanations of MNEs. In addition, more productive offshoring firms integrate in the source country. The paper contributes with novel evidence along these dimensions by using Swedish firm level census data not subject to sample selection bias, in contrast to earlier studies relying on surveys or size-restricted firm data.

Similarly using the Swedish firm level data, Ayça Tekin-Koru studies the selection into FDI and the choice of foreign access strategy in "Multinationals and Corruption: Business as Usual?" By adding the focus on access strategy, this work complements that of Hakkala et al. (2008) using the Swedish MNE data. Two main entry modes are considered, namely, cross-border mergers and acquisitions (M&As) and greenfield investments. If there is a risk that contracts are not enforced, or that corruption makes doing business abroad difficult, then M&As may be preferred over greenfield FDI, as they allow accessing local knowledge and capabilities. Tekin-Koru shows that under specified conditions, the theoretical predictions about the effects of an increase in the level of corruption are that it discourages greenfield

investments, as perhaps expected, but also that it actually encourages cross-border M&As, at least up to a certain corruption level, i.e., a nonlinear effect. The empirical results, using, among others, a bivariate probit model, support this theory, and are also consistent with some predicted secondary effects, e.g., greenfield FDI is more strongly discouraged by corruption for firms with high endowments of mobile skills or assets, such as research and development (R&D), product differentiation, etc., that are not tied to a specific location, whereas investment in M&As is more strongly encouraged for firms with non-mobile skills.

Firm level register data, now from Denmark, are similarly used by Sanne Hiller, Philipp J. H. Schröder, and Allan Sørensen in their paper "Export Market Exit and Firm Survival: Theory and First Evidence," where the idea is to endogenize firm exit in the Melitz (2003) model. More productive firms are more likely to export, and survive longer. These empirical findings support the theoretical results presented. The theory further predicts that exiting firms will leave the export market and continue producing for the domestic market for a while before closing down completely. In the data, the annual fraction of exiting firms that stay domestically active relative to all firms leaving the export market ranges between one third and one half. As the authors note, these fractions might be higher in countries where the domestic market is relatively more important to firms. Further, as they also note, and which would be an interesting question for further theoretical and empirical inquiry, it is possible that a higher number of foreign destinations serviced by a firm might indicate greater opportunities for cross-subsidization and hence predict longer survival.

The Danish firm level data are studied further by Hale Utar in her paper "Characteristics of International Trade Intermediaries and Their Location in the Supply Chain," complementing the work on intermediaries based on U.S. data by Bernard et al. (2010). Trade in intermediate goods has been growing over the recent decades of globalization, with internationalization of production and increased outsourcing. Utar shows that wholesale trade firms differ systematically depending on whether they trade in intermediate or final consumption goods. Wholesalers exporting intermediate goods command higher unit prices in international trade than manufacturers, whereas those exporting final goods command lower. Those exporting intermediate goods are larger, more skill-intensive, and pay higher wages than those exporting final goods. Further, wholesale trade firms differ systematically from manufacturing firms. They employ fewer but more educated workers, and export to fewer countries, but with more products, and higher total sales. Their share in export is higher in industries characterized by greater heterogeneity, in terms of product fragmentation, and lower entry costs, as indicated by firm size and capital intensity. Thus, the paper adds to the documentation of the role of heterogeneity at all levels, including both within and between industries.

4 Financial Markets and Aggregate Fluctuations

Financial markets and aggregate fluctuations are studied in the third part of the volume. The financial markets are among those where the effects of globalization are most evident. Some areas where limitations to globalization remain are the "home bias" in stock markets, restrictions on pension funds to invest domestically, and rigidities due to differential tax regimes across countries, but by and large, the financial markets are among the most globalized. This has implications for the access to capital and opportunities to invest, but also for the control of financial matters, e.g., the possibility for a country to set both its own monetary policy and its exchange rate. The behavior of foreign currency exchange rates is extremely important for the globalization process, and among the key issues are whether, when, and how currency unions should be formed. Of course, the currency market is closely linked to the bond market, also with international participants, and the interaction of both markets with the macroeconomy and the business cycle is affected by globalization.

In the paper "Cyclical or Structural? Evidence on the Sources of U.S. Unemployment," Jinzhu Chen, Prakash Kannan, Prakash Loungani, and Bharat Trehan examine the importance of the structural component of unemployment. This is unemployment stemming from a lack of matching of workers and firms, e.g., due to sectoral shifts, so that both workers and firms may be searching in vain. In contrast, cyclical unemployment is due to an overall lack of demand. Increases in the cross-industry dispersion in stock returns is used as a forward-looking indicator of shocks that affect sectors differently, and are expected to be permanent, thus warranting labor reallocation. Using vector autoregressions, the authors show that the portion of unemployment fluctuations attributable to structural factors, based on the stock return dispersion measure, rises from about 25 % to 40 % when focusing on long-term rather than overall unemployment. The remainder may be associated with cyclical factors, such as shocks to aggregate demand, oil prices, interest rates, inflation, and monetary and fiscal policies. Since reallocation takes time, long-term unemployment seems a relevant criterion. Indeed, stock prices should be moved mainly by long-run rather than cyclical factors, as argued by Fisher Black (1987), and the use of the stock return dispersion measure is based on this idea. It would be interesting to compare with results using implied volatility backed out of financial option prices as an alternative uncertainty measure, as this is specifically forward-looking (based on market participants' expectations). In his study of output and employment fluctuations, Bloom (2009) has shown that implied volatility is highly correlated with the cross-sectional dispersion measure, and Christensen and Nielsen (2007) have used it as a forward-looking measure to explain long-run stock market fluctuations. Overall, viewed jointly with the papers tying frictional unemployment to international trade and sourcing, the empirical results of Chen, Kannan, Loungani and Trehan on the importance of the structural component of unemployment underscore the strong role of globalization for the domestic economy.

Cathrine L. Mann and Oren Klachkin, in "U.S. Treasury Auction Yields Before and During Quantitative Easing: Market Factors vs. Auction-Specific Factors" use a unique data set on U.S. Treasury auctions to investigate the changes following the onset of quantitative easing in March 2009. China, Japan, and the U.K. are large players in the market for U.S. Treasury securities. The auction data do not directly include information on bidder nationality, but instead on bidder type, where foreigners are proxied by indirect bidders, as distinct from primary dealers (the 20 market makers, large banks, and broker-dealers) and direct bidders (with a trading relationship with the Treasury, e.g., large money and mutual fund managers). Auction allocations are observed, not bids. Little evidence of differences between bidder types is found. Larger allocations are expected to be associated with higher bids, and thus lower observed yields. This pattern is confirmed for bonds during the boom-bust period prior to quantitative easing, but not for bills, and disappears afterward. The effects of market factors are as expected throughout, e.g., higher option-implied risk (VIX) induces a flight to Treasuries and so lower yields. Auction-specific factors known prior to the auction may be used to forecast the outcome, and the best forecast is the market yield on the matched instrument on the previous day. Viewing this as an efficient market result, the fact that received coefficients are slightly different from unity makes sense, since the martingale property should apply to the discounted price, not the yield, and under the risk-neutral or pricing measure, only, cf. Harrison and Kreps (1979). Thus, future research could investigate risk premia, and the possibility of strategic bidding. Liquidity could be examined using a measure such as the limited dependent variable transaction cost estimate of Lesmond et al. (1999). Based on the reported results, the expected negative effects on yields of higher allocations occur after quantitative easing rather than before, once the previous day's yield is controlled for, documenting that this policy measure, implemented in the secondary market for U.S. Treasuries, did affect the primary (auction) market, too.

In their paper "Exchange Rate Volatility, Euro Effect and the Two Margins of Trade: Evidence from Monthly Trade Data," Florian Johannsen and Inmaculada Martinez-Zarzoso analyze data on daily exchange rates and monthly bilateral trade. Using a sample selection correction based on the inverse Mill's ratio to account for zero trade flows in a gravity model, following Helpman et al. (2008), and allowing for fixed or random importer and exporter effects, they separate the impacts of explanatory variables on the extensive and intensive margins of trade. Calculating exchange rate volatility as the standard deviation of daily rates over a moving interval, the authors show that higher volatility reduces both the probability of trade, and its amount, for all categories considered, namely, capital goods, intermediates, and final consumption goods. Being a member of the European Union (EU) promotes trade via both margins, whereas being in the Euro zone does so only via the intensive margin, and only for intermediates and final goods. Indeed, Euro membership actually reduces the probability of trade in all categories, and also the amount of trade in capital goods. Policy makers are thus left with a dilemma: On the one hand, the formation of a currency union eliminates exchange rate volatility within the union, and thus promotes trade, according to the received results on

volatility. On the other hand, the results on Euro membership warn that a common currency may reduce investment and thus trade in capital goods.

In her paper "Sovereign Debt and Austerity in the Euro Area: A View from North America," Anne Krueger reviews some of the literature on optimal currency areas. Mundell's original criteria for the success of a potential currency area included labor and capital mobility within the area, a mechanism for risk sharing, e.g., through fiscal policy, and synchronization of business cycles. These are still relevant, and more criteria have been added over time, as discussed in the paper. An unavoidable trilemma is that a country cannot in the long run have a fixed exchange rate (e.g., be in a currency union), choose its own macroeconomic policy, and at the same time have an open capital account. Here, in addition to the papers referenced by Krueger, some of the early, influential contributions are those by Fleming (1962) and Mundell (1963). Besides discussion of these issues, the paper provides a historical overview of some past programs of the International Monetary Fund (IMF), and some background and reasons for the challenges for the Euro area associated with the economic problems in Greece. The historical discussion is illuminating and extensive. An additional important aspect (although not part of the exposition) is the role of the OEEC/OECD for the success of economic integration, liberalization, and growth in the post-war years.[2] The high quality of this organization's staff and its attention to structural issues were considered highly valuable in many of the more successful European countries. With regard to the Greek case, it appears that the Mundell criteria with the later additions and the trilemma should again be considered. The author attributes the crisis to fiscal excesses. It may also be that Greece, even without extravagant public finances, but abundant in services, low wages, high unemployment, rigid business practices, and so on, would not satisfy the criteria discussed for membership of a common currency area with, for example, Germany.

5 International Integration and Migration

International migration is a pressing issue within globalization, and it is closely connected to the way in which countries choose to integrate. Similarly to the gains from trade, there should be potential gains to increased mobility of labor. The chance of realizing such gains depends on the labor market absorption of immigrants, and an important issue is whether actual migration improves the matching of abilities of workers and skills needed by firms. Generous welfare states financed through high taxes offer both redistribution and insurance, and may seem attractive to many

[2]The Organisation for European Economic Cooperation (OEEC) was formed in 1948, to help administer the Marshall plan, and by 1961 transformed into the Organisation for Economic Cooperation and Development (OECD), with membership extended to non-European states, to stimulate economic progress and world trade.

migrants. This raises adverse selection issues regarding which migrants approach a country, and which to accommodate, as well as moral hazard issues regarding those accommodated. The question is whether and in which form welfare states can remain in a world with continuing migration and integration through the formation of unions. This is an area of complicated interplay between economic, demographic, political, and legal issues, as reflected by the papers in the fourth part of the volume.

Oded Stark in "Possible Policy Responses to a Dark Side of the Integration of Regions and Nations" shows that when two regions or countries merge, aggregate relative deprivation is higher (i.e., worse) in the merged region than the sum of the aggregate relative deprivations of the two regions before the merger. Here, aggregate relative deprivation refers to the average relative deprivation of individuals in a region, with relative deprivation of an individual defined as the sum of income differences between the individual and those in the region with higher income. This is intended to capture a notion of social distress. The received increase in social distress is a dark side to the integration of regions or countries. It would be interesting to extend the analysis to the risk of poverty from EU integration. Further extensions could endogenize incomes and thus allow for gains from integration, e.g., due to increased productivity, see Rivera-Batiz and Romer (1991). In terms of policy responses, Stark goes on to device publicly financed policies to retain either aggregate or each individual relative deprivation at the pre-merger level. These policies are constructed very explicitly, as algorithms, and numerical illustrations are offered.

Peter Nannestad in his paper "Where Have All the Flowers Gone?" uses data on migration within the EU-15 countries to test the model of Roy (1951) and Borjas (1987). Earlier tests have relied on data on migration between more dissimilar countries, e.g., from Mexico to the U.S., or to the EU from the outside, and turned up mixed evidence. The idea of focusing on within EU-15 migration is that barriers are low, as assumed in the theory tested, and data sufficiently detailed that variables may be measured in the manner called for in the theory. Specifically, the position of migrants in the skill distribution of the sending country is regressed on the ratio of the Gini coefficients for the income distributions in the receiving and sending countries, along with various controls. The results show that the higher the income dispersion in the receiving country relative to the sending, the more likely it is that immigrants are drawn from the high end of the skill distribution in their country of origin. This result on the skill-based locational sorting of immigrants is consistent with the theory tested. The policy implications present a challenge to highly distributive welfare states like the Nordic, as they would not stand to receive the highest-skilled migrants, according to these findings.

6 Globalization and Policy Issues

Besides labor migration, also product market integration and the ways in which countries shape and run unions and other international organizations involve a host of economic, political, and legal issues. The fifth part of

the volume includes papers on policy issues relating to the role of the EU, international conflicts, the environment, fiscal policy, economic integration, and the WTO.

Thorvaldur Gylfason and Per Magnus Wijkman in their paper "Which Conflicts Can the European Neighbourhood Policy Help Resolve?" address the important question of whether the prospect of economic integration, in particular with respect to trade, with a major economic power, here the EU, can reduce political and military conflicts among member and candidate countries. The consequences for economic welfare of integration between small and large countries has been considered previously, see Kowalczyk (2000), and the additional political and national security dimensions are obviously of great interest, too. Such thinking has led to prominent initiatives and institutions, including the Marshall Plan, GATT, the World Bank, and the IMF. The authors identify five factors that are critical for economic integration with the EU to lead to conflict resolution: Access to facilitators of compromise, a well-functioning market economy, strong common institutions to realize potential intra-regional free trade, the prospect of accession to the EU, and the absence of third parties to the conflicts. The authors argue that this indicates encouraging prospects for the Balkans, but neither so for Eastern partners, such as the Ukraine, where Russia is involved as a third party, nor for Mediterranean partners, such as Israel, for lack of an active facilitator of compromise in the conflict over Palestine. The EU has failed to assume the latter role, due to, among others, disagreement among its members, and the need for larger international coordination, in particular, with the U.S. Ultimately, the resulting policy recommendations include that the EU adopt a long-term perspective and provide greater support in its neighborhood in response to changes there.

Ellen Margrethe Basse in "The European Union's Role in the Development of Global Environmental Law" demonstrates that the EU through its actions is actively pushing the conceptualization of a new, more comprehensive, and thorough, approach to the management and regulation of the global environment. Such global environmental law may be seen as developing in parallel with existing national and international environmental law, cf. Yang and Percival (2009). Climate changes represent serious global threats, and the EU is among the leaders of international efforts to reduce greenhouse gas emission. The author explains how EU initiatives in the area of emissions from flights, ships, and road transport reflect ambitions to secure and strengthen global norms in cases where regimes and agreements in place are ineffective. Indeed, the European Court of Justice supports the EU's right to act independently of its member states in trying to export EU environmental norms. In effect, the EU is playing the role of a global actor in this respect, combining its internal regulatory power and external economic power to pursue its environmental agenda. Thus, while Gylfason and Wijkman identify some shortcomings with respect to the EU's taking on its global responsibilities, Basse is more positive regarding its engagement.

In their paper "Product Market Integration, Tax Distortions and Public Sector Size," Torben M. Andersen and Allan Sørensen consider a Ricardian trade model, with two countries trading some of a large number of goods. Following Dornbusch et al. (1977) and Eaton and Kortum (2002), trade frictions provide a measure of the degree of market integration, with globalization driven by both political and technological changes reducing frictions. In the model, there are iceberg costs to trade, i.e., frictions imply that more than one unit must be shipped in order to supply one unit to the export market, and reductions in iceberg costs are used to capture increasing product market integration. A larger public sector (higher tax rate) tends to increase wages and reduce wage competitiveness. The authors show that product market integration does not necessarily reduce public sector size, in part due to the increased tax base stemming from gains from trade. Indeed, a country with a strong preference for public spending may gain from integrating with a country with low such preference, and integration need not lead the country with the large public sector to downsize this. Clearly, these results require an asymmetric model, with cross-country differences in preferences and technologies, and hence the present asymmetric extension of previous analysis by Andersen and Sørensen (2012) of the symmetric case is essential. For the Nordic welfare states, one policy implication of the new results is that integration and globalization need not lead to their dismantlement. Thus, Andersen and Sørensen are less pessimistic regarding integration than Stark.

Tristan Kohl in "The WTO's Effect on Trade: What You Give is What You Get" revisits the question of the impact of WTO membership on trade, where Rose (2004) somewhat unexpectedly found no effect. Kohl contributes to this literature by incorporating data on political relations, in particular, a comprehensive set of economic integration agreements (EIAs), correcting for selection and omitted variable bias using these, and correcting for the over-dispersion and large number of zeroes in the trade data by using a zero-inflated negative binomial maximum likelihood estimation procedure. With these adjustments, he finds a strong positive impact of WTO membership on trade. The effect is stronger for developed countries than for developing, and what you give is what you get: Although WTO membership promotes trade among partners, even if not all of them are members, formal members gain the most. Finally, EIAs promote trade. As Kohl's results are more in line with what would be expected from theory than those in the earlier literature, it appears that his methodological innovations are promising.

7 Globalization and Development

Globalization is not only a process among developed, industrialized countries. Developing countries are very much affected by globalization, too. They trade, engage in North-South partnerships, send migrants, receive official development

aid (ODA) and FDI, etc. Does globalization help or hurt development, according to criteria such as growth, inequality, education, and others? These and related issues are treated in the sixth part of the volume.

In "The Cycle of Development in Africa: A Story About the Power of Economic Ideas," Martin Paldam analyzes the last 60 years of development in Sub-Saharan Africa. There are three main phases, divided by kinks in 1972 and 1994. Growth was negative between these 2 years, although fairly satisfactory before and after this period of growth tragedy. Paldam considers a number of hypotheses about possible explanations of this cyclical growth pattern, testing them based on a set of socio-economic variables. African poverty has been studied, e.g., by Acemoglu and Robinson (2010), but Paldam deliberately focuses on theories that he finds more operational, compared to the primacy of institutions school. The hypotheses considered are related to political dynamics, oil and debt crises, geography, history, wars, and governance, among others. The best fit is obtained through the shifts in dominating development strategy. The market systems preferred by the colonial powers and prevailing until around 1965 were replaced by African socialism, with import substitution industrialization, trade protection, and state owned enterprises (SOEs). The goal here was self-reliance and insulation from the world market. From around 1985, market systems returned, through the adoption of trade liberalization and privatization of SOEs. Accounting for implementation lags, the fit with the growth tragedy phase is close. Paldam notes that African GDP per capita nevertheless has doubled since 1950, but that without the growth tragedy, it would have tripled. The policy recommendation is one of market orientation and openness.

Annette Skovsted Hansen in "Developing Globalization: How Japanese ODA Created Channels for Globalization Processes, 1954-2012" documents an interesting case of development assistance leading to globalization. Over the past 60 years, Japanese ODA has financed training courses in management and technology in Japan, leading to the formation of personal networks kept together through formal newsletters, societies, conventions, etc., and forming permanent channels of communication and exchange. The networks are connected to the host country, but may also encourage South-South exchanges. The formation of personal networks may reflect preferential mechanisms, see Barabási and Albert (1999), with the host country selecting course participants among already well-connected elites from developing countries. Hansen is able to access information through newsletters and carry out a qualitative case study. There are examples of exchanges among network members within or between their home countries in the South, and with Japanese stakeholders, career shifts, and export increase. The paper contributes with a new aspect of how foreign development aid actually spurs globalization at the micro level.

In "Globalization and Female Labor Force Participation in Developing Countries: An Empirical (Re-)Assessment," Konstantin M. Wacker, Arusha Cooray and Isis Gaddis investigate the impact of FDI and trade on female labor force participation. Goldin (1995) has argued for a U-shaped relation between development and the labor force participation of married women, with high participation among

poor, agricultural workers, falling rates as (husband's) income increases, then rising rates as female education improves with further increases in income, the substitution effect beginning to outweigh the income effect. Previous empirical literature, often relying on case studies or cross-country regressions, indicates that globalization supports female labor force participation. In this paper, the authors consider a panel data set of 80 countries over 25 years, controlling for fixed effects and thus unobserved heterogeneity. The results indicate that both FDI and trade have a negative impact on female labor force participation, and that this is stronger for younger cohorts. Effects are more positive in countries with a higher share of industry in value added. The authors discuss the possibility that female well-being is not necessarily harmed by globalization, based on previous findings that trade reduces female illiteracy rates and improves access to education. Household income may have increased sufficiently for females to stay home. Nevertheless, in this case, the policy implication emerges that a reduction in female labor force participation following an opening to globalization should be met with continuing education programs, more flexible work schedules, or similar policies, to support employability.

8 Globalization: What We Know and What We Would Like to Know

The seventh and last part of the volume consists of two panel discussions. The first, on business perspectives on globalization, offers an extensive presentation by Jørgen Vig Knudstorp, CEO, LEGO Group, ranging from broad issues of trade, power, culture, digitalization, demography, and the environment, over consideration of specific effects of globalization on business, such as free trade, complexity of management, cycles, and inequality, to discussion of strategies adopted by businesses facing the challenge of globalization, including choice of operating model, location decisions, and marketing. This is followed by discussion spurred by questions from Keith Maskus and David Teece. The last panel offers some reflections and indications of the road ahead, with remarks by Avinash Dixit on the collapse of trade after the financial crisis, bad governance as a trade barrier, and the interaction between internet use and globalization, remarks by Ronald W. Jones on globalization and the "sclerotic economy," and remarks by Dale T. Mortensen on unemployment, reallocation, and frictions. The ensuing discussion ends the volume.

References

Acemoglu, D., & Robinson, J. A. (2010). Why is Africa poor? *Economic History of Developing Regions, 25*, 21–50.

Andersen, T. M., & Sørensen, A. (2012). Globalization, tax distortions, and public sector retrenchment. *Scandinavian Journal of Economics, 114*, 409–439.

Antràs, P., & Helpman, E. (2004). Global sourcing. *Journal of Political Economy, 112*, 552–580.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512.

Bernard, A. B., Jensen, B., Redding, S. J., & Schott, P. K. (2010). Wholesalers and retailers in U.S. trade. *American Economic Review Papers and Proceedings, 100*, 408–413.

Black, F. (1987). *Business cycles and equilibrium*. New York: Basil Blackwell.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica, 77*, 623–685.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review, 77*, 531–553.

Bunzel, H., Christensen, B. J., & Mortensen, D. T. (2017). *Firm heterogeneity, labor markets, and international trade—evidence from Danish matched employer-employee data*. Berlin: Springer.

Christensen, B. J., Lentz, R., Mortensen, D. T., Neumann, G. R., & Werwatz, A. (2005). On-the-job search and the wage distribution. *Journal of Labor Economics, 23*, 31–58.

Christensen, B. J., & Nielsen, M. Ø. (2007). The effect of long memory in volatility on stock market fluctuations. *Review of Economics and Statistics, 89*, 684–700.

Coase, R. (1937). The nature of the firm. *Economica, 4*, 386–405.

Diamond, P. A. (1982). Wage determination and efficiency in search equilibrium. *Review of Economic Studies, 49*, 217–227.

Dornbusch, R., Fischer, S., & Samuelson, P. (1977). Comparative advantage, trade and payments in a Ricardian model with a continuum of goods. *American Economic Review, 67*, 823–839.

Eaton, J., & Kortum, S. (2002). Technology, geography and trade. *Econometrica, 70*, 1741–1779.

Fleming, J. M. (1962). Domestic financial policies under fixed and floating exchange rates. *IMF Staff Papers, 9*, 369–379.

Goldin, C. (1995). The U-shaped female labor force function in economic development and economic history. In T. P. Schultz (Ed.), *Investment in women's human capital and economic development* (pp. 61–90). Chicago: University of Chicago Press.

Grossman, G. M., & Helpman, E. (2002). Integration versus outsourcing in industry equilibrium. *Quarterly Journal of Economics, 117*, 85–120.

Hakkala, K., Norbäck, P.-J., & Svaleryd, H. (2008). Asymmetric effects of corruption on FDI: Evidence from Swedish multinational firms. *Review of Economics and Statistics, 90*, 627–642.

Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory, 20*, 381–408.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics, 123*, 441–487.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*, 305–360.

Kowalczyk, C. (2000). Welfare and integration. *International Economic Review, 41*, 483–494.

Krugman, P. (1995). Growing world trade: Causes and consequences. *Brookings Papers on Economic Activity, 26*, 327–362.

Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A new estimate of transaction costs. *Review of Financial Studies, 12*, 1113–1141.

Melitz, J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*, 1695–1725.

Mortensen, D. T. (1982). Property rights in mating, racing, and related games. *American Economic Review, 72*, 968–979.

Mundell, R. A. (1963). Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economics and Political Science, 29*, 475–485.

Rivera-Batiz, L. A., & Romer, P. M. (1991). Economic integration and endogenous growth. *Quarterly Journal of Economics, 106*, 531–555.

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review, 94*, 98–114.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers, 3*, 135–146.

Yang, T., & Percival, R. (2009). The emergence of global environmental law. *Ecology Law Quarterly, 36*, 615–664.

Part I
Search, Matching, and Globalization

Product and Labor Market Entry Costs, Underemployment and International Trade

Spiros Bougheas and Raymond Riezman

Abstract We develop a small, open economy, two-sector model with heterogeneous agents and endogenous participation in a labor matching market. There are two types of agents: workers and entrepreneurs. Both populations are heterogeneous. Workers are distinguished by their potential ability as skilled workers and entrepreneurs by their potential ability to manage a firm. To capture the notion of decentralized labor markets we assume random matching. Those agents on the long side of the market who are not matched find employment in the unskilled sector as do those agents who decided not to attempt to enter the matching market. The output of matched pairs is a function of the two partners' abilities. We find that disparities in labor institutions become a source of comparative advantage. The exact patterns will depend not only on the costs of entering the skilled sector but also on the mechanism used for dividing the surplus. We analyze the implications of asymmetric market entry costs for the patterns of international trade and underemployment. We find that if labor market inefficiencies are sufficiently strong trade liberalization can lead to welfare losses. We also examine the robustness of our results when we allow for complementarities in the production function and for alternative matching mechanisms.

1 Introduction

Establishing a competitive advantage in high-skilled sectors at the national level requires that a number of conditions must be met. The Ricardian theory of international trade emphasizes the need for technological know-how while from the Heckscher-Ohlin-Vanek model we learn that a sufficient endowment of skilled labor is necessary. While endowments and technologies are necessary pre-conditions they are by no means sufficient. Neoclassical trade theory is silent about the product and labor market institutions which play an important role in bringing the factors

S. Bougheas (✉)
School of Economics, University of Nottingham, Nottingham NG7 2RD, UK
e-mail: spiros.bougheas@nottingham.ac.uk

R. Riezman
University of Iowa, GEP, CESifo, Iowa City, IA, USA

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_2

of production together. In particular, both the entry of workers into skilled labor markets and the establishment of new enterprises are costly.

When these costs are sufficiently high they discourage market participation. For example, Brixiova et al. (2009) and Fan et al. (1999) suggest that the reluctance of workers to enter skilled labor markets can explain shortages of skilled labor in emerging economies and the consequent slow development of their private sector. In contrast, relatively low skill acquisition costs and minimal labor market frictions can potentially explain the phenomenon of overeducation and mismatch observed by researchers in many European countries, United States and Canada.[1] For example, Maynard et al. (2006) summarizing research results by sociologists, psychologists and operation researchers report that in the United States and the United Kingdom at least one in five workers experience underemployment. Further evidence comes from studies looking at the labor market impact of the economic crisis that began in 2007 which find that underemployment levels have substantially risen since the onset of the crisis reaching 25 % in many parts of the world (McKee-Ryan and Harvey, 2011; Bell and Blanchflower, 2011).

Looking at the other side of the labor market, Djankov et al. (2002) provide evidence that market entry costs incurred by start-up firms are significant and vary widely across countries. They find that "The official cost of following required procedures for a simple firm ranges from under 0.5 % of per capita GDP in the United States to over 4.6 times per capita GDP in the Dominican Republic, with the worldwide average of 47 % of annual per capita income."

In addition to market entry costs, we also need to consider frictions arising during the matching process of skilled workers to firms. The decision of young people to acquire skills is going to depend, in addition to any direct costs, on their expectations about the probability of getting a job in the skilled sector and, given that they do find a job, on the quality of the match. Similarly, the decision of potential entrepreneurs to establish new firms will depend on their expectations about the future availability of skilled labor and the latter's level of skills. Furthermore, both parties decisions will depend on the allocation of the surplus generated by the match.

These issues are well understood by labor economists.[2] In this paper, we analyze some of the implications for international trade. We develop a two-sector model with three factors of production; namely, unskilled labor, skilled labor and entrepreneurial ability.[3] One sector produces a low-tech good with a constant returns to scale technology that requires only unskilled labor. The second sector is a high-

[1] See McGuiness (2006) for a review of this literature.

[2] For example, the need for coordination between skill acquisition and job creation in order to avoid situations where the economy is locked in a low-skill/bad-job trap is emphasized by both Snower (1996) and Redding (1996).

[3] A simplified version of the model with one-sided uncertainty has been used by Bougheas and Riezman (2007) to examine the relationship between the distribution of endowments and the patterns of trade and by Davidson and Matusz (2006) and Davidson et al. (2006) to examine redistribution policy issues.

tech sector. To establish a production unit in that sector a skilled worker needs to be matched with an entrepreneur.

There are two types of agents: workers and entrepreneurs. Both populations are heterogeneous. Workers are distinguished by their potential ability as skilled workers and entrepreneurs by their potential ability to manage a firm. Initially, each type must decide whether to enter the matching market. Workers who decide to enter incur a fixed cost related to the acquisition of skills. Entrepreneurs who opt to enter incur a cost for establishing a new firm. To capture the notion of decentralized labor markets we assume random matching. Those agents on the long side of the market who are not matched find employment in the unskilled sector as do those agents who decided not to attempt to enter the matching market. The output of matched pairs is a function of the two partners' abilities.

Not surprisingly, we find that disparities in labor institutions become a source of comparative advantage. The exact patterns will depend not only on the costs of entering the skilled sector but also on the mechanism used for dividing the surplus. This suggests that in addition to traditional sources of comparative advantage, i.e. endowments and technologies, we also need to take into account those costs related to the acquisition of skills and those costs related to the creation of firms and the institutional structure of labor markets (unions, minimum wages, etc.). Thus, our work is related to a group of papers suggesting that differences in labor market rigidities across nations can be a major driving force of comparative advantage (Krugman 1995; Davis 1998a,b; Kreickemeier and Nelson 2006). Research in this area has paid particular attention to rigidities that have a direct impact on wage formation. In contrast, our main interest is on cross-country differences in (a) the costs of establishing new firms, and (b) the costs of entering skilled labor markets. Finally, our work is also related to some recent theoretical work that explores the implications of trade liberalization for inequality and labor market outcomes.[4]

Traditionally, matching models also include a search process thus generating frictional unemployment (see, for example, Davidson et al. 1999; Davidson et al. 2008; Felbermayr et al. 2011; Felbermayr et al. 2013). Our main concern is to examine issues related to long-term mismatch associated with underemployment. Our model generates either underemployment of skills or firm capacity that is not utilized depending of which side of the market is long. In particular, we show that unless the sharing rule satisfies a condition that it is equivalent to the one derived in Hosios (1990) for search models using the matching function, there will always be some type of imbalance. Clearly, as Mortensen and Wright (2002) argue, there is no good reasons to believe that when the sharing rule is taken as a primitive it will satisfy the Hosios condition. We demonstrate that the effect of trade liberalization on

[4]In Helpman et al. (2010) although both populations of firms and entrepreneurs are heterogeneous it is only the participation of the second group that is derived endogenously. Egger and Kreickemeier (2012) analyze a model with one heterogenous population and generalized endogenous participation where agents in addition to their level of skills also decide in which sector to be employed. In our model, both workers and entrepreneurs can choose whether or not to enter the matching market.

underemployment will depend on the pattern of trade. More specifically, we find that trade increases underemployment when the country has a comparative advantage in the high-tech sector. The level of underemployment will also depend on the sharing rule that divides the surplus between workers and entrepreneurs. Here, we find that the likelihood that the small-open economy has a comparative advantage in the high-tech sector is decreasing with the level of underemployment in autarky.

Most of our analytical results are derived from a benchmark version of our model that includes a linear production technology and a one-to-one matching mechanism. In Sect. 2 we develop the model and examine the autarky case and then in Sect. 3 we open the small-economy to international trade. In Sect. 4 we analyze two extensions of the benchmark version of our model. First, we allow for complementarities in the production function and we use this extended version to explore the welfare implications of trade liberalization. We show that trade can potentially be welfare reducing. We also identify conditions under which the patterns of international trade are not optimal. Second, we also examine alternative matching mechanisms and show that our results are fairly robust. We offer some final comments in Sect. 5.

2 The Closed-Economy Benchmark Model

The economy is populated by two types of agents and produces two goods. The two types of agents, workers and entrepreneurs, are each of unit mass. The first good, the numeraire, is a high-tech product and its production requires the joint efforts of an entrepreneur and a worker. The second good is a primary commodity and all types of agents can produce one unit should they decide to seek employment in that sector. Let P be its price in numeraire units. All agents are risk neutral, form expectations rationally and have identical Cobb-Douglas preferences allocating equal shares of their income on each good which implies that real income is equal to nominal income divided by $\sqrt{P}$.[5]

The populations of both workers and entrepreneurs are heterogeneous. Workers are differentiated by their ability α to work in the high-tech sector and entrepreneurs by their ability z to manage in the high-tech sector. Both α and z are private information and are randomly drawn from uniform distributions with support $[0, 1]$. Both workers and entrepreneurs have to incur a fixed utility (real income) cost $0 < \gamma < 1$ and $0 < c < 1$, respectively, to enter the high-tech sector.[6] Entrepreneurs and workers that have incurred the fixed entry costs are randomly matched. If the

[5] Let X denote the level of consumption of the high-tech product, Y the level of consumption of the primary commodity and I the level of nominal income. By maximizing $\sqrt{XY}$ subject to $I = PX + Y$, we obtain the solutions $X = \frac{I}{4P}$ and $Y = \frac{I}{4}$, which after substituting them back in the utility function and multiplying by 2 (because (a) the marginal utility of income is equal to 1, and (b) the measure of agents is equal to 2) we obtain the solution in the text.

[6] In an earlier version of the paper, we had the costs denominated in numeraire units (units of the high-tech good). As a result the relative price of the two goods depended on the size of these costs.

two masses are not equal then unmatched agents enter the primary sector. Matched pairs produce $\alpha + z$ units of the high-tech product.

To complete the description of the model we need to specify how matched pairs divide their joint output. The division of surplus normally depends on the outside options of the two parties and their relative bargaining power. Given that we have assumed away any recontracting the outside options of the two sides are the same and equal to P the income they will receive in their alternative employment option. Denote by β the share of output allocated to entrepreneurs. In this section we shall assume that all pairs divide the surplus equally, i.e. $\beta = \frac{1}{2}$. As we will see below, assuming equal division is analytically convenient and allows for analytical derivations. We will explore numerically the consequences of relaxing this restriction.[7]

Given that an agent's expected payoff is increasing in her own ability there exist two cut-off ability levels α^* and z^* such that all workers with ability levels less than α^* and all entrepreneurs with ability levels less than z^* do not incur the high-tech sector entry costs and find employment in the primary sector. Thus, a mass of workers of $1-\alpha^*$ and a mass of entrepreneurs of $(1-z^*)$ will enter the matching market. The decisions to enter the high-tech sector, and the cut-off levels, will depend on each agent's belief about their likelihood of being matched. Thus, there are three cases to consider that correspond to three potential rational expectations equilibria, namely **matching market clearing** $(1-\alpha^*) = (1-z^*)$, **surplus of entrepreneurs** $(1-\alpha^*) < (1-z^*)$, and **surplus of workers** $(1-\alpha^*) > (1-z^*)$. The one that prevails will depend on the values of the various model parameters. In particular, it will depend on the difference between the two entry costs and the level of bargaining power. When $\beta = \frac{1}{2}$ in the benchmark model, the equilibrium type depends only on the relative size of the two entry costs. Thus, without any loss of generality we assume that $c < \gamma$ in which case in equilibrium, as we verify below, there will be a mass of entrepreneurs who incur the fixed cost of entry but are not matched.

By definition an entrepreneur with ability z^* is indifferent between investing and market search and directly entering the primary sector. Given that the income of this threshold agent is equal to z^* if matched and equal to P if unmatched, the equilibrium condition for the cut-off level is given by

$$\frac{1}{2}\left(\frac{1-\alpha^*}{1-z^*}\right)\left(z^* + \frac{1+\alpha^*}{2}\right) + \left(1 - \frac{1-\alpha^*}{1-z^*}\right)P - c = P \tag{1}$$

where $\frac{1-\alpha^*}{1-z^*}$ is the probability the entrepreneur is matched with a worker and $z^* + \frac{1+\alpha^*}{2}$ is equal to the expected output of a matched pair where the entrepreneur has

By eliminating this bias we have simplified many derivations and we were able to derive some additional results.

[7] Acemoglu (1996) also employs Nash bargaining in a random matching environment similar to the one in this paper.

ability equal to z^* keeping in mind that only those workers with ability higher than α^* are attempting to enter the high-tech sector. The first term is multiplied by $\frac{1}{2}$ which is equal to the share of output received by each member of a matched pair. Similarly, α^* is determined by

$$\frac{1}{2}\left(\alpha^* + \frac{1+z^*}{2}\right) - \gamma = P \tag{2}$$

To close the model we need the equilibrium condition for one of the two goods markets. Without loss of generality we focus on the market for the primary commodity

$$2\alpha^* = \frac{2\alpha^* P + (1-\alpha^*)\frac{2+\alpha^*+z^*}{2}}{2P} \tag{3}$$

The left-hand side is equal to the gross supply of the primary commodity. All workers that enter the matching market are matched and thus there are α^* unmatched workers which means there are α^* unmatched entrepreneurs. Therefore, in total there is a mass of $2\alpha^*$ agents that are employed in the primary sector and each produces one unit. The right-hand side is equal to the gross demand. The specification of preferences imply that an agent with income y demands an amount $\frac{y}{2P}$ of the primary commodity. Furthermore, risk-neutrality implies that the marginal utility of income is constant and thus, for the derivation of the gross market demand it suffices to derive aggregate income and divide it by $2P$. Agents employed in the primary sector produce one unit and earn income P and the first term of the numerator on the right-hand side shows their gross income. The second term is equal to the total income of matched pairs.[8]

In the next Proposition we verify that the solution of the above system, that solves for the three endogenous variables α^*, z^* and P, is indeed a rational expectations equilibrium.

Proposition 1 *If $\gamma > c$ then $z^* < \alpha^*$.*

Proof (1) and (2) imply that $\frac{1}{2}\left(z^* + \frac{1+\alpha^*}{2}\right) - \frac{1-z^*}{1-\alpha^*}c = \frac{1}{2}\left(\frac{1+z^*}{2} + \alpha^*\right) - \gamma$. The equality can be written as $\frac{1}{4}(1-\alpha^*)(\alpha^* - z^*) = \gamma - \frac{1-z^*}{1-\alpha^*}c$. For $\gamma = c$ the last expression can be written as $\frac{1}{4}(1-\alpha^*)^2(\alpha^* - z^*) = \gamma(z^* - \alpha^*)$. Given that $\gamma > 0$ it follows that $\alpha^* = z^*$. Next consider the case $\gamma > c$ and let $\gamma \equiv c + \delta$. Now we

[8]For the derivation of the last term, given that the output of a matched pair is equal to the sum of the abilities of its members, it suffices to add individual abilities. Thus, we have that aggregate income of matched pairs equals

$$\int_{\alpha^*}^{1} \alpha d\alpha + \frac{1-\alpha^*}{1-z^*}\int_{z^*}^{1} z dz$$

Notice that second term follows from random matching and $z^* < \alpha^*$.

can write the equality as $\frac{1}{4}(1-\alpha^*)^2(\alpha^*-z^*)=c(z^*-\alpha^*)+\delta$. Given that $\delta>0$ and given that when $\gamma=c$, $\frac{1}{4}(1-\alpha^*)^2(\alpha^*-z^*)=c(z^*-\alpha^*)$ we have $\alpha^*-z^*>0$ which completes the proof.

Remark 1 There is another equilibrium where nobody participates in the matching market given that it is the best response for each type of agent not to participate if she believes that no agent of the other type will participate. However, this equilibrium is unstable given that any small deviation from any of the two types increases significantly the participation payoff of the other type. The interior equilibrium described by conditions (1), (2) and (3) is a unique strict strategy equilibrium. This is because in our model agents are heterogeneous.

2.1 Entry Costs and the Autarky Price

Comparative advantage is completely determined by comparing the autarky price with the foreign price and in the benchmark model the autarky price depends only on the two entry costs. With that in mind, in this section, we examine how changes in these costs affect the autarky price. The following lemma will be useful for subsequent proofs.

Lemma 1 $\alpha^*>\frac{1}{2}$.

Proof See the Appendix.

The following Proposition describes some important comparative static results.

Proposition 2 *Suppose that* $\gamma>c$. *Then, we have (a)* $\frac{d\alpha^*}{d\gamma}>0$, *(b)* $\frac{dz^*}{dc}>0$, *(c)* $\frac{d\alpha^*}{dc}<0$ *and (d)* $\frac{dz^*}{d\gamma}\gtrless 0$.

Proof See the Appendix

It is not surprising that an increase in γ discourages workers from participating in the matching market and thus the overall effect is to increase α^*. Similarly, an increase in c discourages the entry of entrepreneurs in the matching market, i.e. z^* increases. Now, consider the effect of an increase in any of the two entry costs on the entry decisions in the other side of the market. Payoffs depend on the abilities of both agents so any increase in the threshold level of either workers' or entrepreneurs' entry costs increases the expected payoff of the other type of agent and thus their incentive to participate. Thus, an increase in c has a positive effect on workers' payoffs thus providing a positive incentive for workers to participate, (α^* falls.). However, the effect of an increase in γ on the entrepreneurs' entry decision is ambiguous due to a second effect. Given that there are unmatched entrepreneurs, an increase in γ discourages the entry of workers which means that the likelihood of potential entrepreneurs being matched *declines* which discourages their entry. It is clear that this effect is larger the wider the gap between the two entry costs.

Next, we examine how entry costs affect autarky prices.

Proposition 3 *Suppose that $\gamma > c$. Then, (a) $\frac{dP}{dc} > 0$ and (b) $\frac{dP}{d\gamma} < 0$.*

Proof See the Appendix

The effect of a change in c on the autarky price is positive. This is because the decline in the participation rate by entrepreneurs increases the worker's expected payoff thus further increasing their participation rate. Thus, since there is a surplus of entrepreneurs, the supply of the high-tech product increases and this results in an increase in the autarky price. An increase in γ discourages the participation of workers in the matching market and as a consequence both the production of the high-tech product and the autarky price decline.

3 International Trade

We now consider international trade. Let P^T denote the international price. It is clear that if $P^T > P$ the economy will export the primary commodity and if $P^T < P$ the economy will export the high-tech product. The following Proposition follows directly from Proposition 3.

Proposition 4 *Suppose that $\gamma > c$. Then, other things equal, economies with higher labor entry costs will export the primary commodity and economies with higher entrepreneur entry costs will export the high-tech product.*

Remark 2 In the statement of the Proposition the qualifier 'other things equal' is there to remind us that the pattern of international trade will depend not only on cross country differences in the gap between the two costs but also on the levels. The prediction will be reversed if we set entrepreneur entry costs higher than labor entry costs.

Using the results stated in Proposition 4 we are able to make the following generalizations. Consider two countries A and B. If country B's higher entry cost is higher than country A's higher entry cost and country B's lower entry cost is lower than country A's lower entry cost then country B's autarky price will be lower than country A's autarky price. The last statement is due to the symmetry of the model which implies that the autarky price remains the same if we switch the entry costs of the two markets. In contrast, we cannot make any general statements about other rankings of entry costs.

3.1 Underemployment and Trade

We know from the autarky case that when entry costs are asymmetric in equilibrium there are some agents who entered the matching market but were not matched. The total real income loss of unmatched agents due to entry costs $(\alpha^* - z^*)c$ provides

a measure of inefficiency. As the following proposition demonstrates the effect of international trade on inefficiency depends on the pattern of trade.

Proposition 5 *As the economy moves from autarky to free trade the measure of inefficiency declines when the economy exports the primary commodity and increases when the economy exports the high-tech product.*

Proof [9]Setting $P = P^T$, rearranging and totally differentiating equations (1) and (2) we get the new system of equations

$$\frac{1}{2}d\alpha + \frac{1}{4}dz = dP^T$$
$$\left(\frac{1}{4} - \frac{1-z}{(1-\alpha)^2}c\right)d\alpha + \left(\frac{1}{2} + \frac{c}{1-\alpha}\right)dz = dP^T$$

The determinant of the new system is equal to

$$\Delta = \frac{3}{16} + \frac{1}{4}\frac{1-z}{(1-\alpha)^2}c + \frac{1}{4}\frac{c}{1-\alpha} > 0$$

Then,

$$\frac{d\alpha}{dP^T} = \frac{\left(\frac{1}{4} + \frac{c}{1-\alpha}\right)}{\Delta} > 0$$
$$\frac{dz}{dP^T} = \frac{\left(\frac{1}{4} + \frac{1-z}{(1-\alpha)^2}c\right)}{\Delta} > 0$$

Lastly,

$$\frac{d\alpha}{dP^T} - \frac{dz}{dP^T} = \frac{\frac{c}{1-\alpha}\left(\frac{z-\alpha}{1-\alpha}\right)}{\Delta} < 0$$

Suppose that $P < P^T$. In this case the world price is higher than the autarky price so that the economy exports the primary product. The increase in the price will reduce inefficiency.

The intuition for this result is that if the economy exports the primary product when liberalized then trade will draw resources into that sector and out of the high-tech sector. The high-tech sector is where the matching inefficiencies occur and hence efficiency increases as trade increases.[10]

[9]The * have been suppressed.

[10]It is clear that symmetry implies that the result does not depend on which side of the market is short. What matters for our conclusions is the presence of underemployment and not on the type of underemployment.

3.2 Division of Surplus and Trade

To this point we have assumed that workers and entrepreneurs share firm output equally. However, it is clear that any change in the division rule will affect all entry decisions and the autarky price. When the two parties share output equally but worker entry costs are higher than those of entrepreneurs it is not surprising that in equilibrium there is a surplus of entrepreneurs. Below we demonstrate that there always exists a sharing rule such that the two equilibrium cut-off levels are equal, i.e. $\alpha^* = z^* = x$. Denote by β^* the value of β that sets $\alpha^* = z^* = x$. Substituting these expressions in equilibrium conditions (1) and (2) we get

$$\beta^* \left(\frac{1+3x}{2} \right) - c = P$$

$$(1-\beta^*) \left(\frac{1+3x}{2} \right) - \gamma = P$$

Eliminating the autarky price from the first two conditions and rearranging we obtain

$$\beta^* = \frac{1}{2} - \frac{\gamma - c}{1+3x}$$

The solution is very intuitive.[11] When the two entry costs are equal we also need to set the shares allocated to each side equal so that the entry masses of workers and entrepreneurs are also equal. If entrepreneur entry costs are higher then we need to increase the share of output allocated to entrepreneurs. The exact amount will depend on the gap between the two costs and their level.

Two countries that differ in their sharing rules but are otherwise identical will have different autarky prices and thus both can benefit from trade. Hence, we would like to know how a change in the sharing rule, affects a small open economy's pattern of trade. More specifically, suppose that we increase the share of output allocated to entrepreneurs, i.e. β increases. As Table 1 indicates, the effect on the autarky price will depend on the relationship between β and β^*.

The numerical results suggest that when $\beta < \beta^*$, an increase in the share of output allocated to entrepreneurs results in a higher autarky price and when $\beta > \beta^*$ the autarky price falls as β increases. Therefore, the autarky price reaches a maximum when $\beta = \beta^*$. Below we prove that this is indeed the case for values of γ

[11] Our solution for β^* corresponds to the Hosios (1990) condition derived from search models using the matching function.

Table 1 Sharing rule and the autarky price

β	P	α	z
(a) $c = 0.3$, $\gamma = 0.4$			
0.3	0.187	0.608	0.821
0.4	0.313	0.610	0.727
0.467 (β^2)	0.405	0.674	0.674
0.5	0.355	0.700	0.620
0.6	0.225	0.788	0.547
(b) $c = 0.3$, $\gamma = 0.8$			
0.2	0.073	0.875	0.929
0.3	0.209	0.772	0.811
0.348 (β^2)	0.273	0.764	0.764
0.4	0.204	0.813	0.721
0.5	0.094	0.907	0.763
(c) $c = 0.7$, $\gamma = 0.8$			
0.35	~ 0	~ 1	~ 1
0.4	0.061	0.921	0.941
0.472 (β^2)	0.148	0.863	0.863
0.5	0.115	0.891	0.876
0.6	~ 0	~ 1	~ 1

and c sufficiently close.[12] Letting $P(\beta; \gamma, c)$ be the autarky price as a function of the entrepreneur's share of the surplus, we obtain the following result.

Proposition 6 *For γ and c sufficiently close $P(\beta; \gamma, c)$ attains a maximum at $\beta = \beta^*$.*

Proof See the Appendix.

An important implication of the above result is that as the masses of the two types of entrants get closer the likelihood that the country has a comparative advantage in the high-tech sector goes up. This is intuitive given that when the two masses of entrants are equal underemployment and hence, inefficiency in the high-tech sector is minimized.

It is also interesting to note that with a variable sharing rule entrepreneurs are not necessarily on the long-side of the market as a relatively high proportion of output allocated to them can compensate for higher entry costs. Table 1 presents comparative static results for three distinct cases. In Table 1a, c the two entry costs are relatively close but in the former both are low while in the latter both are high. In Table 1b the gap between the two entry costs is relatively large. The results

[12] Demonstrating the result for values of γ and c sufficiently apart has proven to be a very daunting task. However, calibrations of the model (Table 1 provides just a few examples), where we have allowed the two entry costs and the sharing rule to take values in the interval [0, 1], show that the result stated in Proposition 6 is not only valid when we allow the two values to differ considerably but also that the maximum is a global maximum.

suggest that there is a monotonic effect of a change in the sharing rule on the cut-off corresponding to the short-side of the market. Keep in mind that for $\beta < \beta^*$ entrepreneurs are on the short-side of the market while for $\beta > \beta^*$ workers are on the short-side of the market. In contrast, the effect of a change in the sharing rule on the long-side is ambiguous as we have an additional effect first mentioned in Proposition 2. Given that the change affects the short-side it effects the value of a match but also affects the likelihood that an agent on the long-side of the market is matched.

4 Beyond the Benchmark Model

4.1 Skill Complementarity

In this section, we extend the benchmark model by allowing for a more general production function. More specifically, we consider the case where the skills of workers and entrepreneurs are complementary. Now, matched pairs produce $(\alpha + z)^2$ units of the high-tech product. Without any loss of generality, we are going to restrict our attention to the case where $\gamma > c$. To keep the analysis tractable we are also setting $\beta = \frac{1}{2}$. Given these restrictions, once more in equilibrium we must have $z^* < a^*$.

In this case all workers that invest in skills will be matched but only a proportion $\frac{1-\alpha^*}{1-z^*}$ of entrepreneurs will find employment in the high-tech sector. The equilibrium condition for z^* is given by

$$\frac{1}{2}\left(\frac{1-\alpha^*}{1-z^*}\right)\frac{\int_{\alpha^*}^{1}(\alpha+z^*)^2 d\alpha}{1-\alpha^*}+\left(1-\frac{1-\alpha^*}{1-z^*}\right)P-c=P \tag{4}$$

where $\frac{1}{2}\frac{\int_{\alpha^*}^{1}(\alpha+z^*)^2 d\alpha}{1-\alpha^*}$ is equal to the expected payoff of a matched entrepreneur with ability equal to the equilibrium cut-off level. The corresponding condition for α^* is given by

$$\frac{1}{2}\frac{\int_{z^*}^{1}(\alpha^*+z)^2 d\alpha}{1-z^*}-\gamma=P \tag{5}$$

Now, we turn our attention to the goods market equilibrium concentrating again on the market for the primary commodity. As before, the gross supply is equal to $2\alpha^*$. Next, we derive the gross demand of the primary commodity. As before, the specification of preferences imply that an agent with income y demands an amount $\frac{y}{2P}$ of the primary commodity. Agents employed in the primary sector produce one unit and earn income P. What remains is to derive the demand for the primary commodity by those agents who are matched.

The combined income of a matched pair comprising of an entrepreneur with ability z and a worker with ability α is equal to $(\alpha+z)^2$. In order to find the expected income of a matched pair we need to derive the distribution of $\alpha+z$ which is the sum of two independent, non-identically distributed uniform random variables.[13] More specifically, α is uniformly distributed on $[\alpha^*, 1]$ and z is uniformly distributed on $[z^*, 1]$.

Lemma 2 *The distribution density function of* $\alpha+z$ *for* $\alpha^* > z^*$ *is given by*

$$\begin{aligned} &\frac{\alpha+z-\alpha^*-z^*}{(1-\alpha^*)(1-z^*)} \qquad for \qquad \alpha^*+z^* < \alpha+z \leqslant 1+z^* \\ &\frac{1}{(1-z^*)} \qquad for \qquad 1+z^* < \alpha+z \leqslant 1+\alpha^* \\ &\frac{2-\alpha-z}{(1-\alpha^*)(1-z^*)} \qquad for \qquad 1+\alpha^* < \alpha+z \leqslant 2 \end{aligned} \tag{6}$$

Proof Lusk and Wright (1982) provide the derivation when the two random variables are non-identically but independently uniformly distributed on intervals with a lower bound equal to 0. For our more general case we apply the following transformation. Let $Z = z - z^*$ and $A = \alpha - \alpha^*$. Then Z is uniformly distributed on $[0, 1-z^*]$ and A is uniformly distributed on $[0, (1-\alpha^*)]$. Also $\alpha+z = A+Z+\alpha^*+z^*$. So it is sufficient to find the distribution of $A+Z$.

Using the above density functions we can calculate the expected output of a matched pair $(E\{(\alpha+z)^2 \mid \alpha^* \leqslant \alpha \leqslant 1, z^* \leqslant z \leqslant 1\}$. It follows that the primary market equilibrium condition is given by

$$2a^* = \frac{2\alpha^* P + (1-\alpha^*)(E\{(\alpha+z)^2 \mid \alpha^* \leqslant \alpha \leqslant 1, z^* \leqslant z \leqslant 1\})}{2P} \tag{7}$$

The first term, on the right-hand side, is equal to the income of all workers employed in the primary sector. The second term is equal to the aggregate income of matched pairs.

As in the benchmark case, the system of Eqs. (4), (5), and (7) solves for the three endogenous variables α^*, z^* and P. This new system is too complex to solve analytically but numerical calibrations of the model show that the results of Propositions 2–5 derived for the benchmark case are also valid when complementarities are present.[14] Notice that the qualitative results on the pattern of trade do not depend on the exact form of the production function. This is because here we are

[13] This of course requires that this distribution is the same as the realized distribution resulting from random matching. Alós-Ferrer (2002) has shown that this is indeed the case.

[14] The numerical results are provided in a separate Appendix that is available from the authors.

concentrating on cross-country differences in market entry costs. As Bougheas and Riezman (2007), Costinot and Fogel (2009), Grossman and Maggi (2000), Ohnsorge and Trefler (2007), and Sly (2012) have shown, this is not the case anymore when countries also differ in the distribution of endowments.

4.1.1 Matching Efficiency with Skill Complementarities

When the technology is linear what matters for efficiency is who gets matched however, it does not matter with whom they are matched. The reason is that as long as we know who is matched on each side of the matching market we can find aggregate production in that sector by adding their respective ability levels. However, this is not the case when complementarities are present. Our function is a particular case of a super-modular function. As Grossman and Maggi (2000) have demonstrated efficiency requires that we match workers and entrepreneurs with identical abilities. Thus, we are going to use this more general framework to make some observations on the gains from trade and the pattern of trade. More specifically, using an example, we are going to demonstrate that (a) trade can lead to welfare losses, and (b) the patterns of trade may be sub-optimal. What drives these results is that the competitive equilibrium under autarky is inefficient. When our model allows for skill complementarities there are two sources of inefficiency. The first one is underemployment and this source of inefficiency, as we have already demonstrated, is present even in the absence of skill complementarities. The second one is due to mismatch that arises only when the technology is not linear. A social planner by restricting entry can eliminate the first source. However, given that types are private information a social planner cannot eliminate the second source of inefficiency. Nevertheless, this type of inefficiency must be significantly lower in markets with lower search costs. To keep things simple, we have assumed that search costs are sufficiently high so that agents do not have an incentive to separate. As these costs decline matching efficiency should improve. With that in mind, in the example below we compare the competitive equilibrium with a social optimum where both inefficiencies have been eliminated.[15]

Example 1 (Efficiency of Competitive Equilibrium) Consider the example: $c = 0.5$ and $\gamma = 0.6$. We measure aggregate welfare by aggregating individual utilities yielding $W = 2\sqrt{XY}$, where X denotes the level of consumption of the high-tech product and Y the level of consumption of the primary commodity.[16] Aggregate welfare derived in autarky equilibrium, W_A^C, is given by[17]

[15] Our result that trade can potentially lead to welfare losses relies on comparisons between the competitive equilibrium under autarky and the corresponding equilibrium under trade and thus is not affected by this choice.

[16] Keep in mind that the size of the population has measure 2.

[17] See footnote 5.

$$W_A^C = \frac{2\alpha^* P + (1-\alpha^*)(E\{(\alpha+z)^2 \mid \alpha^* \leqslant \alpha \leqslant 1, z^* \leqslant z \leqslant 1\} - c - \gamma)}{2\sqrt{P}} \tag{8}$$

Substituting the above values of entry costs in (4), (5), and (7) we find that $\alpha^* = 0.64$, $z^* = 0.59$ and $P = 0.43$. Finally, substituting these values in the welfare function we find that $W_A^C = 0.84$.

Next, we compare the above solution with the social optimum level of aggregate welfare under autarky, W_A^S. We begin by setting the mass of workers participating in the matching market equal to the corresponding mass of entrepreneurs. Let x^* denote the proportion of agents who decide not to enter the matching market and let X_A^S and Y_A^S denote the representative agent's consumption levels of the high-tech product and the primary commodity correspondingly. These consumption levels are equal to the aggregate quantities produced in the economy divided by 2 (given that the measure of agents is equal to 2) and given by

$$X_A^S = \left(\int_{x^*}^{1} (2x)^2 \, dx - (c+\gamma)\left(1-x^*\right) \right) / 2 \tag{9}$$

and

$$Y_A^S = x^* \tag{10}$$

Given that efficient matching requires that agents of equal ability are matched, the first term in the brackets in (9) captures the level of aggregate production of the high-tech product. The second term is equal to the aggregate cost of entry in the matching market. Equation (10) follows from the fact that each agent employed in the primary sector produces one unit. After we substitute (9) and (10) in the welfare function we maximize the latter by choosing the proportion of agents who will find employment in the primary sector to obtain $x^* = 0.69$. Substituting the solution in (9) and (10) and then those solutions in the welfare function we get $X_A^S = 0.28$, $Y_A^S = 0.69$ and $W_A^S = 0.8746 > 0.84 = W_A^C$.

The above results show that in autarky the market equilibrium is inefficient which is not surprising given that the social planner eliminates underemployment (every agent who incurs the entry cost finds employment in the high-tech sector) and matches agents efficiently. Furthermore, given that the high-tech sector operates more efficiently, optimal participation in that sector is below the corresponding market equilibrium level.[18]

Example 2 (International Trade Under Free Competition May Reduce Welfare) Next, we consider the corresponding welfare levels under international trade when $P^T = 0.38 < P = 0.43$. Given that the international price is below the autarky price

[18]In fact, this is a third source of inefficiency due to a participation externality which is common in many matching and search models.

the small open economy has a comparative advantage in the high-tech product. By substituting the international price in (4) and (5) and solving the system we find the equilibrium cut-off participation rates for the open economy are equal to $\alpha^* = 0.61$ and $z^* = 0.56$. Substituting these values and the international price in the right hand side of (8) we find that $W_T^C = 0.81 < 0.84 = W_A^C$; thus, in this particular case, welfare under international trade is lower than welfare under autarky. The intuition for this result is that when the economy opens to trade it expands the sector in which the inefficiencies arise and in this particular case, the costs due to these inefficiencies exceed the gains from trading at a price that differs from the autarky one.

Example 3 (International Trade is Socially Efficient) We need to be very careful about interpreting the last result. To see why, let us see what a national social planner would have done when facing the same exogenous international price. The social planner, in addition to allocating agents to sectors, decides which goods and what quantities will be traded with the rest of the world. Let $\tau_X \gtrless 0$ and $\tau_Y \gtrless 0$ denote the units traded of each good, where positive numbers indicate imports and negative exports. These quantities must satisfy the trade balance condition

$$P^T \tau_Y = -\tau_X$$

The representative agent's consumption levels of the two goods are given by

$$X_T^S = X_A^S + \tau_X$$

and

$$Y_T^S = Y_A^S + \tau_Y$$

Substituting the above three conditions in the welfare function and choosing the participation rate to maximize welfare we obtain $\tau_Y = 0.024$, $x^* = 0.68$, $X_T^S = 0.27$, $Y_T^S = 0.70$ and $W_T^S = 0.875 > 0.8746 = W_A^S$.[19] This demonstrates that if the inefficiencies arising in the matching market are eliminated, trade always improves welfare.

Thus, if matching inefficiencies exist our results suggest that imposing trade restrictions might be welfare improving. However, the results also suggest that a better policy might be to improve labor and product market institutions thus facilitating more efficient matches. Once this is done, free trade is the preferred policy. So, it is not international trade that lowers welfare, rather it is labor market inefficiencies that cause welfare to fall in moving from autarky to free trade.

[19]Due to the choice of functional forms and parameter values the differences are small, however, they are robust in the sense that the qualitative results are obtained for a wide set of parameter values.

Example 4 (The Patterns of Trade Under Free Competition May be Sub-optimal) In the above example the social planner chooses to export the high-tech product and thus the equilibrium patterns of trade are optimal. But in the absence of a social planner this is not always the case. Consider the following question: what must be the international price so that the social planner would choose not to trade; i.e. $\tau_X = \tau_Y = 0$? It is clear that this would be the price that would induce the social planner to choose the same ability cut-off level as the one chosen in the case for autarky, i.e. $x^* = 0.69$.[20] We denote this price by P^S. This price solves

$$2x^* = \frac{2P^S x^* + \int_{x^*}^{1} (2x)^2 \, dx - (c+\gamma)(1-x^*)}{2P^S}$$

This is similar to (7) but now we have substituted the corresponding demand for and supply of the primary commodity given that production is determined by the social planner's allocation. Substituting the values for c, γ and x^* we obtain $P^S = 0.402$. The implication for trade patterns is that if $P^T > P^S$ then the social planner would choose to export the primary commodity and if $P^T < P^S$ the social planner would choose to export the high-tech product. If the world price, P^T lies between the autarky price under a competitive equilibrium ($P = 0.42$) and the social planner's autarky price ($P^S = 0.402$) then the equilibrium pattern of trade will not be optimal. The intuition is that market inefficiencies affect the autarky price. If the world price lies between the two autarky prices then the patterns of trade are not optimal.

4.2 Alternative Matching Mechanisms

In this section, we examine the robustness of our comparative static results to alternative matching mechanisms. Up to this point we have assumed that exactly one entrepreneur (long-side of the market) is matched with one worker leaving the rest of the entrepreneurs to seek employment in the primary sector. Given our supposition that there is no possibility of recontracting (infinite search costs) we have assumed matched agents share the surplus equally.

Before we consider any alternative mechanisms we will show that our benchmark set-up is equivalent to one in which all unmatched entrepreneurs in the benchmark case are now matched with one single worker while each one of the rest of the entrepreneurs are matched again with one worker. In this new arrangement, the worker who is matched with multiple entrepreneurs is in a strong bargaining position. Given that the production technology requires a single entrepreneur,

[20] This is an application of the second welfare theorem. Suppose that the agents in the economy are allocated to sectors by the social planner (this step follows from the fact that the equilibrium allocation is inefficient) and then exchange goods in competitive markets. The equilibrium price would be the one that decentralizes the social planner's optimal allocation under autarky.

bargaining will push the payoff of all entrepreneurs matched with the singe worker down to the outside option which in this case is equal to the price of the primary commodity. Now all these entrepreneurs are indifferent between staying in the high-tech sector and entering the primary market. Thus, in this new set up, with the exception of one pair, all other workers and entrepreneurs receive the same payoffs as those in the original set-up. The exception is because now there is one additional entrepreneur who receives the low payoff and a worker who receives a payoff that is equal to the total surplus generated by the pair minus the price of the primary commodity. Given that we have assumed that both populations are very large the two versions only differ in a set of measure zero.

Now consider the other extreme.[21] Suppose that all workers (short-side of the market) are again matched but now some of them are matched with one entrepreneur and some of them are matched with two entrepreneurs.[22] Thus, in this alternative arrangement underemployment is more evenly distributed in the economy. To keep this simple, we will ignore complementarities and focus on the linear technology case. Once more, under the supposition that $c < \gamma$, the mass of entrepreneurs who enter the matching market, $1-z^*$, will be higher than the corresponding mass of workers, $1-\alpha^*$. The proportion of workers matched with two entrepreneurs is equal to $\frac{\alpha^*-z^*}{1-\alpha^*}$ and the proportion of entrepreneurs matched with workers who are also matched with another entrepreneur is equal to $2\frac{\alpha^*-z^*}{1-z^*}$. The equilibrium condition that determines z^* is given by

$$2\frac{\alpha^*-z^*}{1-z^*}P+\left(1-2\frac{\alpha^*-z^*}{1-z^*}\right)\frac{1}{2}\left(z^*+\frac{1+\alpha^*}{2}\right)-c=P \tag{11}$$

where the left-hand side is equal to the marginal entrepreneur's expected payoff from entering the market. The equilibrium condition for α^* is given by

$$\frac{\alpha^*-z^*}{1-\alpha^*}\left(\alpha^*+\frac{1+z^*}{2}-P\right)+\left(1-\frac{\alpha^*-z^*}{1-\alpha^*}\right)\frac{1}{2}\left(\alpha^*+\frac{1+z^*}{2}\right)-\gamma=P \tag{12}$$

where if the marginal worker is matched with more than one entrepreneur they receive a payoff equal to the total surplus minus the price of the primary commodity (the entrepreneur's outside option) and if matched with a single entrepreneur they receive half the surplus. Once more, we need the market equilibrium condition (3) to close the model.

Numerical calibration shows that with one exception the comparative static results under this alternative mechanism are the same as those derived from the benchmark case.[23] The only exception relates to the effect of a change in the entry

[21] We are indebted to Carl Davidson for suggesting this alternative mechanism.

[22] Of course, if the measure of entrepreneurs who enter the matching market is more than twice the measure of corresponding workers then all workers will be matched with multiple entrepreneurs.

[23] The numerical results are provided in a separate Appendix available from the Authors.

cost of entrepreneurs on α^* that determines the mass of workers who enter the matching market. In the benchmark case we found that an increase in the entry cost of entrepreneurs has a negative effect on α^* (more workers enter) because the expected ability of entrepreneurs is now higher. This result could be reversed with the alternative matching mechanism because there is an additional effect. Namely, as the mass of entrepreneurs entering the matching market declines the likelihood that a worker will be matched with more than one entrepreneur, and thus receiving the higher payoff, also declines. This effect discourages worker entry, so the net effect is ambiguous.

The original benchmark model is a special case of the matching arrangement considered in this section. Our paper is motivated by the existence of persistent surpluses of either skilled workers or vacancies in certain labor markets, especially in emerging economies and their potential implications for comparative advantage. Our more general version suggests that it is not only underemployment that matters but also its distribution throughout the economy. To see this, think of each agent on the short-side of the market as occupying a distinct location. Agents on the long-side of the market know the locations and choose one at random.[24] There are many potential matching arrangements. Those that we have considered in this section correspond to two extreme cases.

5 Conclusions

Both workers and potential entrepreneurs who want to enter sectors that use advanced technologies must incur entry costs. For workers these costs might capture time and money spent on skill acquisition while for entrepreneurs these costs might be related to the establishment of new technologies or more directly to costly procedures related to the start-up of new enterprises. The decision to incur these costs will depend on expectations about future benefits from participating in these markets. In turn, these benefits will depend on the likelihood of finding a match and thus employment in these markets and on the productivity of that match. Competitive markets can ensure that *ex ante* all entry decisions are optimal but *ex post* it is very likely that some agents will fail to match and thus their new skills or know-how will be underemployed. Having argued that such imbalances are common we have built a simple two-sector model with heterogeneous agents in order to explore their implications for international trade.

Our first task has been to explore the impact of a change in market entry costs on competitiveness and the patterns of international trade. We have found that the results will depend on three factors. First, on the side of the market that faces the change in entry costs, second, on the distribution of underemployment in the econ-

[24] What matters is that they do not know how many others are trying to find a match in the same location.

omy, and third, on the sharing rule for dividing the surplus generated by a match. More specifically, we have found that an increase in the entry costs of the agents on the short-side of the market will decrease the international competitiveness of that sector. However, the effect of an increase in the entry costs of the long-side of the market would depend on the distribution of underemployment in the economy. Furthermore, we have shown that the lower the level of underemployment, where the latter directly depends on the sharing rule, the higher the likelihood that the sector's competitiveness is strong.

Calibrations have shown that our results also hold when we introduce complementarities in the production function. However, now in addition to inefficiencies arising because of social sub-optimal entry decisions we also have matching inefficiencies. Given that the autarkic equilibrium is not Pareto optimal it is not surprising that when the economy has a comparative advantage in the sector affected by those inefficiencies, international trade can be welfare reducing. In fact, we have also demonstrated that even the patterns of trade can be inefficient. We have also argued that the best policy response is to initiate measures that improve the functioning of the labor market rather than imposing restrictions on the cross-border movement of goods.

Appendix[25]

Proof of Lemma 1 The system of Eqs. (1), (2) and (3) can be rewritten as

$$\frac{1}{2}\left(\frac{1+z}{2}+\alpha\right)-\gamma=P \tag{13}$$

$$\frac{1}{2}\left(z+\frac{1+\alpha}{2}\right)-\frac{1-z}{1-\alpha}c=P \tag{14}$$

$$P=\frac{(1-\alpha)(2+\alpha+z)}{4\alpha} \tag{15}$$

Differentiating (15) with respect to α we get

$$\frac{\partial P}{\partial \alpha}=\frac{-2-\alpha^2-z}{4\alpha^2}<0$$

Then the difference $\frac{1}{2}\left(\frac{1+z}{2}+\alpha\right)-P=\frac{1}{2}\left(\frac{1+z}{2}+\alpha\right)-\frac{(1-\alpha)(2+\alpha+z)}{4\alpha}$ is increasing in α. When the expression is evaluated at $\alpha=\frac{1}{2}$ we find that it is equal to $-\frac{1}{8}<0$. Then, the Lemma follows from (13) and $\gamma>0$.

[25] In all proofs we have suppressed the *.

Proof of Proposition 2 By substituting (15) into (13) and (14) we can reduce the above system into two equations in the two unknowns α and z. Totally differentiating the new system we get

$$\left(\frac{1}{2}-\frac{\partial P}{\partial \alpha}\right)d\alpha+\left(\frac{1}{4}-\frac{\partial P}{\partial z}\right)dz=d\gamma \tag{16}$$

$$\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)d\alpha+\left(\frac{1}{2}+\frac{1}{1-\alpha}c-\frac{\partial P}{\partial z}\right)dz$$
$$=\left(\frac{1-z}{1-\alpha}\right)dc \tag{17}$$

where

$$\frac{\partial P}{\partial z}=\frac{1-\alpha}{4\alpha}>0$$

Next, we proceed to show that the determinant Δ is positive.

$$\Delta=\left(\frac{1}{2}-\frac{\partial P}{\partial \alpha}\right)\left(\frac{1}{2}+\frac{1}{1-\alpha}c-\frac{\partial P}{\partial z}\right)-\left(\frac{1}{4}-\frac{\partial P}{\partial z}\right)\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)$$
$$=\left[\frac{3}{16}-\frac{1}{4}\frac{\partial P}{\partial z}-\frac{1}{4}\frac{\partial P}{\partial \alpha}\right]+\left[\frac{1}{2}\frac{1}{1-\alpha}c-\frac{1}{1-\alpha}c\frac{\partial P}{\partial \alpha}+\frac{1}{4}\frac{1-z}{(1-\alpha)^2}c-\frac{1-z}{(1-\alpha)^2}c\frac{\partial P}{\partial z}\right]$$

Lemma 1 implies that $\frac{\partial P}{\partial z}<\frac{1}{4}$ and that $-\frac{\partial P}{\partial \alpha}>\frac{3}{4}$. The two inequalities imply that the expression in the first bracket is positive. The first inequality also implies that $\frac{1}{4}\frac{1-z}{(1-\alpha)^2}c-\frac{1-z}{(1-\alpha)^2}c\frac{\partial P}{\partial z}>0$ which, in turn, implies that the expression in the second bracket is also positive.

Thus, $\Delta>0$.

(a) $\Delta>0$ implies that

$$sign\left\{\frac{d\alpha^*}{dc}\right\}=sign\left\{-\left(\frac{1-z}{1-\alpha}\right)\left(\frac{1}{4}-\frac{\partial P}{\partial z}\right)\right\}=sign\left\{-\frac{1}{4}\left(\frac{1-z}{1-\alpha}\right)\frac{2\alpha-1}{\alpha}\right\}$$

where given that $\alpha>\frac{1}{2}$ is negative.

(b) $\Delta>0$ implies that

$$sign\left\{\frac{d\alpha^*}{d\gamma}\right\}=sign\left\{\frac{1}{2}+\frac{1}{1-\alpha}c-\frac{\partial P}{\partial z}\right\}>0$$

Given that $\frac{\partial P}{\partial z}<\frac{1}{4}$ the expression is positive.

(c) $\Delta > 0$ implies that

$$sign\left\{\frac{dz^*}{d\gamma}\right\} = sign\left\{-\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)\right\}$$
$$= sign\left\{-\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c+\frac{2+\alpha^2+z}{4\alpha^2}\right)\right\}$$
$$= sign\left\{-4\alpha^2(1-z)c+(1-\alpha)^2\left(2+2\alpha^2+z\right)\right\} \gtreqless 0$$

For sufficiently high values of γ (high α) the expression will be negative. For low values of γ the sign will depend on c, and given that an increase in c implies an increase in z (see below), for relative extreme values of c the expression will be positive.

(d) $\Delta > 0$ implies that

$$sign\left\{\frac{dz^*}{dc}\right\} = sign\left\{\left(\frac{1}{2}-\frac{\partial P}{\partial \alpha}\right)\left(\frac{1-z}{1-\alpha}\right)\right\} > 0$$

where the inequality follows from $\frac{\partial P}{\partial \alpha} < 0$.

Proof of Proposition 3 (a) Totally differentiating (13) we get

$$\frac{dP}{dc} = \frac{1}{4}\frac{dz}{dc}+\frac{1}{2}\frac{d\alpha}{dc}$$

Given that $\Delta > 0$ the sign of the above expression is the same as the sign of

$$\frac{1}{4}\left(\frac{1}{2}-\frac{\partial P}{\partial \alpha}\right)\left(\frac{1-z}{1-\alpha}\right)-\frac{1}{2}\left(\frac{1-z}{1-\alpha}\right)\left(\frac{1}{4}-\frac{\partial P}{\partial z}\right)$$
$$= \frac{1}{4}\left(\frac{1-z}{1-\alpha}\right)\left(\left(\frac{1}{2}+\frac{2+\alpha^2+z}{4\alpha^2}\right)-\frac{1}{2}\left(\frac{2\alpha-1}{\alpha}\right)\right)$$
$$= \frac{1}{4\alpha}\left(\frac{1-z}{1-\alpha}\right)\left(2-\alpha^2+2\alpha+z\right) > 0$$

(b) Totally differentiating (13) we get

$$\frac{dP}{d\gamma} = \frac{1}{4}\frac{dz}{d\gamma}+\frac{1}{2}\frac{d\alpha}{d\gamma}-1$$

The first two terms are equal to

$$\left\{-\frac{1}{4}\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)+\frac{1}{2}\left(\frac{1}{2}+\frac{1}{1-\alpha}c-\frac{\partial P}{\partial z}\right)\right\}/\Delta$$

To complete the proof we need to show that the numerator is less than Δ

$$
\begin{aligned}
&-\frac{1}{4}\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)+\frac{1}{2}\left(\frac{1}{2}+\frac{1}{1-\alpha}c-\frac{\partial P}{\partial z}\right)-\\
&\left[\frac{3}{16}-\frac{1}{4}\frac{\partial P}{\partial z}-\frac{1}{4}\frac{\partial P}{\partial \alpha}\right]-\left[\frac{1}{2}\frac{1}{1-\alpha}c-\frac{1}{1-\alpha}c\frac{\partial P}{\partial \alpha}+\frac{1}{4}\frac{1-z}{(1-\alpha)^2}c-\frac{1-z}{(1-\alpha)^2}c\frac{\partial P}{\partial z}\right]\\
&=\left(\frac{1}{2}+\frac{1}{1-\alpha}c\right)\frac{\partial P}{\partial \alpha}-\left(\frac{1}{4}-\frac{1-z}{(1-\alpha)^2}c\right)\frac{\partial P}{\partial z}\\
&=-\frac{1}{16\alpha^2(1-\alpha)}\left((2(1-\alpha)+4c)\left(2+\alpha^2+z\right)+\alpha(1-\alpha)^2-4\alpha(1-z)c\right)\\
&=-\frac{1}{16\alpha^2(1-\alpha)}\big(4(1-\alpha)+2(1-\alpha)\alpha^2+2(1-\alpha)\\
&\qquad\qquad\times z+8c+4\alpha^2c+4zc+\alpha(1-\alpha)^2-4\alpha(1-z)c\big)
\end{aligned}
$$

The proof follows from $4\alpha(1-z)c<4c<8c$.

Proof of Proposition 6 The function $P(\beta;\gamma,c)$ is continuous but not differentiable at $\beta=\beta^*$.

For $\beta>\beta^*$ the equilibrium of the model is determined by

$$(1-\beta)\left(\frac{1+z}{2}+\alpha\right)-\gamma=P \tag{18}$$

$$\beta\left(z+\frac{1+\alpha}{2}\right)-\frac{1-z}{1-\alpha}c=P \tag{19}$$

$$P=\frac{(1-\alpha)(2+\alpha+z)}{4\alpha} \tag{20}$$

Given that the definition of β^* implies that $\alpha^*>z^*$ we have obtained the above system from the system (13), (14) and (15) after setting the entrepreneur's share of surplus equal to β. Let $P^+(\beta)$ denote the price function for $\beta>\beta^*$. We will sow that $\frac{dP^+(\beta)}{d\beta}_{(\beta=\beta^*)}=\left(\frac{\partial P}{\partial \alpha}\frac{d\alpha}{d\beta}+\frac{\partial P}{\partial z}\frac{dz}{d\beta}\right)_{(\beta=\beta^*)}<0$. Notice that given that the market clearing condition (20) does not directly depend on β the expressions for $\frac{\partial P}{\partial \alpha}$ and $\frac{\partial P}{\partial z}$ are the same as those derived in the proof of Proposition 2. After substituting (20) in (18) and (19) and totally differentiating we get

$$\left((1-\beta)-\frac{\partial P}{\partial \alpha}\right)d\alpha+\left(\frac{1-\beta}{2}-\frac{\partial P}{\partial z}\right)dz=\left(\frac{1+z}{2}+\alpha\right)d\beta \tag{21}$$

$$\left(\frac{\beta}{2}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)d\alpha+\left(\beta+\frac{c}{1-\alpha}-\frac{\partial P}{\partial z}\right)dz=-\left(z+\frac{1+\alpha}{2}\right)d\beta \tag{22}$$

Following the same steps as those used in the proof of Proposition 2 we can show that the determinant is positive for β close to $\frac{1}{2}$, and we have already shown that as γ and c come closer together β^* approaches $\frac{1}{2}$. Then,

$$sign\left\{\frac{\partial P}{\partial \alpha}\frac{d\alpha}{d\beta}+\frac{\partial P}{\partial z}\frac{dz}{d\beta}\right\}_{\beta=\beta^*}$$

$$=sign\left\{\begin{array}{c}\frac{\partial P}{\partial \alpha}\left(\left(\frac{1+z}{2}+\alpha\right)\left(\beta+\frac{c}{1-\alpha}-\frac{\partial P}{\partial z}\right)+\left(z+\frac{1+\alpha}{2}\right)\left(\frac{1-\beta}{2}-\frac{\partial P}{\partial z}\right)\right)\\ -\frac{\partial P}{\partial z}\left(\left((1-\beta)-\frac{\partial P}{\partial \alpha}\right)\left(z+\frac{1+\alpha}{2}\right)+\left(\frac{\beta}{2}-\frac{1-z}{(1-\alpha)^2}c-\frac{\partial P}{\partial \alpha}\right)\left(\frac{1+z}{2}+\alpha\right)\right)\end{array}\right\}_{\beta=\beta^*}$$

Setting $\alpha=z=x$ and simplifying we get

$$\frac{\partial P}{\partial \alpha}\left(\frac{1}{2}+\frac{\beta}{2}+\frac{c}{1-x}\right)-\frac{\partial P}{\partial z}\left(1-\frac{\beta}{2}-\frac{c}{1-x}\right)<0$$

The inequality follows from $\frac{\partial P}{\partial z}<\frac{1}{4}$ and $-\frac{\partial P}{\partial \alpha}>\frac{3}{4}$.

To complete the proof we need to show that when $\beta<\beta^*$, $\frac{dP^-(\beta)}{d\beta}_{(\beta=\beta^*)}=\left(\frac{\partial P}{\partial \alpha}\frac{d\alpha}{d\beta}+\frac{\partial P}{\partial z}\frac{dz}{d\beta}\right)_{(\beta=\beta^*)}>0$, where $P^-(\beta)$ denotes the corresponding price function. In this case, the equilibrium is given by

$$(1-\beta)\left(\frac{1+z}{2}+\alpha\right)-\frac{1-\alpha}{1-z}\gamma=P \tag{23}$$

$$\beta\left(z+\frac{1+\alpha}{2}\right)-c=P \tag{24}$$

$$P=\frac{(1-z)(2+\alpha+z)}{4z} \tag{25}$$

Notice that we can obtain (25) by substituting γ for c, c for γ, α for z and z for α in (18). Thus the partial derivatives $\frac{\partial P}{\partial \alpha}$ and $\frac{\partial P}{\partial z}$ are given by

$$\frac{\partial P}{\partial \alpha}=\frac{1-z}{4z}>0$$

and

$$\frac{\partial P}{\partial z}=\frac{-2-z^2-\alpha}{4z^2}<0$$

After substituting (25) in (23) and (24) and totally differentiating we get

$$\left((1-\beta)+\frac{\gamma}{1-z}-\frac{\partial P}{\partial \alpha}\right)d\alpha+\left(\frac{1-\beta}{2}-\frac{1-\alpha}{(1-z)^2}\gamma-\frac{\partial P}{\partial z}\right)dz=\left(\frac{1+z}{2}+\alpha\right)d\beta \tag{26}$$

$$\left(\frac{\beta}{2}-\frac{\partial P}{\partial \alpha}\right)d\alpha+\left(\beta-\frac{\partial P}{\partial z}\right)dz=-\left(z+\frac{1+\alpha}{2}\right)d\beta \tag{27}$$

Because of symmetry the determinant of this system is identical to the one for the previous case for $\beta=\frac{1}{2}$. (Let the old matrix be $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$. Then the new matrix is $\begin{bmatrix} D & C \\ B & A \end{bmatrix}$. The determinant has not been affected by these changes, however, when the signs were different across a diagonal before now the positive sign has become negative and the other way around). Thus, at least for values of β close to $\frac{1}{2}$, the new determinant will also be positive. Then the sign of $\frac{dP^{-}(\beta)}{d\beta}_{(\beta=\beta^*)}$ is the same as

$$sign\left\{\frac{\partial P}{\partial \alpha}\frac{d\alpha}{d\beta}+\frac{\partial P}{\partial z}\frac{dz}{d\beta}\right\}_{\beta=\beta^*}$$

$$=sign\left\{\begin{array}{l}\frac{\partial P}{\partial \alpha}\left(\left(\frac{1+z}{2}+\alpha\right)\left(\beta-\frac{\partial P}{\partial z}\right)+\left(z+\frac{1+\alpha}{2}\right)\left(\frac{1-\beta}{2}-\frac{1-\alpha}{(1-z)^2}\gamma-\frac{\partial P}{\partial z}\right)\right)\\ -\frac{\partial P}{\partial z}\left(\left((1-\beta)+\frac{\gamma}{1-z}-\frac{\partial P}{\partial \alpha}\right)\left(z+\frac{1+\alpha}{2}\right)+\left(\frac{\beta}{2}-\frac{\partial P}{\partial \alpha}\right)\left(\frac{1+z}{2}+\alpha\right)\right)\end{array}\right\}_{\beta=\beta^*}$$

Setting $\alpha=z=x$ and simplifying we get

$$\frac{\partial P}{\partial \alpha}\left(\frac{1}{2}+\frac{\beta}{2}-\frac{\gamma}{1-x}\right)-\frac{\partial P}{\partial z}\left(1-\frac{\beta}{2}+\frac{\gamma}{1-x}\right)>0$$

The above inequality follows from using the same logic as the one used for the proof of Lemma 1 to show that $z>\frac{1}{2}$ and then using the last inequality to show that $\frac{\partial P}{\partial \alpha}<\frac{1}{4}$ and $-\frac{\partial P}{\partial z}>\frac{3}{4}$.

This completes the proof of the proposition.

Acknowledgements We would like to thank participants at the Society for the Advancement of Economic Theory Conference, Ischia 2009, the Mid-West International Economics Meeting, Evanston, 2010, the GEP Annual Conference, University of Nottingham 2010, the CES*ifo* Area Conference on Global Economy, Munich, 2011, the Conference on Worker-Specific Effects of Globalization, Tubingen, 2011, the ETSG Conference, Copenhagen, 2011, the Conference on Globalization: Strategies and Effects, Kolding Denmark, 2011 and seminar participants at ECARES, Brussels, 2011, the University of Southern California, 2012 and the University of Tennessee, 2012 for helpful comments and suggestions. The usual disclaimer applies.

References

Acemoglu, D. (1996). A microfoundation for social increasing returns in human capital accumulation. *Quarterly Journal of Economics, 111*, 779–804.

Alós-Ferrer, C. (2002). Random matching of several infinite populations. *Annals of Operations Research, 114*, 33–38.

Bell, D., & Blanchflower, D. (2011). Underemployment in the UK in the great recession. *National Institute Economic Review 215*, R23–R33.

Bougheas, S., & Riezman, R. (2007). Trade and the distribution of human capital. *Journal of International Economics, 73*, 421–433.

Brixiova, Z., Li, W., & Yousef, T. (2009). Skill shortages and labor market outcomes in Central Europe. *Economic Systems 33*, 45–59.

Costinot, A., & Fogel, J. (2009). Matching and inequality in the world economy. *Journal of Political Economy, 118*, 747–786.

Davidson, C., & Matusz, S. (2006). Trade liberalization and compensation. *International Economic Review, 47*, 723–747.

Davidson, C., Martin, L., & Matusz, S. (1999). Trade and search generated unemployment. *Journal of International Economics, 48*, 271–299.

Davidson, C., Matusz, S., & Nelson, D. (2006). Can compensation save free trade. *Journal of International Economics, 71*, 167–186.

Davidson, C., Matusz, S., & Shevchenko, A. (2008). Globalization and firm level adjustment with imperfect labor markets. *Journal of International Economics, 75*, 295–309.

Davis, D. (1998a). Does European unemployment prop up American wages? National labor markets and global trade. *American Economic Review, 88*, 478–494.

Davis, D. (1998b). Technology, unemployment, and relative wages in a global economy. *European Economic Review, 42*, 1613–1633.

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2002). The regulation of entry. *Quarterly Journal of Economics, 117*, 1–37.

Egger, H., & Kreickemeier, U. (2012). Fairness, trade and inequality. *Journal of International Economics, 86*, 184–196.

Fan, C., Overland, J., & Spagat, M. (1999). Human capital, growth and inequality in Russia. *Journal of Comparative Economics, 27*, 618–643.

Felbermayr, G., Prat, J., & Schmerer, H.-J. (2011). Globalization and labor market outcomes: Wage bargaining, search frictions and firm heterogeneity. *Journal of Economic Theory, 146*, 39–73.

Felbermayr, G., Larch, M., & Lechthaler, W. (2013). Unemployment in an interdependent world. *American Economic Journal - Economic Policy, 5*(1), 262–301.

Grossman, G., & Maggi, G. (2000). Diversity and trade. *American Economic Review, 90*, 1255–1275.

Helpman, E., Itskhoki, O., & Redding, S. (2010). Inequality and unemployment in a global economy. *Econometrica, 78*, 1239–1283.

Hosios, A. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies, 57*, 279–298.

Kreickemeier, U., & Nelson, D. (2006). Fair wages, unemployment and technological change in a global economy. *Journal of International Economics, 70*, 451–469.

Krugman, P. (1995). Growing world trade: causes and consequences. *Brookings Papers on Economic Activity, 1*, 327–362.

Lusk, E., & Wright, H. (1982). Deriving the probability density for sums of uniform random variables. *The American Statistician, 36*, 128–130.

Maynard, D., Joseph, T., & Maynard, A. (2006). Underemployment, job attitudes, and turnover intentions. *Journal of Organizational Behaviour, 27*, 509–536.

McGuiness, S. (2006). Overeducation in the labor market. *Journal of Economic Surveys, 20*, 387–418.

McKee-Ryan, F., & Harvey, J. (2011). I have a job but. . . : A review of underemployment. *Journal of Management, 37*, 962–996.

Mortensen, D., & Wright, R. (2002). Competitive pricing and efficiency in search equilibrium. *International Economic Review, 43*, 1–20.

Ohnsorge, F., & Trefler, D. (2007). Sorting it out: International trade with heterogeneous workers. *Journal of Political Economy, 115*, 868–892.

Redding, S. (1996). The low-skill, low-quality trap: strategic complementarities between human capital and R&D. *Economic Journal, 106*, 458–470.

Sly, N. (2012). International productivity differences, trade and the distribution of factor endowments. *Review of International Economics, 20*(4), 740–757.

Snower, D. (1996). The low-skill, bad-job trap. In A. Booth, & D. Snower (Eds.), *Acquiring skills: Market failures, their symptoms, and policy responses*. New York: Cambridge University Press.

Offshoring, Mismatch, and Labor Market Outcomes

David M. Arseneau and Brendan Epstein

Abstract We study the role of labor market mismatch in the adjustment to a trade liberalization that results in the offshoring of high-tech production. Our model features two-sided heterogeneity in the labor market: high- and low-skilled workers are matched in a frictional labor market with high- and low-tech firms. Mismatch employment occurs when high-skilled workers choose to accept a less desirable job in the low-tech industry. The main result is that this type of job displacement is actually beneficial for the labor market in the country doing the offshoring. The reason is that mismatch allows this economy to reallocate domestic high-skilled labor across both high- and low-tech industries. In doing so, this reallocation dampens both the increase in the aggregate unemployment rate and the decline in aggregate wages that come as a consequence of shifting domestic production abroad. From a policy perspective, this result is perhaps counter-intuitive because it suggests that some degree of job dislocation is actually desirable as it helps facilitate adjustment in the labor market following a trade liberalization.

1 Introduction

Improvements in transportation, information, and communication technologies over the past 30 years have led to greater internationalization of production processes. A key component of this aspect of globalization has been an increase in the willingness of firms to offshore—that is, to hire foreign workers to do jobs that were previously

The views in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

D.M. Arseneau (✉) • B. Epstein
Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue N.W., Washington, D.C. 20551, USA
e-mail: david.m.arseneau@frb.gov; brendan.epstein@frb.gov

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_3

done domestically.[1] One consequence of the increase in offshoring has been a backlash in public opinion regarding whether or not this trend is desirable. Indeed, offshoring—and, in particular, its impact on the labor market—is a controversial topic. The problem is all the more acute when we consider that the effects of offshoring may run beyond actual observed offshoring. Blinder (2009) makes the point that while actual offshoring may have an important effect on wages and employment, the mere fact that certain jobs have characteristics that make them susceptible to offshoring may also be an important driver of labor market outcomes.

There is a wide body of empirical evidence on the impact of offshoring on the labor market.[2] However, it is difficult to extract a consistent theme at the aggregate level primarily because the wage and labor market effects tend to differ greatly depending on what specific notion of offshoring is being considered and on the degree of disaggregation of the data being studied. Indeed, no clear consensus emerges from these papers regarding whether or not a strong backlash against offshoring is, in fact, justified.

One relatively underdeveloped area of the literature concerns the role of trade and job dislocation, or the idea that high-skilled workers are forced into taking low-paying jobs. Earlier work by Jacobson et al. (1993) and Kletzer (2001) suggests that job dislocation has an important role in the adjustment to globalization. More recently, work by Hummels et al. (2011) uses Danish micro-level data to show that the adjustment costs for workers displaced due to offshoring tend to be higher than the adjustment costs for workers displaced for other reasons.

In this paper, we try to better understand the link between job dislocation and offshoring using a theoretical approach. Our analysis builds on Davidson et al. (2008)—henceforth DMS—which, in turn, is an open economy extension of Albrecht and Vroman (2002).

The model consists of two countries, called the North and South, each of which is endowed with a labor force composed of both high- and low-skill individuals. Low-skill individuals are only qualified for low-tech jobs, while high-skill individuals are qualified for both high- and low-tech jobs.[3] Upon entry into a market ex ante identical firms make an irreversible investment decision to adapt one of two possible technologies: low- or high-tech production. Free entry into vacancy posting is

[1]The trend toward increased offshoring and more significant foreign involvement by multinational enterprises more generally is documented in Crino (2009), for example.

[2]For example, early work by Feenstra and Hanson (1996, 1997, 1999) focuses on manufacturing offshoring, while Amiti and Wei (2005) and Gorg and Hansely (2005) focus on the rapid growth in services offshoring. These papers center primarily at the industry level. Taking yet a different approach, both Liu and Treffler (2008) and Ebenstein et al. (2011) try to measure the impact of offshoring using disaggregated data from the Current Population Survey (CPS). In still a different direction, Goos et al. (2014) examine the link between offshoring and changes in the occupational structure of employment.

[3]This definition of skill, taken from Albrecht and Vroman (2002), is also used in Burdett and Mortensen (1998), Mortensen and Pissarides (1999), Gautier (2002), Dolado et al. (2009), and Arseneau and Epstein (2014a).

such that both high- and low-tech firms can exist within any given country. In order to produce, firms match with workers in a labor market that is characterized by search frictions as in Diamond (1982), Mortensen (1982), and Mortensen and Pissarides (1994). Given the different employment opportunities available to high-skill workers, two types of equilibrium configurations are possible. In an ex-post segmented (EPS) equilibrium high-skilled workers only accept high-tech jobs. Alternatively, under certain conditions a cross-skill matching (CSM) equilibrium—that is, one in which high-skilled workers are willing to accept both high- and low-tech (*mismatch*) jobs—can arise.

We make two improvements relative to the original model used in DMS. First, a key limitation of the DMS setup is that there are no explicit international production linkages. Our paper improves on this by introducing a richer production structure. Specifically, we assume that in the North there exists a multinational firm that produces output using a flexibly parameterized constant elasticity of substitution (CES) aggregator of domestic and foreign (offshored) high-tech inputs. The second improvement is that our analysis is carried out in a version of the model that is carefully calibrated using data from both the U.S. and emerging market economies.

The paper has two main results. First, we show that mismatch employment helps mitigate adverse labor market consequences that arise as a result of offshoring. We show this by first qualitatively reproducing the results in DMS under which a trade liberalization triggers a switch from EPS to CSM equilibrium. We then isolate the effect of mismatch by forcing the economy from a post-liberalization CSM equilibrium (with mismatch) to a post-liberalization EPS equilibrium (without mismatch). The difference in allocations across the two post-liberalization outcomes reveal that by allowing domestic high-skilled labor to reallocate across both the remaining high- and low-tech industries, mismatch helps to dampen both the increase in the unemployment rate and the decline in wages that result from shifting domestic production abroad. This finding is essentially an open economy extension of Albrecht and Vroman's (2002) result on the impact of skill-biased technological change in a closed economy setting. From a policy perspective, this result suggests that some degree of mismatch is actually desirable from an aggregate perspective because it helps facilitate adjustment in the labor market following a trade liberalization.

The second result is more technical in nature. We show that under our calibration the size of the parameter space over which it is possible for mismatch to emerge as an equilibrium outcome is limited. In particular, as long as the size of the high-tech offshoring sector remains relatively small—which appears to be the most empirically relevant case—highly-skilled workers will only choose to enter into mismatch employment if domestic and foreign inputs are sufficiently complementary in the multinational's production function. If the two inputs are instead viewed as substitutes, the calibrated economy lies in a part of the parameter space where mismatch does not arise in a post-liberalization equilibrium.

In terms of related literature, beyond the empirical papers cited above, there is a large body of literature, surveyed by Helpman (2006), on offshoring within the context of standard trade theory with frictionless labor markets. Grossman

and Rossi-Hansberg (2008) is a relatively recent contribution. More recently, a growing number of studies have extended the offshoring literature to include equilibrium unemployment. Examples of these studies include Mitra and Ranjan (2010), Arseneau and Leduc (2012), and Ranjan (2013). In a slightly different direction, both Oldenski (2014) and Acemoglu et al. (2012) study the role of offshoring in the polarization of the U.S. labor force—that is, the trend decline in middle-income occupations. This paper is different from this earlier literature because—like DMS, which, in turn, builds on a long research agenda summarized in Davidson and Matusz (2009)—it focuses on the interaction between offshoring and mismatch unemployment.

The remainder of the paper is organized as follows. We present the model in Sect. 2. Section 3 discusses our calibration strategy, and the main results are presented in Sect. 4. Section 5 addresses the sensitivity of our results to alternative parameterizations of the multinational's production function. Finally, Sect. 6 offers some concluding comments.

2 The Model

The world economy consists of two countries that we refer to as the North and the South. Each economy is inhabited by both high- and low-skilled workers, and there is no international migration. In addition, in each economy there is a final-output producing firm that uses domestic low-tech intermediate products, only, as inputs. There is also a vertically integrated multinational headquartered in the North that produces final output as well, but using high-tech intermediate inputs, only. In turn, intermediate inputs are produced using labor. Each country's final output is non-tradable. Thus, trade occurs only in intermediate goods.

Following related literature, we abstract from capital and assume that each intermediate goods producer has only one job. Also, we follow DMS regarding production configurations under autarky and trade. In autarky high-tech intermediate production occurs exclusively in the North. Under trade, offshoring takes place and high-tech intermediate goods are also produced in the South and sold to the North for use as inputs in the multinational's production of final output.

Each economy's labor market is characterized by search frictions. In line with Albrecht and Vroman (2002) and DMS, search is assumed to be undirected and we abstract from on-the-job search. Undirected search simply means that per-period matches between firms and job seekers are the result of an aggregate matching technology that is increasing in aggregate vacancies and aggregate unemployment.[4]

[4]This can also be interpreted as random search, since the assumption of an aggregate meeting function is such that individuals searching for employment may encounter job opportunities that they are unwilling or unqualified to take. Alternatively, in the present context directed search would involve sectoral high- and low-tech meeting functions that take as inputs sector-specific vacancies and searchers instead of their aggregate counterparts. The existence of such sectoral

The labor market is structured such that low-skill workers in either economy are only qualified for employment in jobs operated by the corresponding domestic low-tech intermediate producers. In contrast, high-skilled workers are in principle able to attain employment in either the high-tech multinational—via high-skilled workers in the North employed by high-tech intermediate producers and high-skilled workers in the South employed by offshore (high-tech) intermediate producers—or a domestic low-tech intermediate producer.

Our assumptions regarding the skill content of jobs potentially gives rise to multiple equilibria. Borrowing terminology from Albrecht and Vroman (2002), a cross skill matching equilibrium (CSM) arises when a high-skilled worker is willing to accept a job paying a lower wage in the low-tech sector in the interest of more quickly moving out of the state of unemployment. Throughout the remainder of this paper, we refer to this type of labor market outcome in CSM equilibrium as mismatch employment. In contrast, an ex-post segmentation equilibrium (EPS) arises when high-skilled workers are unwilling to accept low-tech jobs. The conditions under which each of the noted matches can arise are discussed in Sect. 2.3 following development of the model.

Of note, CSM and EPS equilibria are not mutually exclusive: Under certain parameter configurations both equilibria can exist simultaneously. For the Southern economy, in absence of offshoring CSM equilibrium obtains trivially because low-tech jobs are the only source of employment. Beyond that special case, as in Albrecht and Vroman (2002) and DMS we abstract from formal modeling mechanisms that may lead to equilibrium selection.

2.1 The Northern Economy

The Northern economy is inhabited by a population of infinitely lived individuals with unit mass, all of whom always participate in the labor market. A fraction $\psi \in (0,1)$ of the Northern labor force consists of low-skilled workers (L). The remaining $1-\psi$ fraction of the labor force consists of high-skilled workers (H).

The price of final output is normalized to unity. Both final and intermediate producers are price takers in the product market. Final goods producers are also price takers in the inputs market. However, hiring of labor at the intermediate production stage is subject to search frictions implying bargaining power for both workers and intermediate producers. Given this bargaining power, following related search literature we assume that wages are negotiated via Nash bargaining with no commitment to the future path of wages (that is, wages can be instantaneously renegotiated given changes in the economic environment). High- and low-tech

meeting functions imply that workers are indeed able to only search for jobs that they are qualified for and are actually willing to take. See Epstein (2012) and Arseneau and Epstein (2014a) for more on this topic.

intermediate producers fill open positions through vacancy posting. Moreover, free entry into vacancy posting is such that both high- and low-tech intermediate producers coexist.

2.1.1 Production

Final output produced using low-tech (L) intermediate inputs is given by $Y_L = Z_L N_L$, where: Z_L is technology; and N_L is the total amount of low-tech intermediate inputs. As noted earlier, we assume that all product markets—both final goods and intermediate goods—are perfectly competitive and that final goods producers are price takers in the input market. Letting p_L denote the price of low-tech intermediate input, the profit-maximization problem of final goods production using low-tech intermediate inputs is:

$$\max_{N_L} Z_L N_L - p_L N_L.$$

Thus, in equilibrium

$$p_L = Z_L. \tag{1}$$

Since each intermediate producer has only one job, then $N_L = z_L \varepsilon_L + z_M e_M$, where: z_L is the productivity of a match between a low-skill worker and a low-tech intermediate producer; ε_L is the mass of low-skilled individuals working for intermediate low-tech producers; z_M is the productivity of a mismatch employment relationship; and e_M is the mass of mismatch (high-skilled) individuals working for intermediate low-tech producers (in CSM equilibrium $e_M > 0$ and in EPS equilibrium $e_M = 0$). Alternatively, $\varepsilon_L + e_M$ is also the mass of low-tech intermediate producers that are operating at any given point in time. It follows that $y_L = p_L z_L$, where y_L is the value of output for a low-tech intermediate producer matched with a low-skill individual, and $y_M = p_L z_M$ is the value of mismatch output.

The multinational produces final goods Y_H, and can do so using domestic intermediate high-tech (H) inputs and high-tech inputs produced in the South via offshoring (O). We assume the constant elasticity of substitution (CES) function:

$$Y_H = Z_H[\eta(N_H)^\sigma + (1-\eta)(N_O)^\sigma]^{1/\sigma}, \tag{2}$$

where: Z_H is technology; N_H is total amount of domestic high-tech intermediate inputs; N_O is total amount of offshore (high-tech) intermediate inputs; and the parameters η and σ are, respectively, the share of domestic intermediate high-tech inputs and the degree of substitutability between domestic and offshore intermediate

high-tech inputs.[5] In the absence of trade, $\eta = 1$. Note that the multinational production function establishes explicit international linkages between countries, which is an innovation of our modeling approach relative to DMS, where no explicit international linkages are considered.

The profit-maximization problem of the multinational is:

$$\max_{N_H, N_O} Z_H[\eta(N_H)^{\sigma} + (1-\eta)(N_O)^{\sigma}]^{1/\sigma} - p_H N_H - p_O N_O,$$

where: p_H is the price of high-tech Northern intermediate inputs; and p_O is the price of offshore inputs. The first-order conditions imply that:

$$p_H = Z_H \eta (N_H)^{\sigma-1} [\eta(N_H)^{\sigma} + (1-\eta)(N_O)^{\sigma}]^{1/\sigma-1}; \tag{3}$$

and:

$$p_O = Z_H (1-\eta)(N_O)^{\sigma-1} [\eta(N_H)^{\sigma} + (1-\eta)(N_O)^{\sigma}]^{1/\sigma-1}. \tag{4}$$

$N_H = z_H \varepsilon_H$ and $N_O = z_O \varepsilon_O$, with analogous interpretation as earlier. In particular, ε_H is the mass of Northern high-skilled workers employed by Northern high-tech intermediate producers and ε_O is the mass of Southern high-skilled individuals employed by offshore intermediate producers. It follows that for $i \in \{H, O\}$ $y_i = p_i z_i$, where y_i is the value of output for a type-i intermediate producer.

2.1.2 Matching Process

Let m denote matches per period. In line with related literature, m is assumed to be increasing and concave in aggregate vacancies, v, and aggregate unemployment, u. Aggregate vacancies are equal to the sum of low- and high-tech vacancies posted by intermediate producers, respectively v_L and v_H. Similarly, aggregate unemployment is equal to the sum of the masses of low- and high-skill unemployed individuals, respectively u_L and u_H. Note that because the population consists of a unit mass and all individuals participate in the labor force, then u is also the aggregate unemployment rate.

$\gamma_L \in (0, 1)$ is the fraction of unemployed low-skilled individuals relative to total unemployment, that is, $\gamma_L = u_L/u$. In addition, $q = m/v$ is meetings per vacancy, that is, the rate at which intermediate producers with open positions meet workers. It follows that $q\gamma_H$, where $\gamma_H = 1 - \gamma_L$, is the rate at which any intermediate producer with a vacancy meets a low-skill unemployed worker, and $q\gamma_H$ is the rate at which any intermediate producer with a vacancy meets a high-skill unemployed worker. Note that all meetings between workers and firms do not result in a match. Indeed,

[5] $\sigma = 1$ implies perfect substitutability, $\sigma = -\infty$ implies perfect complementarity, and $\sigma = 0$ is the Cobb-Douglas production-function case.

when a low-skill worker and a high-tech intermediate producer meet, no match is formed, which is also the case when a high-skill worker and a low-tech intermediate producer meet in EPS equilibrium.

Let ϕ_L denote the fraction of low-tech vacancies relative to aggregate vacancies, that is, $\phi = v_L/v$. In addition, $f = m/u$ is the rate at which unemployed individuals of any type meet intermediate producers with vacancies of any given type. Thus, an unemployed individual meets a firm with a low-tech vacancy at effective rate $f\phi_L$ and meets a firm with a high-tech vacancy at effective rate $f\phi_H$, where $\phi_H = 1-\phi_L$.

2.1.3 Value Functions

Intermediate Producers For $i, j \in \{L,H\}$ and $i = j$ the value to a Northern intermediate producer of type i from employing a type-j worker, J_i, satisfies:

$$rJ_i = y_i - c - w_i - \delta(J_i - V_j),$$

where: r is the interest rate in the Northern economy; c is the Northern fixed flow cost of posting a vacancy and maintaining a job; y_j is the value of output of a match as defined earlier; w_i is the Nash bargained wage; δ is the exogenous economy-wide rate of job destruction in the North; and V_i is the value of an unfilled vacancy posted by a Northern type-i intermediate producer. The value of mismatch employment for a low-tech firm, J_M, is the same as the expression above but with $i = M$ and $j = L$. Intuitively, the firm's value from a filled position is equal to a worker's marginal value of production net of the wage, the job-maintenance cost, and the expected capital loss owing to job destruction.

Given the meeting process described above, for $i \in \{L, H\}$ the value of a vacancy for a type-i firm, V_i, satisfies:

$$rV_i = -c + q\gamma_i(J_i - V_i)\mathbb{I}_L q\gamma_H(J_M - V_L),$$

where: $\mathbb{I}_L$ is an indicator function whose value is equal to 1 with $i = L$ in CSM equilibrium, while it takes a value of 0 in the event that $i = H$ or in EPS equilibrium if $i = L$.

Workers For $i, j \in \{L,H\}$ and $i = j$ the value to a Northern worker of type i from being employed by a type-j firm, W_i, satisfies:

$$rW_j = w_j - \delta(W_j - U_i).$$

This equation says that the corresponding employment flow value to a worker is equal to the wage paid in the position net of the expected capital loss owing to job destruction. The value of mismatch employment for a high-skill worker W_M is the same as the expression above but with $j = M$ and $i = H$.

For $i \in \{L,H\}$ the value of unemployment for a Northern type-i worker, U_i, satisfies:

$$rU_i = b + f\phi_i(W_i - U_i) + \mathbb{I}_H f\phi_L(W_M - U_H),$$

where: $\mathbb{I}_H$ is an indicator function whose value is equal to 1 with $i = H$ in CSM equilibrium, while it takes a value of 0 if $i = L$ or in EPS equilibrium if $i = H$. It follows that unemployment flow values are equal to the sum of unemployment flow benefits and the expected capital gains from becoming employed.

2.1.4 Surpluses

The surplus of a match is equal to the sum of the capital gains that firms and workers obtain from forming an employment relationship. Thus, for $i, j, k \in \{L, H\}$ and $i = j = k$ match surpluses from high-skill workers matched with high-tech jobs and low-skill workers matched with low-tech jobs, S_j, are given by:

$$S_j = J_j - V_i + W_j - U_k. \tag{5}$$

With $j = M$, $i = L$ and $k = H$ the preceding equation also defines the surplus from mismatch. In all cases, if the value of a particular job is positive for a firm, then free entry into vacancy posting implies $V_i = 0$, which is what allows for the existence of both high- and low-tech intermediate goods producers within any given country.

2.1.5 Wages

Let $\beta \in (0,1)$ denote workers' exogenously determined constant bargaining power. Nash bargaining of wages without commitment to the future path of wages (that is, wages can be instantaneously renegotiated given changes in the economic environment) implies a surplus sharing rule by which a worker's capital gain from becoming employed is a constant fraction β of the associated surplus, and similarly a firm's capital gain derived from forming an employment relationship is a constant fraction $1-\beta$ of the associated surplus. Thus, Nash bargaining implies that for $i,j,k \in \{L,H\}$ and $i = j = k$ the wage w_i is implicitly defined by:

$$(W_j - U_k) = \frac{\beta}{1-\beta}(J_j - V_i).$$

In turn, the mismatch wage satisfies the equation above with $j = M$, $k = H$, and $i = L$.

Also, note that the average wage of low-skill workers, $\bar{w}_L$, is simply equal to w_L, while the average wage of high-skill workers, $\bar{w}_H$, is:

$$\bar{w}_H = \frac{\varepsilon_H w_H + e_M w_M}{1-\psi-u_H},$$

Finally, the Northern aggregate wage, $\mathbb{W}$, is given by:

$$\mathbb{W} = \frac{\varepsilon_L w_L + \varepsilon_H w_H + e_M w_M}{1-u}.$$

2.1.6 Flow Equilibrium Conditions

Three equilibrium flow conditions must be accounted for, which define: (1) the mass of unemployed low-skilled individuals; (2) the mass of unemployed high-skill individuals; and (3) the mass of mismatched individuals. Since all employment relationships end at exogenous rate δ and low-skill individuals exit unemployment at effective rate $f\phi_L$, it follows that in equilibrium the mass of low-skill unemployed individuals satisfies:

$$\delta(\psi-\gamma_L u) = f\phi_L\gamma_L u. \tag{6}$$

Using similar reasoning, the equilibrium mass of unemployed high-skilled individuals satisfies:

$$\delta(1-\psi-\gamma_H u) = f(\phi_H + \mathbb{I}_H\phi_L)\gamma_H u. \tag{7}$$

Furthermore, high-skill workers become mismatch employed at effective rate $\mathbb{I}_H f\phi_L$. therefore, the equilibrium mass of high-skilled workers employed by low-tech firms satisfies:

$$\mathbb{I}_H f\phi_L\gamma_H u = \delta e_M. \tag{8}$$

Of course, the total number of individuals employed by high-tech firms is:

$$\varepsilon_H = 1-\psi-\gamma_H u - e_M,$$

and the total number of low-skilled individuals employed by low-tech firms is:

$$\varepsilon_L = \psi - \gamma_L u.$$

Note that the unemployment rate of low-skill workers is u_L/ψ, and, similarly, the unemployment rate of high-skill workers is $u_H/(1-\psi)$. Moreover, the rate of mismatch employment to total high-skill employment is $e_M/(1-\psi-u_H)$.

2.2 The Southern Economy

Recall that, following DMS, we assume that while the Southern labor force consists of both high- and low-skill individuals, under autarky in the South only low-tech intermediate producers exist whose input is used in the final production of Southern goods. In turn, offshoring of intermediate high-tech production for use as input in the multinational's production of final output in the North introduces a high-tech (offshoring) sector in the South.

Akin to the Northern economy, in the South low-skill workers are only productive in low-tech intermediate firms, while high-skill workers can produce in both low- and high-tech (offshore) intermediate firms. Under autarky the Southern economy is trivially in CSM equilibrium and the Southern ratio of mismatch to total high-skill employment is equal to 1 (given no other employment alternative Southern high-skill workers find it optimal to take low-tech jobs).

For expositional brevity we omit full development of the Southern economy. All notation and assumptions about the Southern economy are analogous to those noted earlier in the development of the Northern economy under the appropriate configurations except for the following. First, variables and parameters corresponding explicitly to the South are distinguished by use of an asterisk. Second, high-tech offshore variables and parameters are denoted with an O (instead of an H). Third, in an equilibrium with trade the value of a job to an offshore intermediate high-tech producer—that is, the value to a high-tech intermediate firm located in the South that employs a Southern high-tech worker and sells its production to the Northern multinational— J_O^* satisfies:

$$r^* J_O^* = y_O - c^* - c_O^* - w_O^* - \delta^*(J_O^* - V_O^*),$$

where: c_O^* is a fixed flow cost specifically associated with maintaining an offshore job and posting offshore vacancies. It follows that the value of an offshore vacancy V_O^* satisfies:

$$r^* V_O^* = -c^* - c_O^* + q^* \gamma_H^* (J_O^* - V_O^*)$$

(note that in an equilibrium with trade total vacancies in the South v^* are equal to the sum of Southern low-tech vacancies v_L^* and offshore vacancies v_O^*; in addition, the fraction of low-tech vacancies is ϕ_L^*, and the fraction of offshore vacancies is $\phi_O^* = 1 - \phi_L^*$). Fourth, in similar spirit to DMS, we assume that autarky is consistent with the value of c_O^* being prohibitively high. However, a decline in c_O^* to a sufficiently small value is consistent with the world economy moving to an equilibrium with trade (for simplicity, we assume that trade is associated with the value of c_O^* having effectively dropped to zero, therefore inducing entry into the offshore market).

2.3 Equilibrium

Following DMS, our baseline world equilibrium assumptions are that: (1) under autarky the North is in EPS equilibrium (while the South is trivially in CSM equilibrium); (2) under trade high-skilled workers in the North become sufficiently pessimistic about their future job market prospects triggering a switch to CSM equilibrium; (3) for simplicity, under trade the wages associated with offshore employment are perceived by Southern high-skill workers as being sufficiently high that they stop accepting low-tech jobs, triggering a switch to EPS equilibrium. Unlike DMS, we later assess the impact of mismatch in the North by considering an alternative trade equilibrium in which both the North and the South are in EPS equilibrium. Moreover, we discuss why in autarky a CSM equilibrium in the North would not emerge under our baseline calibration.

For expositional brevity, in the remainder of this section whenever possible we focus the discussion on the Northern economy. Analogous considerations to those presented, when applicable, are straightforward for the South. Also, recall that in line with related literature, we assume free entry into vacancy posting. This assumption implies that the value of a vacancy for any intermediate producer is equal to zero.

2.3.1 Preliminaries

Key Equilibrium Variables Recall from earlier that the matching function m was assumed to be increasing and concave in aggregate vacancies and aggregate unemployment. Following related literature we assume that $m = Av^{\alpha}u^{1-\alpha}$, where: A is matching efficiency; and $\alpha \in (0,1)$ is the elasticity of matches with respect to aggregate vacancies. Let $\theta = v/u$ denote market tightness. It follows that the rate at which workers meet firms can be stated as $f = A\theta^{\alpha}$, which is increasing in market tightness, and the rate at which firms meet workers is $q = A\theta^{\alpha-1}$, which is decreasing in market tightness. It follows that key equilibrium conditions can be stated in terms of market tightness (rather than aggregate vacancies and aggregate unemployment) and also, given the surplus definitions and the Nash bargaining surplus-sharing rules, in terms of surpluses (rather than agent-specific value functions).

Existence For EPS equilibrium to exist, it must be the case that the surplus from a match between a low-skilled worker and a low-tech intermediate firm is positive and that the surplus from a match between a high-skill worker and a high-tech intermediate firm is positive as well. Solving explicitly for surpluses in EPS equilibrium (which is straightforward given the definition of surplus from earlier as well as relevant value function statements) it follows that $y_L > c + b$ and $y_H > c + b$ are necessary and sufficient conditions for, respectively, $S_L > 0$ and $S_H > 0$ to hold.

For CSM equilibrium to be viable it must be the case that: $S_L > 0$; $S_H \geq 0$; and the surplus from a match between a high-skilled worker and a low-tech intermediate

firm (mismatch) is positive: $S_M > 0$ (again, solving for S_H, S_M, and S_L in CSM equilibrium is straightforward given their definitions and relevant value function statements; in the event that $S_H = 0$ high-tech firms do not post vacancies meaning that with all individuals participating in the labor force $S_L > 0$ and $S_M > 0$ are such that CSM equilibrium emerges trivially). Suppose this is the case. Then, subtracting the surplus S_H from the surplus S_M and rearranging implies that:

$$S_H - S_M = \frac{y_H - y_M}{r + \delta}. \tag{9}$$

In words: the difference between the surplus from a match between a high-skilled worker and a high-tech firm and the surplus from mismatch is simply equal to a weighted difference of the value of output between these matches. Now, rearranging the surplus S_M and substituting in Eq. (9) yields:

$$(r + \delta + f\beta) S_M = y_M - (c + b) - \frac{f\phi_H \beta}{r + \delta}(y_H - y_M).$$

Therefore, S_M is positive if and only if:

$$y_M - (c + b) > \frac{f\phi_H \beta}{r + \delta}(y_H - y_M). \tag{10}$$

It follows from Eq. (9) that if $y_H > y_M$ it is necessarily the case that $S_H > S_M$ (and $S_H > 0$ if $S_M > 0$). Furthermore, inequality (10) is satisfied (and hence $S_M > 0$) if the difference between the output value of mismatch and the sum of unemployment flow benefits and the cost of of maintaining a job is larger than a weighted difference between the value of high-tech and mismatch intermediate output.

Note that, all else equal, the smaller the difference between y_M and y_H is, as well as the smaller ϕ_H is, the more likely it is that inequality (10) is indeed satisfied. Intuitively, as $\phi_H \to 0$ high-tech jobs become increasingly scarce, and in the limit they are entirely unavailable meaning that high-skill workers' outside (high-tech) employment options do not weigh on mismatch surplus. An analogous interpretation follows from considering the case in which $y_M \to y_H$.

As noted earlier, EPS and CSM equilibrium configurations are not mutually exclusive. Indeed, as long as $y_L > c + b$ and $y_H > c + b$, then EPS equilibrium always exists. Should the preceding hold and the economy's parameters be such that CSM equilibrium also exists, then as noted in DMS whether the economy settles on EPS or CSM equilibrium can depend on many factors, including, for instance, high-skill workers' expectations about high-tech employment prospects.

Vacancy Creation Conditions Rearrangement of the Northern vacancy flow value equations along with free-entry into vacancy-posting and the definition of surplus and Nash bargaining imply for $i \in \{L,H\}$ the following Northern vacancy-creation

conditions:

$$(1-\beta)(\gamma_i S_i + \mathbb{I}_L \gamma_H S_M) = c/q \tag{11}$$

These equations hold when the value of a particular type of job to a firm is positive, and simply say that given free-entry into vacancy creation the expected gains from posting low- or high-tech vacancies (the left-hand side) are equal to the expected costs (the right-hand side). In the South, similar reasoning applies leading for $i \in \{L,O\}$ to the following Southern vacancy-creation condition:

$$(1-\beta^*)(\gamma_i^* S_i^* + (1-\mathbb{I}_O^*)\gamma_H^* S_M^*) = c^*/q^*, \tag{12}$$

where: $\mathbb{I}_O^*$ is an indicator function that equals 0 if $i = L$ and the Southern economy is in autarky and equals 1 if offshoring occurs and $i = O$ or $i = L$.

2.3.2 Definitions

Given the development thus far, in all definitions that follow knowledge of the fundamental vectors $\mathbb{S}$ and $\mathbb{V}$ is sufficient to solve for any other endogenous variables and equilibrium flow equations of interest.

Definition 1 (Autarky Equilibrium with North in EPS Configuration and South in CSM Configuration) A steady-state autarky equilibrium with ex-post segmentation in the North and (trivial) cross-skill matching in the South consists of a set of surplus values $\mathbb{S} = \{S_L, S_H, S_L^*, S_M^*\}$ and a vector $\mathbb{V} = \{\theta, \phi_L, \gamma_L, u, \theta^*, \gamma_L^*, u^*, e_M^*, p_H, p_L, p_L^*\}$ such that:

1. The following are solved: the two corresponding Northern surplus equations implied by Eq. (5) and the two analogous surplus equations for the South, the two corresponding Northern vacancy-creation condition equations implied by Eq. (11), the single corresponding vacancy-creation condition for the South implied by Eq. (12), the two Northern price equations (1) and (3) with $\eta = 1$ and the single analogous (low-tech) price equation for the South, Eqs. (6), (7), and the two analogous unemployment equations for the South, and also the single Southern equation defining mismatch employment analogous to its Northern counterpart (Eq. (8)).
2. $S_L > 0, S_H > 0, S_L^* > 0, S_M^* > 0$.

Definition 2 (Trade Equilibrium with Both the North and South in EPS Configurations) A steady-state trade equilibrium with ex-post segmentation in both the North and South consists of a set of surplus values $\mathbb{S} = \{S_L, S_H, S_L^*, S_O^*\}$ and a vector $\mathbb{V} = \{\theta, \phi_L, \gamma_L, u, \theta^*, \phi_L^*, \gamma_L^*, u^*, p_H, p_L, p_L^*, p_O\}$ such that:

1. The following are solved: the two corresponding Northern surplus equations implied by Eq. (5) and the two analogous surplus equations for the South, the two corresponding Northern vacancy-creation condition equations implied

by Eq. (11), the two corresponding vacancy-creation conditions for the South implied by Eq. (12), the three Northern price equations (1), (3), (4) (with $\eta = 0$) and the single analogous (low-tech) price equation for the South, Eqs. (6), (7), and the two analogous unemployment equations for the South.
2. $S_L >, S_H > 0, S_L^* > 0, S_O^* > 0$.

Definition 3 (Trade Equilibrium with the North in CSM Configuration and the South in EPS Configuration) A steady-state trade equilibrium with cross-skill matching in the North and ex-post segmentation in the South consists of a set of surplus values $\mathbb{S} = \{S_L, S_H, S_M, S_L^*, S_O^*\}$ and a vector $\mathbb{V} = \{\theta, \phi_L, \gamma_L, u, e_M, \theta^*, \phi_L^*, \gamma_L^*, u^*, p_H, p_L, p_L^*, p_O\}$ such that:

1. The following are solved: the three corresponding Northern surplus equations implied by Eq. (5) and the two analogous surplus equations for the South, the two corresponding Northern vacancy-creation condition equations implied by Eq. (11), the two corresponding vacancy-creation conditions for the South implied by Eq. (12), the three Northern price equations (1), (3), (4) (with $\eta = 0$) and the single analogous (low-tech) price equation for the South, Eqs. (6), (7), (8), and the two analogous unemployment equations for the South.
2. $S_L > 0, S_H \geq 0, S_M > 0, S_L^* > 0, S_O^* > 0$.

In light of the preceding definitions, the definition of an autarky equilibrium with both the North and South in CSM configurations is straightforward.

3 Calibration

The baseline calibration is summarized in Table 1 and assumes a trade equilibrium with the North in CSM configuration and the South in EPS configuration (this is consistent with equilibrium Definition 3 above). The frequency is monthly, so the interest rates are set to $r = r^* = 0.004$, which is consistent with an annual interest rate of 5 %.

3.1 Northern Economy

As in Arseneau and Epstein (2014a), we assume that individuals with at least some post-secondary education are high-skill workers while those with at most a high school degree correspond to low-skill workers. Bureau of Labor Statistics (BLS) data shows that about half of the U.S. population has at most a high school degree. Therefore, we set the fraction of low-skilled individuals in the Northern economy to $\psi = 0.5$.

As noted earlier, the job market is characterized by a standard Cobb-Douglas matching function. We set the elasticity of the matching function with respect to

Table 1 Baseline calibration at monthly frequency

North		South
Production parameters		
2.45	Multinational technology parameter: Z_H	–
0.99	Multinational share of domestic input: η	–
−41.75	Multinational substitutability parameter: σ	–
1.60	Low-tech final goods technology: Z_L, Z_L^*	0.60
1.00	High-tech intermediate production technology: z_L, z_L^*	1.00
1.00	Low-tech intermediate production technology: z_L, z_L^*	1.00
1.05	Mismatch intermediate production technology: z_M	–
Labor-market, population parameters, and other		
0.004	Interest rate r, r^*	0.004
0.50	Share of low skill individuals: ψ, ψ^*	0.80
0.50	Worker bargaining power: β, β^*	0.50
0.36	Flow unemployment benefits: b, b^*	0.03
0.73	Vacancy and job maintenance flow cost: c, c^*	0.47
0.03	Job destruction rate: δ, δ^*	0.01
0.50	Elasticity of matches with respect to vacancies: α, α^*	0.50
0.87	Matching efficiency: A, A^*	0.44

aggregate vacancies α to 0.5, which is broadly in line with research surveyed in Petrongolo and Pissarides (2001). The matching efficiency parameter A is calibrated as follows. Using aggregate monthly data on unemployment since 1951 that is available publicly from the BLS along with the methodology in Elsby et al. (2009) and Shimer (2012), we find that in the U.S. the probability that an average unemployed individual finds a job within a month is 0.43. Given this figure, assuming an underlying Poisson distribution implies that the monthly rate at which an average unemployed individual finds a job is 0.56. Therefore, the matching efficiency parameter is used to attain the steady state value of $\frac{fu_H + f\phi_L u_L}{u} = 0.56$.

The exogenous job destruction probability δ is calibrated using data on aggregate unemployment from the BLS. These data show that since 1951 the average US unemployment rate is 0.058. Therefore, we pin down the job destruction rate by targeting the steady-state value $u = 0.058$. In turn, the cost of posting vacancies c is used to target the ratio of aggregate vacancies to aggregate unemployment: $v/u = 0.68$. This figure is obtained using data on aggregate job openings from the BLS Job Openings and Labor Turnover Survey since 2000 (when first available) along with the Conference Board's Help-Wanted Index from 1951 through 2000. Combined with the BLS time series for aggregate unemployment, these data imply that in the U.S. the average ratio of aggregate vacancies to aggregate unemployment is 0.68 in the post-war period.

We assume symmetry in bargaining power, so that $\beta = 0.5$. This parameterization has the virtue that, in our model, $\beta = \alpha$ delivers an efficient split of match surplus (see Hosios 1990). Broadly in line with Shimer (2005), unemployment benefits b

are set to deliver a 40 % replacement rate of aggregate wages. In particular, we set $b = 0.4\frac{w_H\varepsilon_H + w_M e_M + w_L\varepsilon_L}{1-u}$.

For production, we normalize the intermediate input technology parameters so that $z_H = z_L = z_O = 1$. In addition, we follow DMS and assume that the technology parameter in the final low-tech goods aggregator is $Z_L = 1.6$ and that the value of intermediate high-tech production in the North is 12.5 % higher than its low-tech counterpart. This last assumption pins down the technology parameter of the final-high tech goods aggregator Z_H and results in a wage premium of correctly match high-skill workers relative to correctly matched low-skill workers of nearly 20 %. In addition, we set the share of domestic intermediate inputs in the multinational's production function to $\eta = 0.9$, broadly in line with Burstein et al. (2008), though lower than used in Arseneau and Leduc (2012). We discuss the choice of the multinational's substitutability parameter σ in the following section, as it is relevant for wage differentials in the South.

Finally, the analysis in Sicherman (1991) suggests that each year of overeducation relative to a job's educational requirements induces an overeducation wage premium of nearly 4 % relative to individuals with the required amount of education. Given that our reference for the model's high-skill workers are individuals with at least some post-secondary education—which can include attaining a college degree and higher—we use the mismatch productivity parameter z_M to target a conservative wage ratio of $w_M/w_L = 1.135$.

3.2 Southern Economy

Our Southern country is meant to be a representative emerging market economy. In particular, we have in mind countries classified as emerging economies by the International Monetary Fund, including Argentina, Brazil, Bulgaria, Chile, China, Colombia, Hungary, India, Indonesia, Latvia, Lithuania, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Romania, Russia, South Africa, Thailand, Turkey, Ukraine, and Venezuela. Publicly available data from the World Bank shows that across our list of emerging economies the average fraction of the population that completed at most a secondary education is about 0.8. Therefore, we set the fraction of low-skilled individuals in the Southern economy to $\psi^* = 0.8$.

As in the North, the Southern job market is characterized by a standard Cobb-Douglas matching function. In line with Petrongolo and Pissarides (2001) we set $\alpha^* = 0.5$. Data on worker flows is quite limited for emerging economies. Therefore, we calibrate the matching efficiency parameter A^* following Bosch and Maloney (2007), who calculate transition rates for Mexico at quarterly frequency. We assume an underlying Poisson distribution and use Bosch and Maloney's findings to back out a representative implied monthly job finding rate for Mexico of 0.11. Thus, akin to the Northern economy we choose A^* so that $\frac{f^*\phi_L^* u_L^* + f^*\phi_O^* u_H^*}{u^*} = 0.11$.

The exogenous job destruction probability δ^* is calibrated using data on aggregate unemployment rates from statistical agencies of nearly all our reference emerging economies. Taken at face value, these data suggest an average unemployment rate across emerging economies of about 10 %. Therefore, we pin down the job destruction rate by targeting the steady-state value $u^* = 0.1$. In turn, the cost of posting vacancies c^* is used to target the ratio of aggregate vacancies to aggregate unemployment: $v^*/u^* = 0.2$. We obtain this figure by taking at face value data on aggregate job openings from several emerging market statistical agencies.[6] In addition, recall that under trade we assume that the offshore-specific flow cost is $c_O^* = 0$.

As in the Northern economy, we assume symmetry in bargaining power so that $\beta^* = 0.5$. Many emerging market economies have no formal unemployment insurance scheme, so we take the midpoint between a lower bound of zero and the replacement rate in the North as a reference point to set: $b^* = 0.2\frac{w_M^* e_M^* + w_L^* \varepsilon_L^*}{1-u^*}$.

Turning towards production, we normalize $z_L^* = 1$. In addition, we follow DMS and assume that the value of Southern low-tech intermediate production is about 40 % of that of Northern high-tech intermediate production, which pins downs the technology parameter Z_L^*.

Finally, we pin down the multinational's substitutability parameter σ by assuming a wage ratio $w_O^*/w_L^* = 1.3$. This wage ratio is higher than its counterpart in the North in the spirit of the assumption that Southern high-tech workers find offshore wages sufficiently high that EPS configuration emerges under trade. As noted in Table 1, the resulting value of σ implies strong complementarity (the baseline calibration is such that for any target $w_O^*/w_L^* > 1$ a negative value for σ always emerges).

4 Quantitative Results

As noted earlier, our baseline exercise involves assessing the impact of a trade liberalization by comparing the benchmark trade equilibrium to an autarky equilibrium in which the North is in EPS configuration and the South is trivially in CSM configuration. When implementing the autarky, we assume that all country-specific parameters remain as in the benchmark trade equilibrium. In addition, when the South is in a pre-liberalization CSM equilibrium we assume a mismatch wage premium of 13.5 % (as was the case for the North in CSM equilibrium). This assumption pins down a value for the Southern mismatch productivity parameter $z_M^* = 1.03$.

We choose to compare the benchmark trade equilibrium to an autarky EPS equilibrium in the North for two reasons. First, doing so allows our results to be more directly comparable to those reported previously in DMS, who, as noted earlier,

[6] In particular, Bulgaria, Lithuania, Malaysia, Philippines, Poland, and Chile.

analyze a trade liberalization under the assumption that high-skilled workers in the Northern economy become sufficiently pessimistic about their future job market prospects such that it induces a shift from EPS to CSM equilibrium. Second—and most importantly—an autarky equilibrium in which the North is in CSM does not exist under our baseline parameterization. This is because our calibration, which is informed by observable data on both U.S. and emerging economy labor markets, implies that under autarky the value of Northern intermediate high-tech production is so much higher than the value of Northern intermediate low-tech production that the mismatch surplus is driven below zero. Hence, mismatch is not supported in the Northern autarky equilibrium under the baseline calibration.

4.1 Main Results

The baseline results are presented in Table 2. The switch from autarky (Case A) to offshoring under the assumption that the North moves to a post-liberalization CSM equilibrium (Case B) results in a decline in both the aggregate Northern unemployment rate, u, and aggregate Northern wage, $\mathbb{W}$. Thus, from the perspective of the Northern economy, the increase in offshoring carries both a benefit and a cost from the perspective of the aggregate labor market—the trade liberalization increases the number of people working, but the aggregate wage declines.

Table 2 Baseline comparisons of autarky and trade

	World economic environment	A. Autarky Benchmark		B. Trade Benchmark		C. Trade Alternative	
	Country	North	South	North	South	North	South
	Equilibrium configuration	EPS	CSM	CSM	EPS	EPS	EPS
1	Price of intermediate high-tech (p_H, p_O)	2.445	–	1.800	0.652	1.119	1.362
2	Price of intermediate low-tech (p_L, p_L^*)	1.600	0.600	1.600	0.600	1.600	0.600
3	Aggregate vacancies (v, v^*)	0.077	0.012	0.039	0.020	0.036	0.036
4	Aggregate unemployment rate (u, u^*)	0.079	0.058	0.058	0.100	0.364	0.095
5	High-tech vacancies (v_H, v_O^*)	0.049	0.000	0.014	0.007	0.003	0.022
6	Low-tech vacancies (v_L, v_L^*)	0.028	0.012	0.025	0.013	0.033	0.014
7	High-skill unemp. rate ($u_H/(1-\psi)$, $u_H^*/(1-\psi^*)$)	0.059	0.058	0.046	0.150	0.606	0.070
8	High-skill mismatch rate ($e_M/(1-\psi-u_H)$, $e_M^*/(1-\psi^*-u_H^*)$)	0.000	1.000	0.634	0.000	0.000	0.000
9	Low-skill unemp. rate (u_L/ψ, u_L^*/ψ^*)	0.099	0.058	0.070	0.087	0.121	0.101
10	Wage: high-skill in high-tech (w_H, w_H^*)	1.636	–	1.007	0.161	0.381	0.825
11	Wage: mismatch (w_M, w_M^*)	–	0.144	0.950	–	–	–
12	Average wage of high-skill ($\bar{w}_H, \bar{w}_H^*$)	1.636	0.144	0.971	0.161	0.381	0.825
13	Wage: low-skill in low-tech (w_L, w_L^*)	0.824	0.127	0.837	0.123	0.814	0.122
14	Aggregate wage ($\mathbb{W}$, $\mathbb{W}^*$)	1.238	0.130	0.905	0.130	0.680	0.266

Disaggregating to the sectoral level, an immediate consequence of offshoring is that the multinational shifts high-tech production away from the Northern economy toward the Southern economy. As a result, high-tech vacancies fall in the North. The resulting reduction in demand for Northern high-skilled labor lowers wages in the high-tech sector. Faced with declining wages and fewer job openings, high-skilled individuals become willing to accept lower quality jobs in the low-tech sector, inducing a switch to CSM equilibrium. Indeed, note that the Northern mismatch rate, $e_M/(1-\psi-u_H)$, rises from 0 to 0.63. In spite of lower vacancies in the high-tech sector, new job opportunities that arise from the willingness of high-skilled individuals to accept employment in the low-tech industry help to push down the high-skilled unemployment rate, $u_H/(1-\psi)$ by nearly 1.5 percentage points.

From the perspective of job creation in the low-tech sector, mismatch generates two opposing effects. On the one hand, because Northern low-tech firms now expect to form matches with high-skill workers—whose wages are higher than those of low-skill workers because of their higher productivity and higher outside option (employment in high-tech firms)—the expected gains from posting a low-tech vacancy are reduced. However, this effect is partially offset by the fact that low-tech firms encounter suitable job candidates faster as a result of high-skilled participation in the market for low-tech jobs. All told, low-tech vacancies decline, but by less than high-tech vacancies, making the share of low-tech vacancies in the total vacancy pool, ϕ_L, rise. Since low-skilled workers find jobs at effective rate $\phi_L f$, the unemployment rate for low-skilled workers declines nearly twice as much as the unemployment rate for high-skilled workers. Moreover, given their higher job-finding prospects, wages for low-skilled workers increase modestly as firms must compensate them for their higher outside option.

All told, these results suggest that a trade liberalization that results in increased offshoring of high-skilled labor generates clear distributional effects in the Northern economy. It flattens the wage distribution as wages in the high-tech sector fall sharply while wages in the low-tech sector increase modestly.[7] Unemployment is lower across the board, but the impact is greater for low-skilled individuals.

In the Southern economy, the shift from autarky to trade results in an increase in both the aggregate wage and aggregate unemployment. Since higher-skilled workers are now only willing to accept offshore (high-tech) jobs, low-tech firms in the Southern economy encounter suitable workers less frequently, which raises expected vacancy-posting costs. Consequently, low-tech vacancies v_L^* decline resulting in a substantial increase in the low-skill unemployment rate, u_L^*/ψ^* and a decline in the Southern low-tech wage w_L^*. In turn, the combination of Southern higher-skilled workers only being willing to accept offshore jobs in conjunction with the limited offshore vacancies posted by the multinational is such that the unemployment rate

[7] In related work, Arseneau and Epstein (2014b) introduces on-the-job search in the present framework to better understand the response of the wage distribution to an increase in offshoring. This approach is closely linked to Burdett and Mortensen (1998), Mortensen (2005), and Dolado et al. (2009) and is supported empirically by Christensen et al. (2005).

of higher-skilled Southern workers $u_H^*/(1-\psi^*)$ rises. However, since the value of offshore production is substantially higher than that of low-tech firms, the average wage of high-skilled workers under trade $\bar{w}_H^*$ is much higher than their autarky wage.

4.2 The Role of Mismatch

We conduct a simple experiment that allows us to parse out the direct role of mismatch in the adjustment of the Northern labor market to increased offshoring: Panel C of Table 2 shows results from the model assuming that the North stays in EPS equilibrium post-liberalization as opposed to switching to CSM equilibrium as in Panel B. Considering this third case allows us to shut down the equilibrium shift that is central to the results presented in the previous subsection as well as the results presented in DMS.

Comparing Panel C to Panel B makes it clear that if the Northern economy remains in EPS equilibrium, wages are significantly lower and unemployment is markedly higher following the trade liberalization. We conclude that the role of mismatch is to substantially mitigate the impact of offshoring on the labor market.

Intuitively, the benefit of mismatch comes from the fact that it allows for sectoral reallocation of the labor market following a large shock. In other words, mismatch opens up a number of additional employment opportunities for high-skilled individuals that would otherwise be absent following the increase in offshoring. As high-skilled individuals take advantage of these opportunities, labor market activity shifts across sectors in a way that cushions the decline in wages and ultimately decreases unemployment. In absence of the ability to engage in this sectoral reallocation, the Northern labor market is fully exposed to the offshoring shock because the low- and high-tech labor markets are effectively segmented.

The role of mismatch in adjusting to trade liberalization in our model echoes earlier results by Albrecht and Vroman (2002), who examine the response to skill biased technological change in a closed economy setting. This may not be surprising as an offshoring shock can be thought of as a form of skill biased technological change in the sense that the reduction in trade costs changes the production frontier of the multinational by making offshoring possible.

5 Sensitivity Analysis

We conduct two sets of sensitivity analyses. The first focuses on how the characteristics of the multinational's production function shape the response of the labor market to an increase in offshoring. The second set focuses on the size of the parameter space that supports mismatch as an equilibrium outcome.

5.1 The Multinational's Production Function

By introducing a final goods producing firm that uses a flexibly parameterized CES aggregator, our model allows international linkages to play a wide role in shaping results. We examine the sensitivity of labor market outcomes to the size of the offshore sector and to the complementarity versus substitutability of Northern and Southern inputs in the multinational's production function.

5.1.1 Size of Offshore Sector (η)

Figure 1 shows labor market outcomes following a trade liberalization that results in a switch from an autarky EPS equilibrium to a trade CSM equilibrium (upper and lower panels on the left) and from an autarky EPS equilibrium to a trade EPS

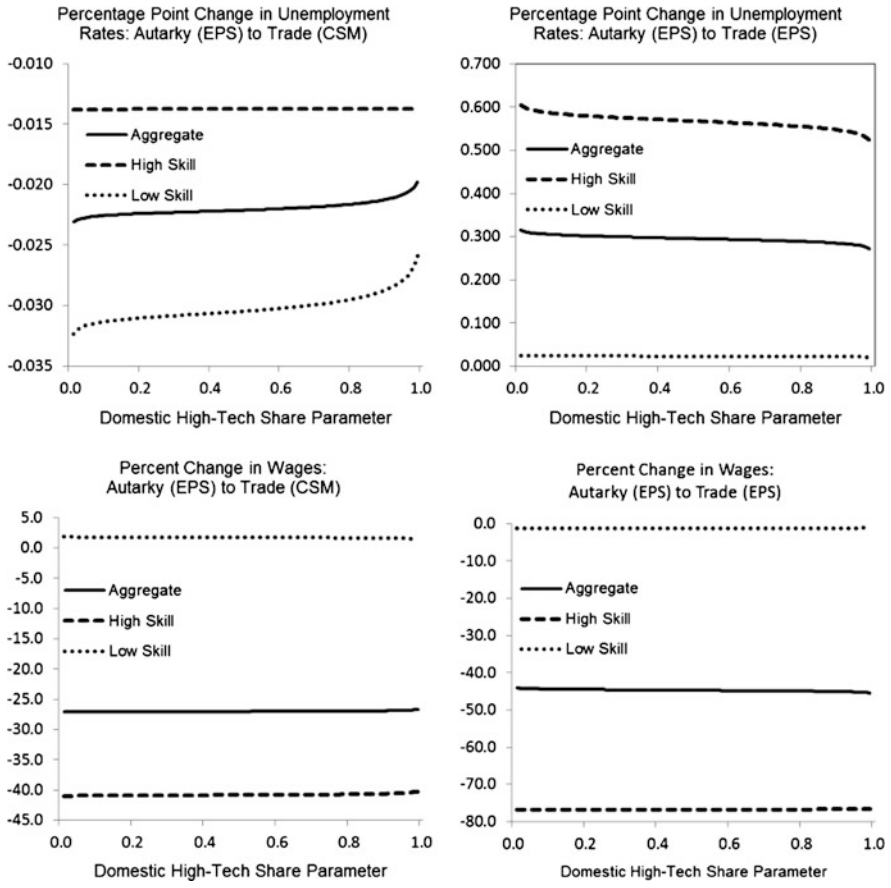

Fig. 1 Sensitivity of labor market outcomes to share of offshoring in multinational production function

equilibrium (upper and lower panels on the right). In each case, the top panels present results for the unemployment rate and the bottom panels present results for the wage. Each panel is constructed by varying the share of offshored production in the high-tech good, η (shown on the x-axis), while holding all other parameters of the model constant.

The figure shows that the baseline results are relatively insensitive to changing the offshoring share parameter. When the increase in offshoring induces a shift from EPS to CSM equilibrium, the upper and lower panels in the left column show that aggregate unemployment decreases and aggregate wages fall sharply regardless of the parameterization for η. Similarly, the upper and lower panels in the right column show that when the economy remains in EPS equilibrium, the unemployment rate rises rather than falls, and wages decline even more sharply. Overall, these results are qualitatively similar to the baseline results and suggest that the equilibrium shift, rather than the specific parameterization of the production function, is much more important in determining the response of the labor market.

5.1.2 Complementarity vs. Substitutability (σ)

Figure 2 conducts a similar exercise as in Fig. 1, but now we vary the complementarity versus substitutability of Northern and Southern inputs in the multinational's production function, only. Compared to the case in which we varied η, changes in σ yield a story that is a bit different. The panels in Fig. 2 show that the baseline results are robust provided Northern and Southern inputs are sufficiently complimentary (i.e., σ is sufficiently negative). But, as the two become more substitutable the effect on the labor market—and, in particular, unemployment—is quite different. For example, when the increase in offshoring induces a shift from EPS to CSM equilibrium, the upper left panel shows that this change drives an increase in low-skilled unemployment for low levels of complementarity. It is also interesting to note that the parameter space over which CSM equilibrium is supported does not exist for values of σ above -4.4 (the shaded region in the figure's two panels). In the next subsection we investigate this existence aspect of the model more closely. The shift from EPS equilibrium in autarky to CSM trade equilibrium shows that when Northern and Southern labor are substitutable, there is a minimal impact of unemployment and the adverse effect of the high-tech wage is notably dampened.

5.2 How General a Result is Equilibrium Mismatch?

Our main results illustrate the importance of whether or not mismatch is a feature of the post-liberalization equilibrium for how the labor market adjusts to offshoring. Implicit is the fact that in our baseline parameterization both EPS (no mismatch) and CSM (mismatch) equilibrium are feasible. How general a result is the coexistence of both types of equilibria under trade?

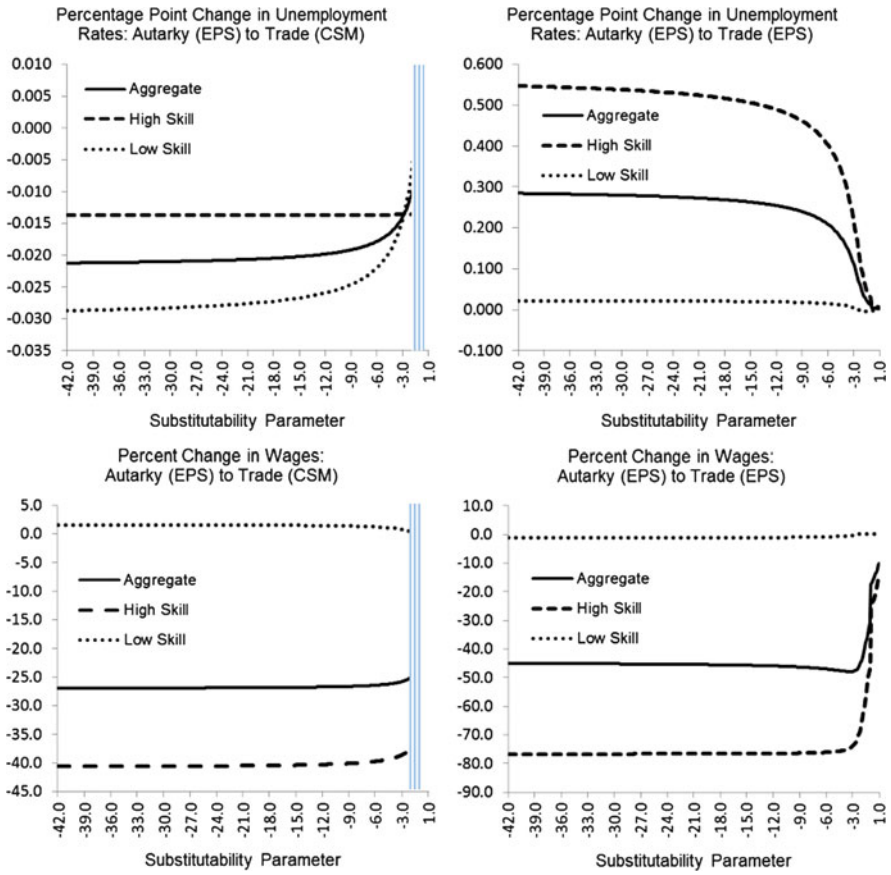

Fig. 2 Sensitivity of labor market outcomes to input substitutability parameter in multinational production function

Figure 3 offers some perspective on this. Holding constant all other parameters, the figure varies the share of offshored production in the high-tech good, η, along the x-axis and the parameter governing the substitutability of Northern and Southern high-skilled inputs, σ, along the y-axis. Every point on the grid evaluates the conditions presented in Sect. 5.2 that tell us whether CSM equilibrium can obtain alongside EPS equilibrium.

The figure is divided into two regions. The (larger) shaded region to the left denotes the part of the parameter space in which both EPS and CSM equilibria coexist. The unshaded (smaller) region to the right denotes that part of the parameter space where mismatch cannot be supported in a trade equilibrium. Finally, the black line denotes that baseline share parameterization of $\eta = 0.9$. The baseline substitutability parameter has σ such that Northern and Southern high-skilled inputs are strong complements in the multinational's production function, putting the

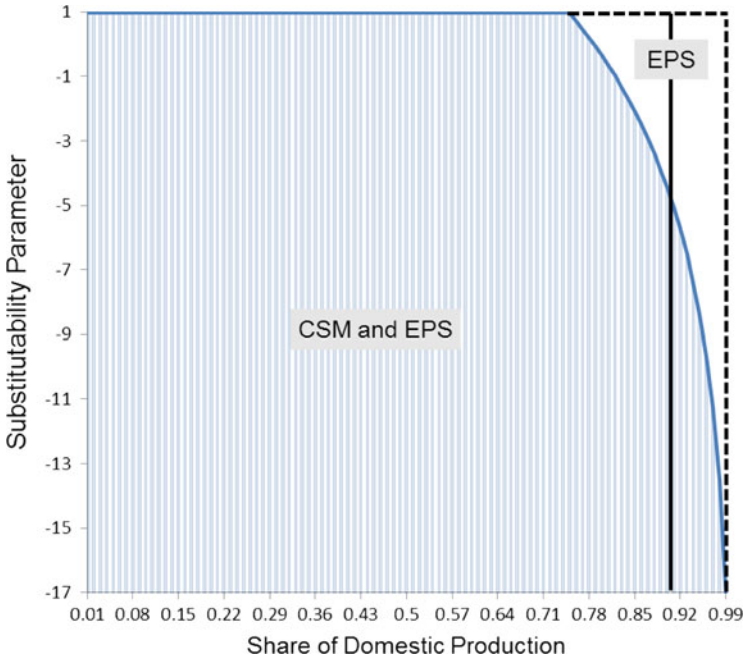

Fig. 3 Existence of CSM and EPS equilibria (substitutability and share of offshoring)

baseline model in a part of the parameter space where both equilibria coexist and mismatch is feasible.

What is interesting about Fig. 3 is that an empirically relevant parameterization is one in which the share of offshored production is relatively small, so that $0.9 \leq \eta < 1$ (see, for instance, Burstein et al. (2008) and Arseneau and Leduc (2012)). Within that range, mismatch only exists as an equilibrium outcome when domestic and foreign high-skilled inputs are sufficiently complementary. While our parameterization of σ satisfies this criteria, there is considerable uncertainty regarding whether offshored production should be viewed as a complement to or a substitute for domestic inputs. Figure 1 suggests that if one believes the inputs are substitutes (or mild complements with a sufficiently low share of offshore activity), the labor market implications of offshoring could be much larger because mismatch will not arise to help cushion the labor market effects following a trade liberalization.

6 Conclusion

We study the role of mismatch in the adjustment of the labor market to a trade liberalization that results in offshoring of high-tech production. Our results suggest that mismatch helps to mitigate the negative labor market consequences that arise

from offshoring. The ability to reallocate domestic high-skilled labor across both high- and low-tech industries dampens the increase in the unemployment rate as well as the decline in wages that result from shifting domestic production abroad. Our analysis also reveals that under our calibration the size of the parameter space over which mismatch can arise as an equilibrium outcome is, in fact, quite limited. Indeed, as long as the share of the offshored production is not too large, high-skilled workers will only choose to accept jobs with low-tech firms if domestic and offshore high-tech production are sufficiently complementary.

In general, we view the limited subset of the parameters space in which mismatch can arise as an equilibrium outcome as a limitation of the model presented in this paper. in order to bring a calibrated version of the model more in line with observed data, the model needs to be extended in a way that makes mismatch a more attractive proposition to both workers and firms following trade reform. We pursue two such extensions in companion work. Arseneau and Epstein (2014b) introduces on-the-job search to the partial equilibrium framework presented here, while Arseneau and Epstein (2014c) extends this modeling framework to a general equilibrium open economy setting with endogenous directed search.

References

Acemoglu, D., Gancia, G., & Zilibotti, F. (2012). *Offshoring and Directed Technical Change*. No. w18595, National Bureau of Economic Research.

Albrecht, J., & Vroman, S. (2002). A matching model with endogenous skill requirements. *International Economic Review, 43*(1), 283–305.

Amiti, M., & Wei, S.-J. (2005). Fear of services offshoring: Is it justified? *Economic Policy, 20*(42), 208–347.

Arseneau, D. M., & Epstein, B. (2014a). *The Welfare Costs of Skill-Mismatch Employment*. Finance and Economics Discussion Series 2014–42, Board of Governors of the Federal Reserve System.

Arseneau, D. M., & Epstein, B. (2014b). *On-the-Job Search and Mismatch in a Model of Offshoring*. Mimeo, Board of Governors of the Federal Reserve System.

Arseneau, D. M., & Epstein, B. (2014c). *Offshoring and Job Dislocation*. Mimeo, Board of Governors of the Federal Reserve System.

Arseneau, D. M., & Leduc, S. (2012). *Threatening to Offshore in a Search Model of the Labor Market*. Mimeo, Board of Governors of the Federal Reserve System.

Blinder, A. (2009). How many U.S. jobs might be offshorable? *World Economics, 10*(2), 41–78.

Bosch, M., & Maloney, W. F. (2007). *Gross Worker Flows in the Presence of Informal Labor Markets: Evidence from Mexico, 1987-2002*. Mimeo.

Burdett, K., & Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review, 39*, 257–273.

Burstein, A., Kurz, C., & Tesar, L. (2008). Trade, production sharing, and the international transmission of business cycles. *Journal of Monetary Economics, 55*(4), 775–795.

Christensen, B. J., Lentz, R., Mortensen, D. T., Neumann, G. R., & Werwatz, A. (2005). On-the-job search and the wage distribution. *Journal of Labor Economics, 23*(1), 31–58.

Crino, R. (2009). Offshoring, multinationals, and labour market: A review of the empirical literature. *Journal of Economic Surveys, 23*(2), 197–249.

Davidson, C., & Matusz, S. J. (2009). *International trade with equilibrium unemployment*. Princeton, NJ: Princeton University Press.

Davidson, C., Matusz, S. J., & Shevchenko, A. (2008). Outsourcing Peter to Pay Paul: High-skill expectations and low-skill wages with imperfect labor markets. *Macroeconomic Dynamics, 12*, 463–479.

Diamond, P. A. (1982). Wage determination and efficiency in search equilibrium. *The Review of Economic Studies, 49*(2), 217–227.

Dolado, J. J., Jansen, M., & Jimeno, J. F. (2009). On-the-job search in a matching model with heterogeneous jobs and workers. *The Economic Journal, 119*, 200–228.

Ebenstein, A. Y., Harrison, A. E., McMillian, M. S., & Phillips, S. (2011). *Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Surveys*. No. w15107, National Bureau of Economic Research.

Elsby, M. W. L., Michaels, R., & Solon, G. (2009). The ins and outs of cyclical unemployment. *American Economic Journal: Macroeconomics, 1*(1), 84–110.

Epstein, B. (2012). *Heterogeneous Workers, Optimal Job Seeking, and Aggregate Labor Market Dynamics*. International Finance Discussion Papers 1053, Board of Governors of the Federal Reserve System.

Feenstra, R. C., & Hanson, G. H. (1996). Globalization, outsourcing and wage inequality. *American Economic Review, 86*(2), 240–245.

Feenstra, R. C., & Hanson, G. H. (1997). Foreign direct investment and relative wages: Evidence from Mexico's Maquiladoras. *Journal of International Economics, 42*(3–4), 371–393.

Feenstra, R. C., & Hanson, G. H. (1999). The impact of outsourcing and high-technology capital on wages: Estimates for the United States, 1979–1990. *Quarterly Journal of Economics, 114*(3), 907–940.

Gautier, P. A. (2002). Unemployment and search externalities in a model with heterogeneous jobs and workers. *Economica, 69*(2), 21–40.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review, 104*(8), 2509–2526.

Gorg, H., & Hansely, A. (2005). Labour demand effects of international outsourcing: Evidence from plant-level data. *International Review of Economics and Finance, 14*(3), 365–376.

Grossman, G. M., & Rossi-Hansberg, E. (2008). Trading tasks: A simple theory of offshoring. *American Economic Review, 98*(5), 1978–1997.

Helpman, E. (2006). Trade, FDI, and the organization of firms. *Journal of Economic Literature, 44*(4), 580–630.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies, 57*(2), 279–298.

Hummels, D., Munch, J. R., & Xiang, C. (2011). Offshoring, transition, and training: Evidence from Danish matched worker-firm data. *American Economic Review: Papers and Proceedings, 102*(3), 424–428.

Jacobson, L. S., Lalonde, R. J., & Sullivan, D. J. (1993). Earning losses of displaced workers. *American Economic Review, 83*(4), 685–709.

Kletzer, L. (2001). *What are the costs of job loss from import-competing industries?* Washington, DC: Institute for International Economics.

Liu, R., & Treffler, D. (2008). *Much Ado About Nothing: American Jobs and the Rise of Service Offshoring to China and India*. No. w14061, National Bureau of Economic Research.

Mitra, D., & Ranjan, P. (2010). Offshoring and unemployment: The role of search frictions and labor mobility. *Journal of International Economics, 81*(2), 219–229.

Mortensen, D. T. (1982). Property rights in mating, racing, and related games. *American Economic Review, 72*(5), 968–979.

Mortensen, D. T. (2005). *Wage dispersion: Why are similar workers paid differently?* Cambridge, MA: MIT Press.

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies, 61*(3), 397–415.

Mortensen, D. T., & Pissarides, C. A. (1999). Unemployment responses to 'skill-biased' technology shocks: The role of labour market policy. *The Economic Journal, 109*(455), 242–265.

Oldenski, L. (2014). Offshoring and the polarization of the U.S. labor market. *Industrial and Labor Relations Review, 67*, 734–761, Supplement.

Petrongolo, B., & Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature, 39*(2), 390–431.

Ranjan, P. (2013). Offshoring, unemployment, and wages: The role of labor market institutions. *Journal of International Economics, 89*(1), 172–186.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review, 95*(1), 25–49.

Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics, 15*(2), 127–148.

Sicherman, N. (1991). Overeducation in the labor market. *Journal of Labor Economics, 9*, 101–112.

Good Times and Bad Times, with Endogenous Trade Policy Responses

T. Huw Edwards

Abstract Globalisation is associated with long periods of sustained economic growth and credit expansion, whereas major recessions tend to lead to falling trade and protectionism. I investigate this, using a model where an important component of trade is search by firms trying out new trade partners. To model this, I set up a schematic model of upstream-downstream matching between firms, where match quality determines profits. Prior to globalisation, Northern upstream and downstream firms work together, to avoid trade costs. As trade is liberalised, new North-South matches begin to develop, but at first these are footloose (since many are experimental in nature), and can be driven out rapidly if there is a demand crisis.

The presence of a highly elastic, searching element of trade, during periods of global liberalisation and fast growth means that countries cannot gain terms-of-trade advantage from protection. By contrast, if a prolonged recession drives out searching firms, the remaining trade is relatively price-inelastic, and beggar-thy-neighbour tariffs become attractive.

Such observations are consistent with the rapid and damaging switch to protectionism after the 1929 Stock Market Crash, and emphasise the need for continued multilateral commitment to trade liberalisation, in the aftermath of the 2008 Great Trade Recession.

1 Introduction

> *'Good times dampen anti-globalization attitudes, while bad times deepen them.' - Jagdish Bhagwati (2004).*

Why do periods of prolonged global growth, such as 1980–2007, tend to be accompanied by even faster growth of World trade, particularly inter-firm, inter-industry trade? Why is the stalling of such growth traditionally associated with protectionism? In addressing these questions, I wish particularly to stress the two-way

T.H. Edwards (✉)
School of Business and Economics, Loughborough University, Ashby Road, Loughborough LE11 3TU, UK
e-mail: T.H.Edwards@lboro.ac.uk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_4

nature of the relationship between globalisation and worldwide growth. During the boom years prior to the Great Trade Recession of 2008, India, China and Eastern Europe were integrated into the World economy, production and supply chains were increasingly split across international borders, as highly productive international partnerships and multinational enterprises spread and took advantage of global differences in comparative advantages. Many of these features (though not the vertical fragmentation of production) replicate those of the pre-1929 globalisation boom. By contrast, the 1930s saw most of these trends go into sharp reversal (Crafts, 2004). Despite some incidences of 'murky protectionism' (in Baldwin and Evenett's 2009 terminology) there has not been a similar lurch into economic nationalism since 2008, and indeed the TTIP negotiations between the EU and USA indicate there is still a will to revive trade liberalisation, probably reflecting an appreciation of the catastrophic implications when the 1929 Wall Street Crash led to a protectionist response. Against this, recent European election results suggest increasing support for nationalist and anti-globalization parties.

One reason why so many countries liberalised trade, particularly following the fall of the Soviet Union, is the (not uncontroversial) observation that more open, market-orientated economies have tended to experience faster growth.[1]

In one respect, this stands in contrast to textbook economic thinking. After all, while trade is good for all economies taken together, in most circumstances textbook analysis would say that it is in the individual interests of 'large' countries to impose a protectionist set of 'optimal' tariffs on imports and exports, in order to improve their terms of trade,[2] so that there is a prisoner's dilemma situation, whereby the noncooperative outcome of tariff-setting games is against the collective interest. By contrast, one might tentatively view the evidence of the pre-2007 boom as suggesting that there is little evidence of a prisoner's dilemma: open trade policies, which favour a country's neighbours, also seem in many circumstances to serve its self-interest. This is a feature which I particularly wish to investigate: is there some characteristic of globalisation booms which leads to a convergence of country and global welfare-maximising policies, and, if so, does such a convergence break down during prolonged recessions or depressions?

In this paper, I link this, in particular, with the high observed income elasticity of trade during the economic 'good times' (Ishii and Yi, 1997). The starting point is that this high elasticity represents a footloose element of supply, which is explicitly linked to the dynamics of a search process, and which may be the first thing to be eliminated during an economic downturn. Once this footloose element is eliminated, the economics of trade protection are turned on their head—essentially, the prisoner's dilemma is returned. This paper is one of a number

[1] Wacziarg and Welch (2008). Dollar and Kraay (2002 and 2003). These studies are not without their critics: for example, Luebker et al. (2002). Winters et al. (2004) give a more overall survey of the benefits of trade liberalisation.

[2] E.g. Gros (1987).

of studies extending the analysis of search frictions—initially developed in the context of labour markets by Mortensen and Pissarides, among others—to inter-firm relationships.[3]

The structure of the paper is as follows.[4] The rest of Sects. 1 and 2 build on the theoretical ideas of existing papers by Ishii and Yi (1997), Rauch and Trindade (2003) and Rauch and Casella (2003) and Grossman and Helpman (2002 and 2005), to develop a partial equilibrium model of a single industry with two-stage production, outlining the difference between growing and declining sectors and that between matched and searching firm pairs. In Sect. 3, I then set up a three-phase time schema, whereby in phase 1 there is little trade, phase 2 sees globalisation and phase 3 sees an economic crisis. Section 4 discusses how either a demand crisis or a prolonged period of demand stagnation can affect this model. Section 5 is perhaps the most important part of the paper, since it develops the model into an explanation of why optimal tariffs may be low during periods of economic growth, but high during periods of economic slowdown. In particular, behaviour can switch from free-trade to protectionist, if there is a crisis, particularly if it is followed by growth resuming at below a critical threshold level. This is illustrated by a numerical simulation model, calibrated to a stylised data set. Section 6 discusses the implications in terms of the return of the prisoner's dilemma in trade protectionism during bad times, and emphasises the importance of maintaining a global commitment to trade liberalisation.

It is also important to ask whether search is, in fact, quantitatively important enough to influence behaviour in the ways suggested in this paper. Evidence from Besedes and Prusa (2006) and Eaton et al. (2007) suggests fast entry and exit. Despite the small scale of firms on initial entry, during the search process, those new entrants who survive rapidly build up to account for perhaps 20–25 % of trade volumes. Tentative evidence linking outsourcing to search is also discussed.

1.1 Fixed Costs, Search Costs and Trade Elasticities

It has become a commonplace that costs of market entry impede global trading patterns, such that economic integration, even in the good times of the 1990s and 2000s, falls far short of what neoclassical theory would predict (Trefler 1995; Obstfeld and Rogoff 2000). Since this primarily reflects limitations on the extensive margins of trade (in other words, the fact that most firms do not trade most products internationally), this is widely explained in terms of a fixed cost of market entry (Tybout, 2002)—firms need to make a particular level of profit in order to justify

[3] See Mortensen and Pissarides (1999) for a survey. Papers applying search and matching to trade include Eaton et al. (2007).

[4] The paper is a revised version of Edwards (2010).

market entry, there are both selection effects in favour of large and successful firms (Bernard et al., 2007) and threshold effects on entry. The latter, threshold effect can also be explained simply in terms of a Ricardian model of comparative advantage (Yi, 2003): once trade costs fall sufficiently, the fragmentation of production becomes possible so that different stages can be carried out in different countries. Consequently, there is a rapid, nonlinear growth of market participation, particularly in terms of vertical tie-ups between firms (Yi, 2003). In addition, where tariffs are ad valorem, effective protection against individual stages of production is magnified (since a tariff may be paid more than once, as the goods cross and re-cross borders), and the same applies to transport costs—so exaggerating the apparent price-sensitivity of intermediates trade.

Many of these arguments are often summarised as the 'new, new' trade theory[5]—i.e. supplementing the models of Krugman (1979) and Grossman and Helpman (1992), with the incorporation of firm-level participation effects and production fragmentation. These elements are now acknowledged to have powerful implications for growth through firm selection,[6] distribution[7] and the observed volatility of trade.[8] By themselves, however, they offer only a partial explanation, containing no real evidence of the nature of the fixed costs of market participation, and under what circumstances they may become sunken rather than just fixed. To explain this, we need to turn to another branch of the literature—the evidence of informational barriers and a search process. This stems from certain stylised facts: the high turnover of trading firms (Besedes and Prusa, 2006) and the evidence of network effects (Rauch and Trindade, 2003), taken with strong evidence that clusters of low-cost producers can go unnoticed by traders for long periods (the noted example being surgical steel production in Sialkot, Pakistan[9]). Informational barriers can generate a search process, and I argue that this process is a primary cause of threshold effects, which in turn helps explain the dynamics of trade.[10] Search takes time and requires confidence and the availability of credit. Critically, firms need to incur a series of ongoing fixed costs during the search process, but once they eventually achieve a satisfactory match, the cost of the past search is sunken. This means that searching firms will indeed be very sensitive to price movements, as in Ishii and Yi (1997), but that, as the search process goes on, firms achieve matches and become more heterogeneous and less subject to threshold effects. This process indicates a sensitivity of trade during the process of globalisation to international

[5]To use Srinivasan and Archana's (2009) terminology.

[6]Melitz (2003).

[7]Feenstra and Hanson (1999).

[8]Ishii and Yi (1997).

[9]Schmitz (1999).

[10]Ishii and Yi (1997) use fixed costs of vertical specialization to explain the high observed income elasticities of trade—which they argue cannot plausibly reconciled with more orthodox models.

traded prices, global aggregate demand and capital availability,[11] but the trading patterns of long-established firms, which are likely to be much less sensitive,[12] and it is this nonlinearity—missed by the existing literature—which may underlie the danger of policy shifts during and after economic crises.

An important empirical question is how much trade, in practice, can be considered to be searching. Evidence on this is somewhat hard to come by, partly because most firm-level trade databases have until recently covered timespans too short to assess long-run industry dynamics thoroughly. Nevertheless, Besedes and Prusa's (2006) study of US imports suggests a 1 year average turnover for new products entering the market. In a model of industry dynamics, these would then be replaced by other new entrants in a repeating cycle. Eaton et al.'s (2007) study of Colombian exporters shows even faster exit rates (83 % of entrants in 1997 left after a year), but continuing high exit rates in the second year (38 %) and third year (10 %) after entry. The number of firms entering in any one year is very large compared to the overall number of firms in the market (for example, in 2005, there were 5060 firms which had been exporting for less than a year, out of 11, 720 exporters): however, firms exporting in the first year tend to export very small amounts. The scale of survivors each year then increases steadily over at least a decade, as they expand their networks of customers. Given the continuing build-up of customers per firm over 10 years, we could also view all firms with less than ten years in the market as being searchers (which would imply nearly 25 % of total sales). A slight caution here: Colombian exports tend to be dominated by small flower-growers, who can enter foreign markets easily and are less tied to a single customer (e.g. by blueprints) than exporters from other countries in other industries might be.

An alternative argument that search is quantitatively significant is given in Edwards and Ferrett (2009), based upon the idea that merger with a downstream supplier usually gives better coordination and security, whereas inter-firm trade benefits from flexibility, and so is associated with search or experimentation. This is supported by evidence from Spencer (2005), showing that outsourcing in China preceded the growth of FDI, but that perhaps 20 % of firm-level trade is still accounted for by outsourcing rather than FDI.

2 An Illustrative Model of Firm-Level Trade

I set up a simple, stylised, partial equilibrium model of a monopolistically competitive industry in a two-country world—the two countries being the North and the South, the former being characterised by higher skill endowments per head.

[11] In terms of a macroeconomic model, market search has elements of capital formation, and so it should be no surprise that it has many of the characteristics of investment demand—in terms of cyclical sensitivity.

[12] This reflects in part the heterogeneity both of firms (Melitz, 2003) and of trading match quality (Rauch and Casella, 2003).

The main market for final goods is in the North. Production requires two stages, which I name upstream (u) and downstream (d). Typically these are carried out by a pairing of firms (which may or may not be vertically integrated by merger), where u sells a semi-finished good to d, who then completes the manufacture and sells it on to final consumers. The two firms are of equal size and *ex ante* expected efficiency: however, productivity varies depending on the goodness of fit of the match, μ_i. As in Rauch and Casella (2003) or Rauch and Trindade (2003), μ, potentially follows a uniform, rectangular distribution between 0 and 1, and firms do not know μ_i before entering a match i, though they know its overall distribution.[13]

Trade between the North and South develops over time. The historical setup takes three phases. In phase 1, trade costs are high, so that there is little trade, and most goods are produced by pairings of firms type u and d within the North. However, the South is assumed to have a potential comparative advantage in upstream production, while the North has a comparative advantage in downstream production. We then enter phase 2: a period of growth and global integration, spurred by a technological or policy change reducing trade costs. Some Northern, downstream firms (though not all) will now search for upstream partners in the South. For example, garments might be manufactured by an upstream firm in China, but according to designs from the downstream firm in a Western economy, which then completes the marketing and distribution worldwide. Phase 3 represents an unanticipated crisis, where credit ceases to be available and growth stalls for a protracted period.

Concentrating on phase 2, the period of globalisation, the growth of the outsourcing trade is impeded by search friction deriving from an assumed need for at least one firm to make a relationship-specific investment: in order to avoid a potential hold-up problem,[14] this generally requires a contractual relationship for at least some minimum period, which I characterise by a fixed contract period, t, during which the two firms have an exclusive relationship.

Firms employ labour in the form of fixed and variable elements. The cost of the latter is normalised at $C = C_N$ for North-North pairings, and at $C = C_S$ ($C_S < C_N$) for North-South pairings. The elasticity of substitution between final goods varieties is ε, which will also closely approximate the own-price elasticity for the output sold by firm pairings, at least as long as the number of firms, N, is large.

The model is driven by an assumed coordination cost which is inversely linearly related to the quality of the match μ_i between firms in pairing i. For simplicity, I concentrate on a rather stylised model where quality of match affects fixed cost, rather than unit variable costs. More precisely, I assume that fixed costs are $F - \mu_i$, where $F > 1$ and $0 \leqslant \mu_i \leqslant 1$. In addition, I assume initially that both firms in a pairing split profits equally.

[13]This setup is derived from Salop's circular cylinder, and is standard in firm-level matching models. Note that Grossman and Helpman's (2002) model is similar, except that firms know with certainty the location of potential partners, and always match with the nearest.

[14]See Hart (1995).

2.1 The Matching Process

We now want to look at the matching process in more detail.

The paper follows Rauch and Trindade (2003) and Rauch and Casella (2003) in using a matching framework based upon Salop's circular cylinder (note that Grossman and Helpman 2002, use a similar setup). This is shown in Fig. 1. Position on the circle refers to some firm-specific characteristics. The essence of the Salop model is that firms are *ex ante* equal in efficiency, but that firm performance is determined by the degree of fit with the match partner: the aim is to match with a firm directly opposite on the cylinder. Hence, match quality, μ_i, is measured by the circumference distance between the two firms. Firms only have a single partner at any time.

The key difference between the match-searching model employed here and the models in the earlier papers is the ability of a firm in an unsatisfactory match to renew search, after a given contract period, t, during which firms face a discount rate r. The parameters r and t help to determine a reservation match quality, μ_R. Firm pairings are also subject to sudden, random death with a constant probability of δ. For simplicity, I normalise the discount rate in terms of the contract period,

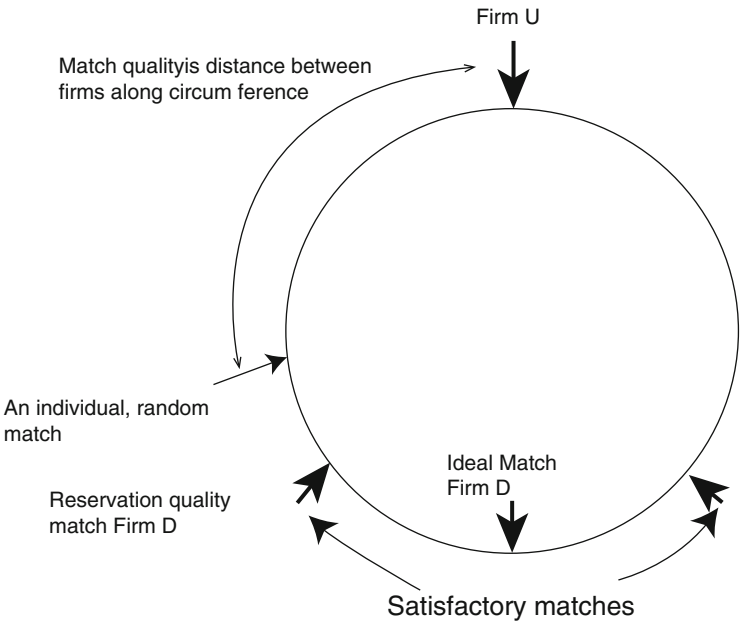

Fig. 1 The Salop matching framework

defining the discount rate and depreciation rate per contract period,

$$\rho = (1+r)^t - 1; \tag{1}$$

$$D = \frac{1}{(1-\delta)^t} - 1. \tag{2}$$

Note that the depreciation rate reflects the probability of either firm in a pairing dying during the contract period.

Any match with a circumference length greater than μ_R will yield a profit great enough for the firms to choose to continue. This gives a probability of acceptance of $1-\mu_R$. However, of these pairings, proportion D will naturally expire anyway during the first contract period, so any surviving firm will have to renew search. By contrast, proportion μ_R of initial pairings will be unsatisfactory, and be dropped after one contract period.

We note that the profit of a reservation quality match ($\mu_i = \mu_R$) is zero. Profit increases linearly with respect to match quality, so that the expected profit of a successful match (where $\mu_R < \mu_i < 1$) is $\frac{1+\mu_R}{2}$ minus the profit of a reservation quality match. This yields an expected profit of $\frac{1-\mu_R}{2}$. Likewise, an unsatisfactory match will have an expected profit of $-\frac{\mu_R}{2}$ (see Table 1). Finally note that a renewed search will yield a present discounted value of zero.

We wish now to solve this problem, based upon the assumption that monopolistic competition will equate the expected return from entering a search process to zero.

Discounted profit with a successful match = one period expected loss with a poor match;

$$\frac{(1-\mu_R)^2}{2\Phi} = \frac{\mu_R^2}{2(1+\Phi)}, \text{ where } \Phi = \rho + D;$$

$$\mu_R = \frac{2(1+\Phi) \pm \sqrt{4(1+\Phi)^2 - 4(1+\Phi)}}{2},$$

$$\mu_R = 1 + \Phi - \sqrt{\Phi(1+\Phi)}, \tag{3}$$

Table 1 Expected returns for the matching process

		Probability	Expected profit above reservation
Satisfactory match	Both firms survive	$(1-\mu_R)(1-D)$	$\frac{1+\mu_R}{2} - \mu_R = \frac{1-\mu_R}{2}$
	One or more dies	$(1-\mu_R)D$	0
Unsatisfactory match		μ_R	$\frac{\mu_R}{2} - \mu_R = -\frac{\mu_R}{2}$

since only the negative root lies between zero and unity. Expanding Eq. (3), the reservation match quality will satisfy

$$\mu_R = 1 + \rho + D - \sqrt{(\rho + D)(1 + \rho + D)}, \tag{4}$$

which is declining as ρ or D increases.

From (4), we can conclude that, with zero discount rates and zero probability of firm death ($r = d = \rho = D = 0$), firms will be infinitely patient in their search, so that the long-run equilibrium is for all firms to be equally and successfully matched. By contrast (and more realistically), with positive time preference, the long-run equilibrium has a uniform distribution of surviving matches, of quality ranging from μ_R to 1, with profits likewise varying. A consequence is that firm pairings are heterogeneous in the long-run, both in terms of profits and trade volumes, with heterogeneity increasing when long-run interest rates are higher or contract periods are longer.

2.2 *The Role of Monopolistic Competition*

The industry is assumed to be monopolistically competitive, on the lines of Krugman (1979). There are both fixed and variable costs. Subject to these, firms can enter or exit the market, although they need a partner (existing or new) in order to produce saleable goods. The elasticity of substitution between final goods varieties is ε (> 1),[15] which closely approximates the own-price elasticity for the output sold by firm pairings, at least as long as the number of firms, N, is large.

To summarise the love of variety model, we start with a Dixit-Stiglitz utility function for utility from consumption of the industry's good:

$$\overline{Y} = \left(\sum_{i=1}^{N} Y_i^{\frac{\varepsilon - 1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon - 1}}. \tag{5}$$

The first order condition for an optimum is

$$\frac{\partial \overline{Y}}{\partial Y_i} = \left(\sum_{j=1}^{N} Y_j^{\frac{\varepsilon - 1}{\varepsilon}} \right)^{\frac{1}{\varepsilon - 1}} Y_i^{\frac{-1}{\varepsilon}} = \frac{P_i}{P^*};\ i \in j, \tag{6}$$

[15]The restriction $\varepsilon > 1$ is associated with consumers' assumed 'love of variety', and also helps ensure finite pricing by firms.

where P^* is the aggregate CES price of utility. We can replace the term in brackets by substituting from (5) into (6), so that

$$\frac{\partial \overline{Y}}{\partial Y_i} = \overline{Y}^{\frac{1}{\varepsilon}} Y_i^{\frac{-1}{\varepsilon}} = \frac{P_i}{P^*}. \tag{7}$$

Or, by rearranging,

$$Y_i = \overline{Y}(\frac{P_i}{P^*})^{-\varepsilon}. \tag{8}$$

The industry is described as 'thick' in the sense that the number of firms producing differentiated goods is large.

When the number of firms, N, is large, then a change in Y_i will have negligible impact upon U, in which case $\frac{\partial \ln Y_i}{\partial \ln P_i} = -\varepsilon$. Applying the standard profit-maximising formula for a firm with constant demand elasticity and a constant marginal cost, C_i, firm pairing i will charge a price of

$$P_i = \frac{\varepsilon}{\varepsilon - 1} C_i. \tag{9}$$

2.3 The Aggregate Price Level and Number of Firms

Substituting from (9) into (8) and rearranging, we obtain an equation for the aggregate price level,

$$P^* = \left(\sum_{j=1}^{N} P_j^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}. \tag{10}$$

When all firms are identical, $P_j = \overline{P}\ \forall j$. Consequently,

$$P^* = N^{\frac{1}{1-\varepsilon}} \overline{P} = N^{\frac{1}{1-\varepsilon}} \frac{\varepsilon}{\varepsilon - 1} \overline{C}, \tag{11}$$

which is decreasing with respect to N for $\varepsilon > 1$, demonstrating the love of variety effect.

The profit of firm i, before taking account of fixed cost, is

$$\pi_i = (P_i - C_i) Y_i = \frac{1}{\varepsilon - 1} C_i Y_i = \frac{1}{\varepsilon - 1} C_i \overline{Y} (\frac{\frac{\varepsilon}{\varepsilon - 1} C_i}{P^*})^{-\varepsilon}, \tag{12}$$

$$\pi_i = k \overline{Y} P^{*\varepsilon} C_i^{\,1-\varepsilon},$$

where

$$k = \frac{1}{\varepsilon}\left(\frac{\varepsilon}{\varepsilon - 1}\right)^{1-\varepsilon}.$$

Profit after fixed cost

$$\Pi_i = k\overline{Y}P^{*\varepsilon}C_i^{\,1-\varepsilon} - F + \mu_i, \tag{13}$$

where F (> 1) is fixed cost with zero match quality, and μ_i is match quality, which varies randomly between 0 and 1.

Where all firms have identical marginal costs

$$\Pi_i = kP^{*\varepsilon}\overline{C}^{1-\varepsilon} - F + \mu_i,$$

$$\Pi_i = \frac{\overline{Y}}{\varepsilon - 1}N^{\frac{\varepsilon}{1-\varepsilon}}\overline{C} - F + \mu_i. \tag{14}$$

In the monopolistically competitive model, a firm pairing with the reservation match quality, $\mu_i = \mu_R$, will just break even after its fixed costs. Hence, in an equilibrium where all firms have the reservation match quality,

$$\frac{\overline{Y}}{\varepsilon - 1}N^{\frac{\varepsilon}{1-\varepsilon}}\overline{C} - F + \mu_R = 0;$$

$$N = \left(\frac{(\varepsilon - 1)(F - \mu_R)}{\overline{YC}}\right)^{\frac{1-\varepsilon}{\varepsilon}}. \tag{15}$$

Substituting into (11), we obtain P^*,

$$P^* = N^{\frac{1}{1-\varepsilon}}\frac{\varepsilon}{\varepsilon - 1}\overline{C},$$

$$P^* = \left(\frac{\overline{Y}}{(\varepsilon - 1)(F - \mu_R)}\right)^{-\frac{1}{\varepsilon}}\frac{\varepsilon}{\varepsilon - 1}\overline{C}^{\frac{\varepsilon - 1}{\varepsilon}}. \tag{16}$$

Note that, in a model with constant and identical marginal cost, all firms will be of the same scale, so all will set prices and output at the level that a firm pairing with reservation match quality would set. Hence (15) and (16) describe the equilibrium with identical firms.

Another way of writing (11) is to take

$$\pi_i = k\overline{Y}P^{*\varepsilon}\overline{C}^{1-\varepsilon} = F - \mu_R;$$

$$P^* = \left(\frac{F - \mu_R}{k\overline{Y}\overline{C}^{1-\varepsilon}}\right)^{\frac{1}{\varepsilon}}. \tag{17}$$

2.4 Higher-Level Demand

Our model is partial equilibrium, so the aggregate price, P^*, refers to a single industry. Consumption of the industry good, at price P^*, leads to utility of $\overline{Y}(P^*)$. This is assumed to be isoelastic, so that

$$\overline{Y} = \overline{Y}_0 P^{*-\upsilon}, \tag{18}$$

where $\upsilon > 0$. Hence, substituting into (16),

$$\begin{aligned} P^* &= P^{*\frac{\upsilon}{\varepsilon}} \left(\frac{\overline{Y}_0}{(\varepsilon-1)(F-\mu_R)}\right)^{-\frac{1}{\varepsilon}} \frac{\varepsilon}{\varepsilon-1} \overline{C}^{\frac{\varepsilon-1}{\varepsilon}}; \\ P^* &= \left(\frac{(\varepsilon-1)(F-\mu_R)}{\overline{Y}_0}\right)^{\frac{1}{\varepsilon-\upsilon}} \left(\frac{\varepsilon}{\varepsilon-1}\right)^{\frac{\varepsilon}{\varepsilon-\upsilon}} \overline{C}^{\frac{\varepsilon-1}{\varepsilon-\upsilon}}. \end{aligned} \tag{19}$$

For utility to be decreasing with respect to $\overline{C}$, we need the parameter restriction that $\varepsilon > \upsilon$.

Assuming a Cobb-Douglas higher-level nesting of the economy, then, as long as the industry is 'small' in proportion to the overall economy, we can take overall national income as given, and hence assume $\upsilon = 1$. Hence, (19) becomes

$$P^* = \left(\frac{(\varepsilon-1)(F-\mu_R)}{\overline{Y}_0}\right)^{\frac{1}{\varepsilon-1}} \left(\frac{\varepsilon}{\varepsilon-1}\right)^{\frac{\varepsilon}{\varepsilon-1}} \overline{C}. \tag{20}$$

2.5 Firm Threshold Prices

An existing, successfully-matched firm pairing will have a match quality $\mu_R < \mu_i < 1$.

We define the **reservation market price**, P_i', for a firm with match quality μ_i, as being the value of P^* at which a firm with match quality μ_i will break even. Beyond the initial search period, the firm will only choose to stay open if $P^* \geqslant P_i'$.

From (13) this condition is

$$\begin{aligned} k\overline{Y}P^{*\varepsilon}{C_i}^{1-\varepsilon} - F + \mu_i &= 0; \\ P' &= \left(\frac{F-\mu_i}{k\overline{Y}{C_i}^{1-\varepsilon}}\right)^{\frac{1}{\varepsilon}} \end{aligned} \tag{21}$$

When $\mu_i = \mu_R$, this is satisfied by P^* as calculated in the previous section. By contrast, when $\mu_i = 1$, this corresponds to

$$P' = \left(\frac{F-1}{k\overline{Y}{C_i}^{1-\varepsilon}}\right)^{\frac{1}{\varepsilon}}, \tag{22}$$

which is lower than P^*, firstly because $F-1<F-\mu_R$, and also because the term on the denominator, is an increasing function of $\overline{Y}$, which should improve as P falls. The intuition is that the most efficient matched firm pairings will withdraw from the market at a lower price than the threshold for new firm entry, and this difference is greater the lower is μ_R, and hence the more heterogeneous are existing matched pairings.

Since $k=\frac{1}{\varepsilon}(\frac{\varepsilon}{\varepsilon-1})^{1-\varepsilon}$, assuming a top-level own price elasticity of unity for the aggregate industry produce, and substituting into (20) we can write

$$P_i'=(\frac{F-\mu_i}{kC_i{}^{1-\varepsilon}})^{\frac{1}{\varepsilon}}\overline{Y}_0^{\frac{1}{\varepsilon}}P'^{\frac{1}{\varepsilon}};$$

$$P_i'=(\frac{F-\mu_i}{kC_i})^{\frac{1}{\varepsilon-1}}\overline{Y}_0^{\frac{1}{\varepsilon-1}}C_i. \tag{23}$$

Hence, we have a situation where P_i' is proportional to C_i. Also note that the reservation prices for market withdrawal when $\mu_i=\mu_R$ and $\mu_i=1$, in the case where marginal cost, $C_i=\overline{C}$ are

$$P^*=(\frac{F-\mu_R}{kC_i})^{\frac{1}{\varepsilon-1}}\overline{Y}_0^{\frac{1}{\varepsilon-1}}\overline{C};$$

$$P'=(\frac{F-1}{kC_i})^{\frac{1}{\varepsilon-1}}\overline{Y}_0^{\frac{1}{\varepsilon-1}}\overline{C};$$

$$\frac{P'}{P^*}=(\frac{F-1}{F-\mu_R})^{\frac{1}{\varepsilon-1}}. \tag{24}$$

3 The Stages of Development of North-South Interfirm Trade

I discuss the three phases of the stylised time-scheme in turn.

3.1 Phase 1: The Pre-Globalisation, Northern Economy

I start by concentrating on the initial phase 1 of our time-scheme, when trade costs are sufficient to ensure that there are no matches between Northern and Southern partners. I consider an equilibrium, where the industry in the North is initially static, with demand and prices constant.

I start by characterising firm pairs on this initial equilibrium growth path as 'established' or 'matched', if they are already in a settled match ($\mu_i \geq \mu_R$) at the start of a contract period—otherwise, we define them as 'searching'.

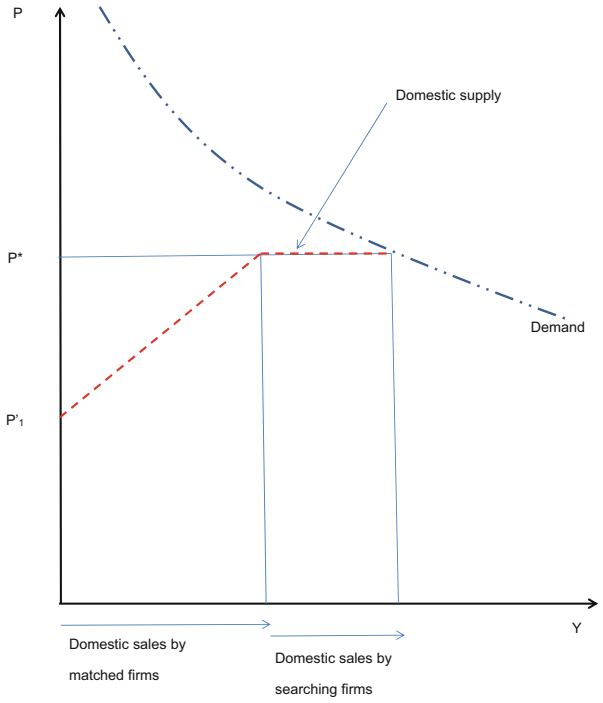

Fig. 2 Equilibrium in the pre-globalisation, Northern economy

The basic situation is shown in Fig. 2. Since unit variable cost with Northern pairings is 1, firm pairs will supply at a price of

$$P_1^* = \left(\frac{F-\mu_R}{k}\right)^{\frac{1}{\varepsilon}}. \tag{25}$$

Searching firms will only engage in search if the price is at least P_1^*: free entry will keep the price to this level, as long as the industry is not shrinking. Matched firm pairings will also supply at this price, though most of these are intramarginal matches, which would continue to produce if the price were to fall (hence the upward-sloping supply curve for matched firms). The number of firms in the industry will satisfy (14), and also balance demand and supply at the price P_1^*: I will denote this initial demand level for the industry's products as Y_1^*.

If firms were infinitely-lived, the economy would tend towards a steady-state equilibrium where all firms were matched ($\mu_i \geq \mu_R$). However, if proportion D of firm pairings suffer in a given contract period from one or both firms randomly expiring, then there will be some new firms entering and searching. In equilibrium, I denote the number of matched firms in period 1 as N_{m1} and the number of searching firms as N_{s1}. In the next contract period, proportion D of existing matched pairings will expire, while proportion $(1-\mu_R)$ of searching pairings will find established

partners. In a static equilibrium, where N_m is constant over time, this implies

$$\frac{N_{s1}}{N_{m1}} = \frac{D}{1-\mu_R}. \tag{26}$$

If the number of firms in the industry were growing at rate G, this would become

$$\frac{N_{s1}}{N_{m1}} = \frac{G+D}{1-\mu_R}. \tag{27}$$

This situation is shown in Fig. 2. The price in equilibrium, P_1^*, is given by the entry price for new, searching pairings. However, there is a kink in the supply curve, since searching pairings will all enter or leave the market at $P_1 = P_1^*$, while matched firms are heterogeneous in the price at which they would exit the market. The most efficient firm pairing would potentially leave the market at $P_1 = P_1^*$, as given in Eq. (25). The ratio of searching to matched firms in equilibrium is given by Eq. (26) or (27).

3.2 *Phase 2: The Globalising Economy*

Now assume that trade costs fall, so that upstream production in the South becomes more competitive. In particular, assume that unit variable costs, including trade costs, for a North-South pairing are C_S, as opposed to 1 for a North-North pairing ($C_S < 1$). However to enter the market, foreign firms of type u have to engage in a costly search process, and so will only do so at a World reference price of

$$P_2^* = P_1^* C_S^{\frac{\varepsilon-1}{\varepsilon}}, \tag{28}$$

at which a firm in a reservation-quality pairing breaks even.

The effect on existing North-North pairings is variable. The N_{S1} searching Northern pairings would not expect to break even at the new, lower price, and consequently will exit the market, except for a fraction who find their existing partners profitable and become matched.

The N_{m1} existing matched North-North pairings are heterogeneous, since they vary in match quality between μ_R and 1. The price at which a matched pairing of quality μ_i will exit the market is given by

$$P_i' = (\frac{F-\mu_i}{k})^{\frac{1}{\varepsilon}} = P_1^* (\frac{F-\mu_i}{F-\mu_R})^{\frac{1}{\varepsilon}}. \tag{29}$$

The last firm will exit when

$$P_1' = P_1^*(\frac{F-1}{F-\mu_R})^{\frac{1}{\varepsilon}}. \tag{30}$$

I assume North-South pairings are not sufficiently low-cost to drive the price down to this level.

At price P_2^*, the critical match quality at which existing matched North-North pairings which will withdraw is

$$\mu_i = F - (\frac{P_2^*}{P_1^*})^{\varepsilon-1}(F-\mu_R); \tag{31}$$

at which price the proportion of North-North pairings which will continue in the market is

$$\begin{aligned}\sigma_{N2} &= \frac{1-\mu_i}{1-\mu_R} = 1 - \frac{\mu_i - \mu_R}{1-\mu_R}, \\ &= 1 - \frac{(1-\overline{C}_S^{\varepsilon-1})(F-\mu_R)}{1-\mu_R}.\end{aligned} \tag{32}$$

This is declining as C_S falls. Output of these pairings will also be lower than before the start of globalisation by proportion $(\frac{P_2^*}{P_1^*})^{-\varepsilon}$, since the industry aggregate price has fallen. However, total industry output will be higher, due to the fall in prices: I denote the own-price demand elasticity for the industry's output as $-\eta$, so that

$$\frac{Y_2^*}{Y_1^*} = (\frac{P_2^*}{P_1^*})^{-\eta}. \tag{33}$$

The residual will be made up by new, North-South pairings. Initially, these will be searching, but over time, in a model with static overall demand, the proportion of these still searching will decline, as some gain matches, until eventually the ratio of searching to matched pairs among the North-South pairings will equal that in Eq. (25).

Figure 3 shows the situation in an economy during the globalisation phase. Searching North-South (foreign) pairings are prepared to enter the market at price P_2^*, which is lower than P_1^*, but assumed to be above P_1'. Consequently, all North-North (domestic) searching pairings exit the market, as do some matched North-North pairings. Once some North-South pairings find successful matches, the result is a supply curve with three segments (i.e. two kinks). First, between P_2' and P_1' (the left segment), only the most successful North-South pairings will be prepared to supply. Between P_1' and P_2^* we are summing horizontally the supply curves of

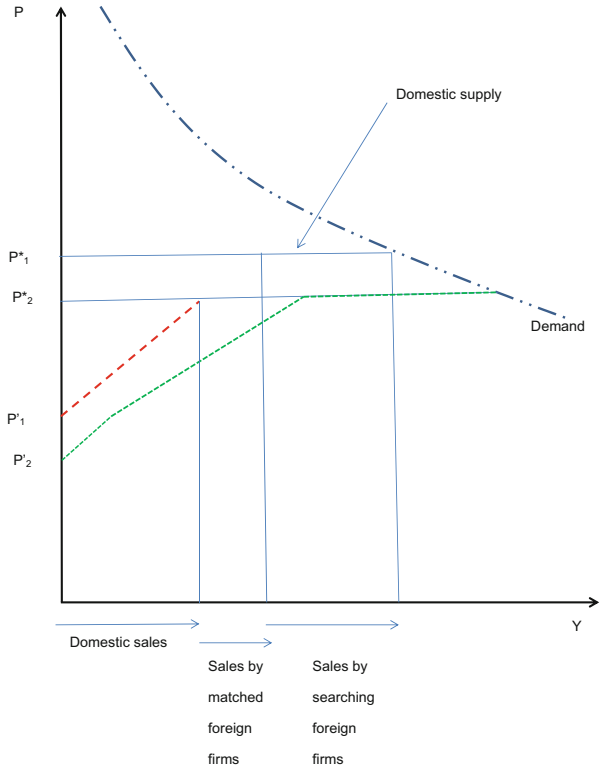

Fig. 3 Demand and supply in the globalising economy

North-North and North-South matched pairings, so the gradient is somewhat less steep. Finally, when the industry aggregate price, $P^* = P_2^*$, there is a horizontal segment made up of searching North-South pairings.

The kinks in this supply curve are an important element in explaining the differential effects of economic shocks, and the potential change in policy following such shocks—as explained in the subsequent sections.

Figure 4 summarises some simulations on the evolution of the proportions of different types of firm pairings over time, following a 10 % reduction in the cost of North-South pairings. These are based upon the parameter assumptions in Table 2.

Simulations are carried out for underlying demand growth rates of 2, 5 and 8 %. After 20 periods, the proportions of searching *NS* pairings vary between 15 and 30 % , depending on demand growth, while matched *NS* pairings are 56–62 % and *NN* pairings 8–29 % of the market.

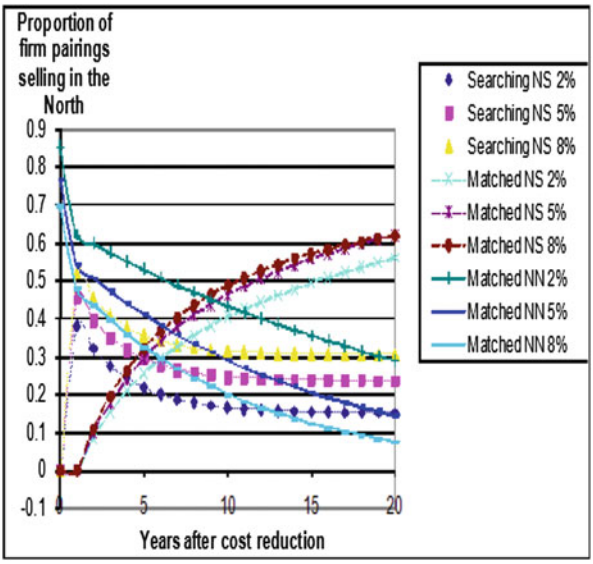

Fig. 4 Evolution of different types of pairings after a reduction in trade costs starts a globalisation phase, related to the underlying growth of demand per annum

Table 2 Parameter assumptions for simulations

Time preference per contract period	ρ	0.08
Death rate of one firm in pairing	D	0.02
Basic fixed cost of market participation	F	1
Subs elasticity between firms	ε	5
Own price elasticity for industry output	η	-1
Implied reservation match quality	μ_R	0.768

4 The Impact of a Negative Demand Shock

We now want to consider the effect of shocks to an economy which has been undergoing the process of globalisation (in the sense of entry by searching foreign upstream firms, and the gradual development of successful, importing partnerships). Take the situation in Fig. 3 and assume that there is a sudden inward shift in the demand curve, caused by a recession. The situation is shown in Fig. 5: essentially, searching foreign firms are footloose, and will simply exit the market at the end of their existing contracts (except for the minority who find profitable contracts). This is a component of import supply which is very sensitive to demand changes. Consequently, unless the demand shock is very large, it can be accommodated simply by the exit of these firms: the remaining foreign and domestic firms see very little change in their demand or prices. Only larger shocks will force established firm pairings back down their supply curves.

Note that, at the point where searching North-South pairings have been eliminated, the elasticity of imports with respect to the industry aggregate price falls from infinite to $\frac{\varepsilon}{\mu_R}$, and continues to rise thereafter, as more firms are driven out.

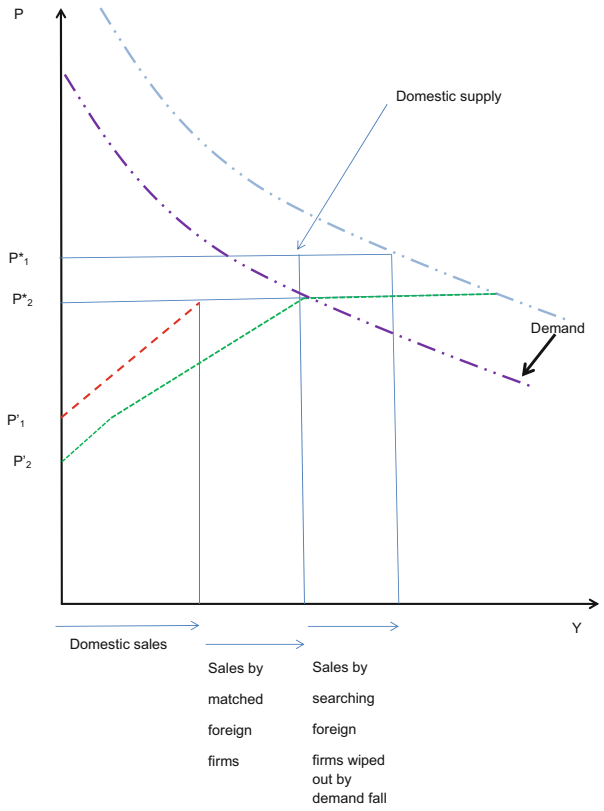

Fig. 5 A negative demand shock which wipes out searching firms, leaving matched firms unchanged

The situation with a larger negative demand shock is that, only after searching firms have been eliminated will prices be driven down. At this point, existing domestic pairings, as well as established importing pairings will be faced with falling prices, and the least efficient will be eliminated (so driving the firms down their supply curves).

Note that an implication is that the first firm pairings to exit the market are searching, North-South pairings, so that trade is especially vulnerable to a medium shock. In the larger shock in Fig. 4, matched North-North pairings will exit at the margin faster than matched North-South pairings in the ratio

$$\frac{\sigma_{N3}}{\sigma_{S3}} = \frac{P_2^* - P_1'}{P_2^* - P_2'} < 1. \tag{34}$$

One should perhaps not be surprised by this finding. Trade search can be viewed as a kind of capital formation, with firms prepared to undergo losses in the early

years of search, in order to make an expected positive return thereafter, once they are established. Like any form of capital formation, we would expect search capital formation to be sensitive, both to changes in the cost and availability of credit, and to accelerator-type changes in overall demand growth. That is what this model indicates.

5 Endogenous Policy Responses to a Boom, a Shock or Stagnation

We have so far established that, during a globalisation boom, where demand and imports are rising fast, there should be at any one time a relatively high proportion of trade which is accounted for by searching firms, whose presence in the market is footloose, and who are vulnerable to either demand and/or credit shocks. This, by itself, implies that the globalisation process can be suddenly halted by unexpected shocks, regardless of any policy response. Moreover, since search is necessary for the long-run growth of trade, a shock may have a prolonged negative impact upon trade.

We now want to consider how trade policy might respond to economic circumstances, given a search model, where various components of foreign supply differ greatly in terms of their price sensitivity. In this context, it is worth bearing in mind that, where trade is between firms who are specifically matched, all countries can be considered 'large' from the point-of-view of trade policy, since trade with a matched partner is differentiated from trade with any other potential partner, yielding potential monopsony power to the importing country (and monopoly power to the exporting country). We will follow a modified version of the standard 'large country, partial equilibrium' exposition of optimal tariff policy (see, for example, Van Marrewijk 2007, Sect. 8.3).

Figure 6 shows schematically the effects of imposing a tariff on an economy with a large amount of initial importers who are searching (and hence price-sensitive). It has been drawn slightly differently to Figs. 1, 2, 4, 5 and 6, since domestic supply is now subtracted from consumers' demand, to give a demand curve for imports. (Note that the domestic supply curve increases less quickly above P^*, since a rise in price does not affect the number of domestic firms, while a fall in price will cause further exit.) The elastic portion of the supply curve, accounted for by the initial searching importers, means that a tariff has to be substantial before it can begin to lower the import price. The welfare effects of this are that there is a rise in consumer prices (causing a welfare loss), and no fall in import costs. There is, however, some tariff revenue from the higher price paid by consumers (though there is still a deadweight loss). Since the model is monopolistically competitive, and prices exceed marginal costs, there is also a profit-shifting effect from the increase in market share borne by domestic producers (see Brander and Krugman, 1983). (This effect will only be

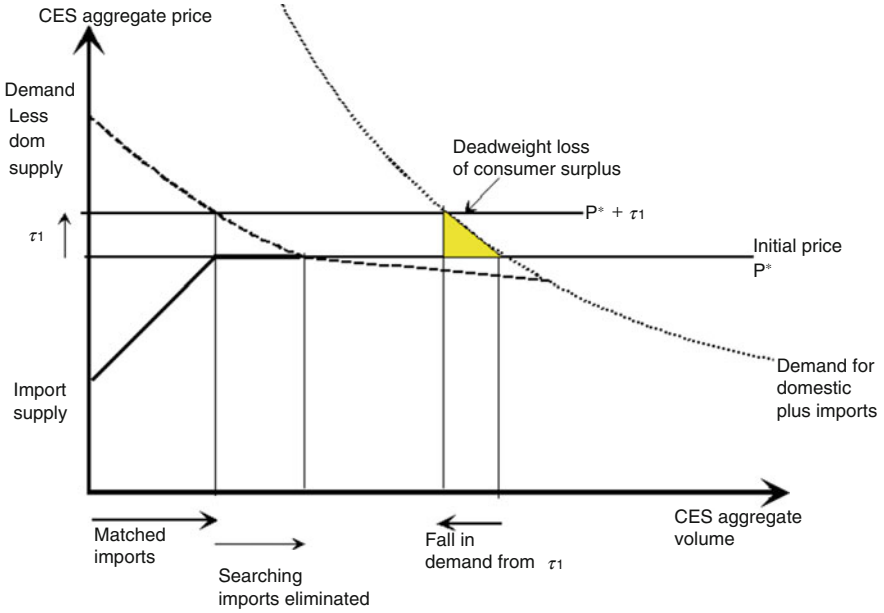

Fig. 6 Effect of a low tariff, τ_1

enough to outweigh the loss of consumer surplus when the elasticity of substitution, ε, is low enough to generate high profit margins, yet still large enough for domestic sales to increase significantly.)

Figure 7 shows a larger tariff, which begins to lower import prices (cf. the large, open economy partial equilibrium model in many trade textbooks). The fall in import prices (by driving out the less efficient importers) means that some of the tariff revenue is at the expense of importers rather than consumers. However, the presence of the flat segment (the threshold effect) means that this effect (when combined with the profit-shifting effect) needs to be large to offset the loss of consumer surplus, since only sizeable tariffs will begin to reduce traded prices.

By contrast, where there are few initial searching firms, a tariff begins much more quickly to impact upon traded prices. The implication is that there is likely to be a discontinuity in tariff-setting: when the sector is shrinking, or growing slowly, optimal tariffs are positive, while, when the sectoral rate of import growth exceeds a threshold rate, optimal tariffs fall to zero.

The analysis here is somewhat simplified, since in reality one needs to consider tariffs in a dynamic setting. In due course, even with a tariff, the natural death of existing matched firms will lead the economy back to a position where searching firms begin to reenter the market—under some circumstances, we would expect this to lead to a cycle of optimal tariffs over time. Nevertheless, the situation with stalled growth indicates a significant difference between eras of prolonged trade growth, where protection is not favoured, and eras of stagnation, where it is.

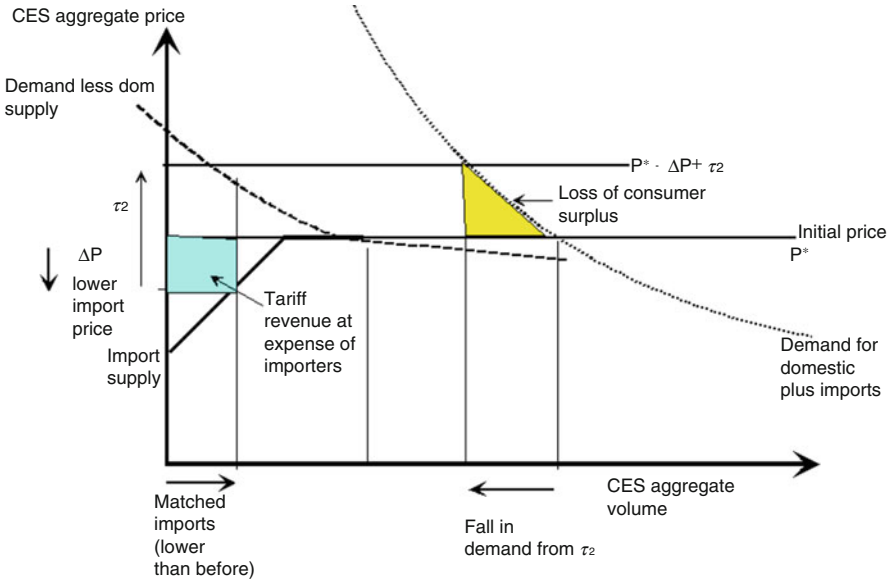

Fig. 7 Effect of a higher tariff, τ_2

Table 3 Parameters for numerical simulation

Discount per contract period	ρ	0.08
Fixed cost	F	1
Elasticity of substitution between firms	ε	10
Share of non-trading firm pairs (base)	S_D	0.5
Death rate of firms per contract period	D	0.02
Marginal cost of *NN* pairing	C_N	1
Marginal cost of *SN* pairing	C_S	$\frac{1}{1.1}$
Total expenditure on the product	M	1000

5.1 Numerical Simulation Analysis

To examine the effects of a crisis and/or a period of stalling of growth, I set up a numerical simulation of the partial equilibrium model outlined in Figs. 6 and 7. The model is static only. Key parameter values chosen are shown in Table 3.

An elasticity between firms of $\varepsilon = 10$ is consistent with many recent studies of trade elasticities (Anderson and van Wincoop, 2004).

The model is calibrated to a series of equilibrium starting growth paths for a variety of growth rates between -2 and $+9\,\%$ per annum. The former case corresponds to the situation where $G + D = 0$, and so is a situation where in equilibrium there are no searching firms. The higher the initial growth rate, the more searching firms there will be.

To summarise the results of the simulations:

Given the parameter assumptions, a tariff that eliminates searching firms reduces national welfare. This is because, while the tariff is not large enough to remove more than the price-sensitive, searching component of supply, import prices do not fall. Consequently, the tariff revenue plus profit-shifting are insufficient at the margin to outweigh the loss of consumer surplus. The results are shown in Table 4.

Note that, where the elasticity of substitution is lower, or the initial share of domestic production is higher, then the profit-shifting effect will be larger, and may outweigh the loss of consumer surplus.

Driving searching firms out is costlier to welfare, the higher is the initial growth rate of trade. This should not be seen as surprising, since a higher tariff is required when there are more searching firms to start with.

Once searching firms have been driven out, a further rise in the tariff first increases and then reduces welfare. It is worth remembering that, once searching firms have been driven out, the analysis is closer to the typical large country model of trade, where a tariff is initially welfare-improving, by reducing import prices. Only later do the marginal costs of a tariff outweigh the benefits.

This result is important in the immediate aftermath of a crisis, where a fall in demand or a collapse of credit have resulted in the exclusion of searching firms. In this case, there are no initial searching firms, and welfare will always benefit from some tariff (in this case, reaching a non-cooperative 'optimal tariff' of around 4 %, reducing trade by around 32 % in total).

Whether or not a tariff can improve national welfare depends on the initial number of searching firms. The faster the initial growth rate of the economy, the more searching firms have to be driven out before a tariff can start to improve the country's terms of trade. Consequently, the inverted-U curve of welfare starts from a lower point. Beyond some threshold rate of growth (in this case, approximately

Table 4 Effect of steady-state growth rate on share of searching firms, tariff required to exclude them and associated national welfare loss

Growth rate (%)	Share of searching firms (%)	Tariff to exclude searchers (%)	Net welfare loss
−2	0.0	0.0	0.0
−1	4.1	0.3	0.107
0	7.9	0.6	0.243
1	11.5	0.9	0.401
2	14.7	1.2	0.574
3	17.8	1.4	0.760
4	20.6	1.6	0.954
5	23.2	1.8	1.154
6	25.7	1.9	1.357
7	28.0	2.1	1.562
8	30.2	2.2	1.768
9	32.2	2.4	1.973

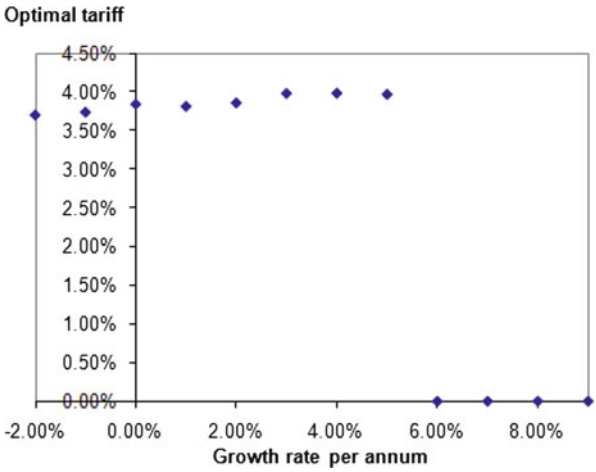

Fig. 8 Optimal tariff related to growth rate of trade, $\rho = 0.08$, $\varepsilon = 0.1$, $D = 0.02$

$G = 5\,\%$), a country will switch from charging an 'optimal tariff' to free trade. This is also summarised in Fig. 8, which shows the simulated 'optimal tariff' for different initial growth rates:

For growth rates below about 5 % per annum, countries will choose noncooperative tariffs of around 4 % on trade. By contrast, above this threshold growth rate, there is a sudden, nonlinear shift to a zero noncooperative tariff, as countries prefer free trade, in order to attract the more footloose, searching firms.

Again, this result is dependent upon the substitution elasticity assumed, and upon the initial import share (a lower substitution elasticity and/or lower import share would raise the growth threshold needed to make tariffs unattractive, while a higher substitution elasticity or higher import share would do the reverse).[16]

The evidence that there is a threshold growth rate required to induce countries to maintain open markets is significant. As long as $G + D > 0$, new firm entry into trade will eventually resume following a crisis. However, many economic shocks—such as Latin America in the early 1980s, Japan in the early 1990s and the Asian Tigers in 1997—tend to be followed by either stagnation (a 'lost decade') or at least reduced growth for a prolonged period. If growth falls below a critical trigger, the analysis of this section suggests that open trade policies may be harder to sustain.

5.2 Is Search Sufficient to Trigger a Policy Switch?

While the model in the preceding section is somewhat sensitive to parameter choices, it does indicate certain important properties: in particular, a prolonged

[16] See author for details.

period of high global trade growth will favour little or no protectionism, but only if the growth rate exceeds a critical threshold (maybe around 5 % per annum).

The Eaton et al. (2007) estimates of search would suggest values (during the growth cycle) quite close to that required on our parameter assumptions to trigger a switch to zero protectionism. This may indicate that countries would be likely to favour free trade in faster-growing industries, while maybe not in others. Following the crisis of 2008–2009, it is likely that few industries would still have enough search to make unilateral free trade attractive.

It is worth bearing in mind that, in a CES demand formulation, the profit-shifting effect (which makes tariffs more attractive) is higher in industries with a lower import share.

6 Implications: Return of the Prisoner's Dilemma in Trade Liberalisation?

Trade liberalisation is often portrayed as a prisoner's dilemma situation: liberalisation benefits countries collectively, but large countries, individually, have an incentive to cheat and impose optimal (terms-of-trade-improving) tariffs. By contrast, in recent decades (and despite the failure of the Doha Round talks), trade liberalisation has undoubtedly made substantial strides—particularly with respect to developing countries. In general, the argument that high tariffs benefit national income—at least, when based upon terms of trade arguments—has rarely been used. An exception may be in agricultural commodities.

Part of the reason may be a trend towards accepting that the price elasticity of demand for traded commodities is perhaps higher than we used to think—for example, Anderson and van Wincoop's (2004) survey cites elasticities of between 5 and 10 for many commodities: a far higher number than traditionally used in Armington CGE models of the 1980s and 1990s.[17] These higher trade elasticities have been theoretically reconciled with observed trade data by the acceptance of substantial Ricardian, fixed-factor elements in trade,[18] including firm heterogeneity, which mean that supply elasticities are reduced. However, a search-based interpretation would say that supply elasticities differ considerably *ex ante* and *ex post*, and that expanding supply may be more elastic than contracting supply.

A prolonged global recession could potentially change this situation for the worse (as memories of the 1930s indicate).[19] In terms of the very schematic analysis of this paper, we are moving from the situation in Fig. 6, where there is a substantial volume

[17] Anderson and Winters (2008) provide an excellent discussion of the effects of moving from traditional to more modern CGE models in assessing trade policy.

[18] Eaton and Kortum (2002).

[19] Eichengreen and O'Rourke (2009) summarise the trade experience of the current economic crisis, with a roughly 20 % fall in World trade (April 2008 to a year later).

of footloose trade, to that in Fig. 7, where the supply of price-elastic, searching trade is insufficient to maintain the incentive to keep tariffs low. In such a situation, the interests of individual countries (which may wish to try to manipulate their terms of trade by protection) and those of the global economy are no longer aligned, hence raising the risks of triggering a round of beggar-thy-neighbour policies.

In sum: the pre-2008 wave of globalisation was sustained and strengthened by the coming-together of national and global interests, when trade is relatively footloose and credit is available. It seems that, since 2008, the advice of economists to learn from history and avoid repeating the mistakes of the 1930s has largely been heeded. However, a resumption of growth is probably the best insurance against a resurgence of protectionist forces.

References

Anderson, J. E., & van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature, 42*(3), 691–751.

Anderson, K., & Winters, L. A. (2008). The Challenge of Reducing International Trade and Migration Barriers. CEPR Discussion Papers 6760, C.E.P.R. Discussion Papers.

Balrwin, R., & Evenett, S. (2009). The collapse of global trade, murky protectionism, and the crisis: Recommendations for the G20. CEPR. London: VoxEU.

Bernard, A. B., Redding, S. J., & Schott, P. K. (2007). Comparative advantage and heterogeneous firms. *Review of Economic Studies, 74*(1), 31–66, 01.

Besedes, T., & Prusa T. J. (2006). Ins, outs and the duration of trade. *Canadian Journal of Economics, 39*(1) 266–295.

Bhagwati, J. N. (2004). In *Defense of globalization*. Oxford: Oxford University Press.

Brander, J., & Krugman, P. R. (1983). A reciprocal dumping model of international trade. *Journal of International Economics, 15*, 313–323.

Crafts, N. (2004). Globalisation and economic growth: A historical perspective. *The World Economy, 27*(1), 45–58.

Dollar, D., & Kraay, A. (2002). Growth is good for the poor. *Journal of Economic Growth, 7*, 195–225.

Dollar, D., & Kraay, A. (2003). Institutions, trade and growth. *Journal of Monetary Economics, 50*(1), 133–162.

Eaton, J., Eslava, M., Kugler, M., & Tybout, J. (2007). *Export Dynamics in Colombia: Firm-Level Evidence*. Borradores De Economia 003957, Banco De La Republica.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica, 70*(5), 1741–1779.

Edwards, T. H. (2010). Globalisation as a 'Good Times' phenomenon: A search-based explanation. *Economics: The Open-Access, Open-Assessment E-Journal, 4*, 2010–2024. http://dx.doi.org/10.5018/economics-ejournal.ja.2010-24

Edwards, T. H., & Ferrett, B. (2009). *The Search for Trading Partners and the Cross-Border Merger Decision* Discussion Paper Series 2009_15, Loughborough University (on REPEC).

Eichengreen, B., & O'Rourke, K. H. (2009). *A Tale of Two Depressions, commentary on Voxeu.org*. http://www.voxeu.org/index.php?q=node/3421.

Feenstra, R., & Hanson, G. H. (1999). The impact of outsourcing and high-technology capital on wages: Estimates for the U.S., 1972–1990. *Quarterly Journal of Economics, 114*(3), 907–940.

Gros, D. (1987). A note on the optimal tariff, retaliation and the welfare loss from tariff wars in a framework with intra-industry trade. *Journal of International Economics, 23*(3–4), 357–567.

Grossman, G. M., & Helpman, E. (1992). *Innovation and growth in the global economy*. Cambridge, MA: MIT Press.

Grossman, G. M., & Helpman, E. (2002). Integration versus outsourcing in industry equilibrium. *Quarterly Journal of Economics, 117*(1), 85–120.

Grossman, G. M., & Helpman, E. (2005). Outsourcing in a global economy. *Review of Economic Studies, 72*(1), 135–159.

Hart, O. (1995). *Firms, contracts and financial structure. Clarendon lectures in economics*. Oxford: Clarendon Press.

Ishii, J., & Yi, K-M. (1997). *The Growth of World Trade*. Research Paper, Federal Reserve Bank of New York.

Krugman, P. K. (1979). Increasing returns, monopolistic competition and international trade. *Journal of International Economics, 9*(4), 469–479.

Luebker, M., Smith, G., & Weeks, J. (2002). Growth and the poor: A comment on Dollar and Kraay. *Journal of International Development, 14*(5), 555–571.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, Econometric Society, 71*(6), 1695–1725.

Mortensen, D. T., & Pissarides, C. A. (1999). New developments in models of search in the labor market. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (1th ed., Vol 3, Chap 39, pp. 2567–2627). Amsterdam: Elsevier.

Obstfeld, M., & Rogoff, K. (2000). The six major puzzles in international macroeconomics: Is there a common cause? In *NBER Macroeconomics Annual, 15 (2000)*. Cambridge, MA: MIT Press.

Rauch, J. E., & Casella, A. (2003). Overcoming informational barriers to international resource allocation: Prices and ties. *Economic Journal, 113*, 21–42.

Rauch, J. E., & Trindade, V. (2003). Information, international substitutability, and globalization. *American Economic Review, 93*, 775–991.

Schmitz, H. (1999). From ascribed trust to earned trust in exporting clusters. *Journal of International Economics 48*(5), 139–150.

Spencer, B. J. (2005). International outsourcing and incomplete contracts' (presidential address, Canadian Economic Association). *Canadian Journal of Economics, 38*(4), 1107–1135.

Srinivasan, T. N., & Archana, V. (2009). *India in the Global and Regional Trade: Determinants of Aggregate and Bilateral Trade Flows and Firms' Decision to Export*. Indian Council For Research on International Economic Relations Working Paper No. 232. ICRIER, Delhi.

Trefler, D. (1995). The case of the missing trade and other mysteries. *American Economic Review 85*(5), 1029–1046.

Tybout, J. (2002). Plant- and firm-level evidence on new trade theories. In J. Harrigan (Ed.), *Handbook of international economics*, Vol 38. Oxford: Basil Blackwell.

Van Marrewijk, C. (2007). *International economics*. Oxford: Oxford University Press.

Wacziarg, R., & Welch, K. H. (2008). Trade liberalization and growth: New evidence. *World Bank Economic Review, 22*(2), 187–231.

Winters, L.A., McCulloch, N., & McKay, A. (2004). Trade liberalization and poverty: The evidence so far. *Journal of Economic Literature, 42*(1), 72–115.

Yi, K-M. (2003). Can vertical specialization explain the growth of world trade? *Journal of Political Economy, 111*(1), 52–102.

Part II
Multinational Firms

Dynamic Capabilities and the Multinational Enterprise

David J. Teece

Abstract Contemporary theories of the MNE have made limited progress in explaining interfirm heterogeneity. A fuller explanation should account for the uncertainties that firms face about market demand and about the costs of competing organizational alternatives. One approach is to allow a role for entrepreneurial managers who assess uncertainties on an ongoing basis and make non-routine decisions that create unique paths for the organizations they lead. This paper presents the dynamic capabilities framework, a multidisciplinary approach to analyzing the characteristics of individual firms and the sources of firm-level competitive advantage. Dynamic capabilities encompass the non-routine actions of entrepreneurial managers and the creative output of a firm's expert talent. These capabilities are also embedded in "signature" organizational routines and processes rooted in an organization's unique history. A dynamic capabilities approach supports a richer understanding of the distinctive characteristics of MNEs. A case study is presented that shows how the exploration division of a large multinational oil firm introduced capability awareness into the management of its global staff of geophysicists.

1 Introduction

A multinational enterprise (MNE) is a business organization that sets strategy, manages operations, and generates sales in more than one country in pursuit of profits. Any theory of the MNE is closely related to the theory of the business enterprise more generally. A robust theory of the business enterprise should be able to help us understand not only its general nature but also its geographical scope and the basis of its sustained competitive advantage, if any.[1]

[1]By sustained competitive advantage, I mean the ability of a firm to earn better than its cost of capital over the longer run.

D.J. Teece (✉)
Haas School of Business, University of California, Berkeley, CA, USA
e-mail: teece@haas.berkeley.edu

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_5

Until recently, economic theories of the firm, whether multinational or otherwise, have not recognized, much less explained, interfirm heterogeneity. Hence, they cannot address issues relating to the formation of firm-level competitive advantage. This is a consequence of the exclusion of capabilities, differential technologies, entrepreneurship, and leadership from most of these models.

Economic theory often assumes markets are homogeneous. This isn't true of national markets, and the degree of heterogeneity in markets that MNEs confront is almost always very high. This is, in part, because the needs and desires of consumers in different countries have not (yet) become irrevocably homogenized. It is also because labor is not fully mobile, and many national institutions, regulations, and policies, remain distinct and home-country bound. The world is, and is likely to remain, only semi-globalized (Ghemawat 2003). Put differently, the world is not by any means completely "flat."

The reality of firm-level heterogeneity and semiglobalization, coupled with relentless change in markets and technology, animate the dynamic capabilities framework. The dynamic capabilities approach to understanding the MNE (Teece 2014a) provides a multidisciplinary perspective on the nature of global corporations and the management activities required to produce sustained competitive advantage and supernormal profits. Dynamic capabilities encompass the distinct routines and culture that characterize an organization, and also the non-routine actions of entrepreneurial managers and expert talent that help organizations grow, adapt, and prosper. They undergird a theory of heterogeneous firms.

The paper begins with a review and critique of the leading economic theories of the MNE. This is followed by an introduction of the dynamic capabilities framework, including how capabilities interact with strategy to impact firm performance. While dynamic capabilities are supported by distinctive organizational routines, the non-routine actions of entrepreneurial managers and expert talent are also analyzed, followed by a case study of how the exploration division of a large multinational oil and gas company applied a capabilities approach to the deployment of its staff of scientists. A concluding section summarizes the key points.

2 Contemporary Theories of the MNE

For at least three decades, the theoretical literature on the MNE in economics and business has been largely dominated by the internalization perspective (Dunning and Lundan 2008). This approach focuses on "market failure" considerations to explain why firms internalize transactions (i.e., invest directly rather than use arm's-length contracts) across national borders.

This approach has its roots in a (non-multinational) analysis by Coase (1937), who offered a simple efficiency condition: firms will internalize transactions when the anticipated bureaucratic cost of doing so is less than the anticipated cost of contracting. Williamson (1981, 1985) deepened understanding of the economic foundations of market failure and associated contractual issues. He also developed

the concept of asset specificity, which, when combined with the possible opportunism of the parties involved, helps explain what he called "recontracting hazards," which, in turn, are the essence of Williamsonian "transaction costs." Elements of the Coase-Williamson framework were applied to the multinational enterprise (MNE) by Buckley and Casson (1976), Dunning (1981), Rugman (1981), Teece (1981a, 1986a), Williamson (1981), and others.

Foreign direct investment in a corporate subsidiary is efficient, in the Coase-Williamson view, when contracting with independent agents to effectuate the same goals would be expected to lead to recontracting hazards. For example, if the overseas activity involves investing in transaction-specific assets that cannot readily be redeployed to other uses, then the MNE relying on an overseas contractor to make the dedicated investment would be exposed to the possibility of future unfavorable renegotiation. Direct foreign investment may increase other risks, but, in the Coase-Williamson formulation, recontracting risks are paramount.

Recent economic research on international trade (e.g., Antràs and Helpman 2004) has adopted a slightly different version of contracting, namely, the property rights approach of Grossman and Hart (1986). Agency theory has also been used (e.g., Grossman and Helpman 2004). The contractual model is coupled with the heterogeneous firms approach, introduced by Melitz (2003), to analyze trade, investment, and internalization decisions in a formal general equilibrium model (see Spencer 2005, for a review). However, in much of this work, firms are depicted as investing and competing in markets with known demand characteristics. And the relative costs of various organizational configurations (e.g., integration versus outsourcing) are usually assumed to be fixed and known, so that each firm need only apply a formula to find the cost-minimizing organizational configuration for its productivity level. In short, these models assume away, for the most part, any role for entrepreneurial managers or for capability formation.

This recent work, like the Coase-Williamson view of the MNE that still dominates international business research, falls short in at least two critical dimensions to be discussed in more detail. First, it fails to recognize the entrepreneurial role that MNEs and their managers play in creating and co-creating new markets for products and services that did not previously exist because supply and/or demand were nonexistent in host countries until they were generated by the MNE.[2] Second, it takes a simplistic view of control, implicitly equating control with internalization, a view that is no longer applicable in today's world of tightly coordinated global supply chains.

[2] Vernon's (1966) theory of trade and investment over the product cycle comes close to recognizing market creation because it posits that multinationals invest in offshore production as their product technology matures. However, it hypothesizes that offshore plants will go into markets that are already importing the goods, which, like Coase-Williamson theories, skips over the market creation function of the MNE.

2.1 Entrepreneurship and Market Co-creation Suppressed

In the MNE context, the internalization question is framed by most international business scholars as one of several "market entry mode" choices. Alternative ways to enter a foreign market include supply contracts, joint ventures, and wholly owned subsidiaries. In recent decades, MNE theory has put much emphasis on these alternative market entry modes (e.g., Brouthers 2013; Hennart 2009; Zahra et al. 2000). The underlying assumption is that markets sometimes "fail" to be the efficient form of organization, not in the sense that government action is called for, but in the sense that internal organization ("integration") is required to help solve an economic problem that is generally delimited as a transaction cost issue.

This contract-oriented "market failure" paradigm ignores the even more significant problem that markets often quite simply fail to emerge and need to be created or co-created by entrepreneurially managed business enterprises (Pitelis and Teece 2010). The entrepreneurial functions of market creation and co-creation have always been fundamental activities of the MNE. Organizational capabilities are integral to the market creation and co-creation process. This type of market "failure," i.e., the very lack of an (intermediate) market, is completely ignored in the conventional Coase-Williamson formulation.

Some consideration of market creation is already present in Casson's important work on entrepreneurship (1982, 1997, 2005). However, market-making in his theory is rather neoclassical and overly focused on individual action. In particular, Casson's approach does not seem to recognize the importance of the capabilities the enterprise and its management may have to shape markets, influence trends, shape demand, and assemble the complements needed for new markets to be viable. Because the market for information/knowledge about new opportunities isn't well developed (Gans and Stern 2010; Teece 1981b), entrepreneurs and managers must build organizational capabilities for the creation and application of such knowledge on a global scale. The neglect of these entrepreneurial/managerial functions in the theory of the firm would appear to be a serious omission.

2.2 Decoupling of Control from Ownership Overlooked

Another weakness in some theories of the MNE is that they implicitly assume, or explicitly state (e.g., Hennart 2010), that strategic control comes through ownership and resides with the parent, i.e., ownership is both necessary and sufficient to achieve the alignment of goals. This view of the MNE tends to see wholly-owned subsidiaries as the preferred organizational form when firm-specific assets need to be protected from recontracting hazards.

In reality, the mere ownership of a business unit doesn't eliminate incentive problems, nor does it necessarily achieve control. This is particularly true in the MNE context, where the alignment of host country managers with the best interests

of the MNE shareholder is less than perfect. This is well recognized by international business scholars, but most MNE theories struggle to find an elegant way to take it into account.

In an earlier time, integration was often required to achieve coordination (Chandler 1977). However, cross-border communication has improved dramatically in recent decades, and supplier capabilities have grown dramatically. As a result, coordination in the supply chain seems to require less internalization than it once did.

Consider Apple, which is known for its hardware design prowess. It owns none of the manufacturing plants that manufacture its electronics goods to demanding specifications. It has tight supply relations with many companies, some pure contractors (e.g., Foxconn, headquartered in Taiwan with factories in China) and at least one competitor (Samsung, headquartered in Korea). Apple helps provide financing to some of its suppliers and may obtain exclusive purchase arrangements from them for short periods (e.g., 3 years). It may even purchase some of the equipment to be used in the factories. Contractual arrangements appear to suffice for Apple to achieve the necessary coordination, while retaining the flexibility needed to respond to market forces thanks to a low level of commitment to dedicated plant and equipment.

These outsourcing arrangements have been studied under several names, including "international production networks" (e.g., Ernst and Guerrieri 1998), "global value chains" (e.g., Gereffi et al. 2005), and the "global factory" (Buckley 2007). Supply networks put great emphasis on both the operational ("supply chain") capabilities of the buyer and the production capabilities of the supplier.

As I discuss in the next section, the reason for internalizing the supply chain and other aspects of an overseas operation might not be control, in the sense of oversight, but rather the ability to capture learning benefits.

3 A Capabilities Framework for the MNE

A different form of market "failure" that involves resource transfer costs and learning issues has been less well developed in the literature.[3] Problems of this nature are addressed by the capabilities view of the MNE. Recent work in international business (some of it generated by advocates of the transaction cost paradigm) has already begun to focus on an entrepreneurial/capabilities approach, albeit with widely varying degrees of engagement (e.g., Buckley 2009; Casson 2000, 2005; Dunning and Lundan 2010; Pitelis 2004; Pitelis and Teece 2010; Rugman and Verbeke 2003).

[3] As before, failure is benchmarked against a hypothetical world of complete, fully-contingent markets and perfect competition. Accordingly, the "failure" discussed is not policy-relevant and does not require government intervention.

Resource transfers are generally less costly to effectuate when they occur inside the MNE (Teece 1975). Overseas resource transfers between independent parties are inherently more complex and costly to contract for and to execute, although the cost difference diminishes as a technology becomes more standardized and/or is re-engineered for modularity (Sturgeon 2002). Internal transfers and learning are facilitated by the common (organizational) culture of the enterprise and the ease of coordination inside the firm relative to coordination through the market. Worries about the leakage of valuable knowledge are mitigated and future benefits from experiential learning have a better chance of being captured by the firm itself. Internalization also simplifies the interchange of personnel across borders, which is often a necessary accompaniment to resource transfers.

Recognition of the knowledge transfer efficiency and appropriability properties of the MNE has given rise to a broader, knowledge-based theory of the evolution of the MNE. In these formulations, the expansion of the enterprise requires the creation of new knowledge as part of an ongoing, social process that reflects the firm's unique history and resource set (Kogut and Zander 1993). Learning is a touchstone of the capabilities approach to the MNE.

The learning envisaged in the capabilities approach is both broader and more significant than mere technological accumulation, such as that discussed in Cantwell (1989). In fact, given the ever-greater global dispersion of technology, reliance on in-house R&D as the sole basis of technology development is no longer tenable. Technologies from both within and beyond the enterprise must be effectively orchestrated to achieve timely delivery of differentiated products and services that customers value (Augier and Teece 2007; Pitelis 2004). This involves asset orchestration capabilities, one of the mainstays of the dynamic capabilities framework.

In the dynamic capabilities framework, top management at headquarters performs a global asset orchestration function, allocating the financial resources needed for the MNE divisions to create markets outside of the home jurisdiction and leaving operational matters to lower levels of the organization. The headquarters can further enhance the firm's capabilities by allowing and facilitating technology transfers amongst the divisions and by encouraging and supporting the exploitation of complementarities.

Subsidiaries also play a vital role in the firm's dynamic capabilities. They can generate know-how and capabilities from their own histories that can be transferred to other business units at home or abroad. This may not have received much attention in the economic view of the MNE, although it's a significant component of the international management literature (e.g., Birkinshaw 2000).

As this suggests, the MNE need not be rigidly hierarchical. Rather, it can function as a network, within which valuable new knowledge can arise anywhere for potential application everywhere. This is reflected in the "transnational" model of Bartlett and Ghoshal (1989). In the dynamic capabilities framework, the MNE is viewed as a portfolio of distinct resources that can be centrally orchestrated and still maintain channels for knowledge sharing in all directions within the network.

In the sections that follow, I provide definitions and illustrations of key concepts, including resources, ordinary capabilities, dynamic capabilities, and strategy.

3.1 Definitions and Core Building Blocks

3.1.1 Resources

Resources are the tangible and intangible assets, broadly defined, that the firm can develop and control. They include the skills of the firm's employees, its equipment, and the collective skills of the organization. Penrose (1959) notes that the services of most "resources" can be applied to multiple lines of endeavor. At any point in time, a firm is likely to have more resource services available than are needed for the output that it is currently producing. A firm with excess resources may find it more profitable to monetize those services via a new use (i.e., product diversification) rather than through a market transaction that leases access to the surplus services to an independent party (assuming such a transaction would even be feasible).

Despite Penrose's abiding interest in the international firm, she did not pay particular attention to the application of her theory to the case of the MNE (Pitelis 2007). Moreover, although Penrose did recognize the importance of entrepreneurship, she did not fully address the roles of entrepreneurs in designing business models or organizations (Augier and Teece 2007). Rather it was Wernerfelt (1984) and Barney (1986, 1991) who attempted to articulate the relationships between firm resources and competitive advantage. They focused on the possession of the right resources as the main mechanism for the generation of economic rent. The resources most likely to support sustainable competitive advantage were defined by Barney (1991) as those meeting his criteria of valuable, rare, inimitable, and non-substitutable (VRIN).

The most important class of VRIN assets is that of intangibles, or what might be referred to as intellectual capital (Teece 2000). Ownership (or control) of intangibles and their complements allows innovating firms to differentiate and establish some degree of competitive advantage. Important classes of intangibles include technological knowledge, patents, trade secrets, brands, customer relationships, business models, and organizational know-how.

A capability is the capacity to utilize resources to perform a task or an activity. The organizational know-how of the firm and its ability to transfer technology (embedded in routines and resources) across distances and borders are very much implicated in the MNE's capabilities.

3.1.2 Ordinary Capabilities

Ordinary capabilities, which encompass operations, administration, and governance of the firm's activities, make a firm capable of producing and selling a defined (and static) set of products and services. They support technical efficiency in performing a fixed group of activities, regardless of how well- or ill-suited the outputs are to the firm's competitive needs (Teece 2007: 1321). Routines for new product

development, quality control, knowledge transfer, and performance measurement are clusters of activities that can be part of a firm's ordinary capabilities.

Routines, however, are never the whole story where creative activities such as product development are concerned. At Apple, for example, product development is a mixture of creativity and routines. The late Steve Jobs described this in an interview (Burrows 2004):

> . . . there is no system. That doesn't mean we don't have process. Apple is a very disciplined company, and we have great processes. But that's not what it's about. Process makes you more efficient. But innovation comes from people meeting up in the hallways or calling each other at 10:30 at night with a new idea, or because they realized something that shoots holes in how we've been thinking about a problem. It's ad hoc meetings of six people called by someone who thinks he has figured out the coolest new thing ever and who wants to know what other people think of his idea. And it comes from saying no to 1000 things to make sure we don't get on the wrong track or try to do too much. We're always thinking about new markets we could enter, but it's only by saying no that you can concentrate on the things that are really important.

In some cases, the resources (assets, practices, etc.) involved in ordinary capabilities will meet the VRIN criteria. But, more often, the practices can be diffused and, hence, quickly adopted by competitors (Teece 2014a).

Ordinary capabilities can be manipulated and reshaped by a firm's dynamic capabilities. In particular, the organization's ordinary capabilities are continuously developed, augmented, combined, and winnowed using dynamic capabilities. This is necessary to align the firm with the requirements of its strategy and to assist it in achieving evolutionary fitness (Helfat et al. 2007: 7).

3.1.3 Dynamic Capabilities

Dynamic capabilities, which can be weak or strong, govern the capacity of the enterprise to sense opportunities and to profitably build and renew resources and assets that lie both within and beyond its boundaries, reconfiguring them as needed to innovate and respond to (or bring about) changes in the business environment (Teece et al. 1997; Pisano and Teece 2007). They reside in the culture and practices of the organization and in the cognitive skills of its managers. Dynamic capabilities govern how the enterprise and its top management develop conjectures about the evolution of consumer preferences, business problems, and technology; validate and fine-tune them; then act on them by realigning assets and activities so as to enable continuous innovation and change. Dynamic capabilities cannot generally be bought and therefore tend to get "built" over time. They are difficult for rivals to imitate.

Strong dynamic capabilities, by definition, enable firms to orchestrate their resources effectively in harmony with their strategy. The real competitive strengths of a firm lie in dynamic capabilities and in strategy formulation and execution, underpinned by value-enhancing "signature" processes and other VRIN resources that the firm and its management can command. Signature processes embody positive elements of a company's unique history, experience, culture, and creativity (Gratton and Ghoshal 2005).

Ordinary capabilities can, in most cases, be imitated with some effort, whereas dynamic capabilities, by their very nature, are hard to replicate. The imitability of organizational processes is a key parameter in the dynamic capabilities framework (Teece et al. 1997). Replicability (i.e., imitability by the firm itself, sometimes referred to as scalability) is often a source of advantage, particularly for companies seeking to expand abroad, whereas imitability by rivals can undermine advantage. A resource or capability that is easily imitated by others will clearly not be able to support superior financial returns.

An exception applies to MNE subsidiaries in certain contexts. If competition is suppressed by governmentally or privately imposed entry barriers, or by weak physical and social infrastructure that prevents ordinary capabilities from quickly diffusing throughout the economy, then strong ordinary capabilities (but weak dynamic capabilities) may be adequate to support long-run competitive advantage.

3.2 Applying Dynamic Capabilities

Dynamic capabilities can usefully be disaggregated into three clusters of processes and entrepreneurial/managerial orchestration activities (Teece 2007): (1) identification and assessment of opportunities at home and abroad (*sensing*), (2) mobilization of resources globally to address opportunities and to capture value from doing so (*seizing*), and (3) continued renewal (*transforming*). These activities are required of the firm's management if the firm is to sustain itself as markets and technologies change.

Sensing involves exploring technological possibilities, probing markets, listening to customers, and scanning the business environment. It requires management to be entrepreneurial and to build hypotheses about market and technological evolution, with an emphasis on identifying unmet needs. Sensing activities must be embedded throughout the company, with knowledge flowing from the farthest reaches of the organization to the top management team.

Seizing occurs after opportunities are properly sensed and calibrated. To seize opportunities, the firm deploys resources to create and capture value from discoveries, inventions, or innovations in which top management has identified the greatest promise. The capabilities behind seizing involve identifying, establishing control or influence over, then coordinating complementary assets, for example, by building a global supply chain, establishing alliances and joint ventures, and much more.

A key "seizing" activity for top management is development of business models that will be hard for rivals to imitate. A business model captures the logic of a how a business creates and delivers value to customers while earning a profit for itself (Chesbrough and Rosenbloom 2002; Teece 2010). Business models typically require adjustment or even wholesale transformation, perhaps repeatedly, especially in unsettled markets with lots of entry by rivals.

Transformation requires capabilities for selectively phasing out old products, adjusting lines of communication, and changing, as needed, business models, methods, and even organizational culture. As part of Carlos Ghosn's turnaround of Nissan, for example, the company launched a 3-month cross-functional study of its supplier relationships, which had been keeping its cost of parts higher than its rivals (Santos et al. 2009). Following the analysis, the company divested its interest in numerous suppliers and adopted a more arm's-length relationship. Some procurement was moved to new joint ventures with Renault. These structural changes were a key part of the company's revival.

Transformational capabilities are needed most obviously when major threats and opportunities need to be addressed. But they are also needed periodically to soften the rigidities that develop over time from asset accumulation and the development of standard operating procedures. The fast-feedback loop enabled by social media has provided a new high-speed imperative for organizational responsiveness.

In the early 2000s, network equipment company Cisco shifted from a highly centralized process, in which "all decisions came to the top 10 people in the company" to a more decentralized and collaborative management system, with a network of councils and boards entrusted and empowered to launch new businesses, and incentives to encourage executives to work together (McGirt 2008). According to CEO John Chambers, the new system required only "one week to get a [business] plan that used to take six months" (ibid.). However, over time, the structure became sclerotic and, beginning in 2009, Cisco reduced the number of councils from 12 to 3, while dissolving the associated boards, in a renewed push to speed up decision making (Clark and Tibken 2011).

Decentralization is often a component of strong dynamic capabilities. In the multinational context, this involves granting more local autonomy. The coffee specialist, Starbucks, began to enter different countries in the late 1990s. Top management recognized and responded to the need to adapt the products and overall experience for a variety of tastes and cultural habits in Europe (Alderman 2012) and, even more so, in China (Burkitt 2012). In 2011, this location-aware approach necessitated an organizational change to ensure a better balance between local voices and corporate initiatives, leading the company to change from a functional to a regional management structure.

4 Dynamic Capabilities and Strategy

Strategy is separate from, and strongly complementary to, dynamic capabilities. Dynamic capabilities must be exercised in aid of a good strategy in order to be effective, and a strategy must be supported by strong dynamic capabilities to have a chance of succeeding.

Strategy, when developed successfully, provides the specifics of how the firm will diagnose issues and trends, overcome obstacles, support customers, and outmaneuver rivals. It defines the path ahead, which may be circuitous. Strong

Table 1 The interrelation of dynamic capabilities and strategy

Strategy kernel	Diagnosis	Guiding policy	Coherent action
Related dynamic capabilities clusters	Sensing	Seizing/Transformation	Seizing/Transformation
Nature of managerial orchestration	Entrepreneurial	Administrative	Leadership

dynamic capabilities enable the enterprise to move down the path quickly and effectively. Strategy is not about goals; it is about how to overcome the obstacles standing in the way of achieving goals.

A strategy can be defined as "a coherent set of analyses, concepts, policies, arguments, and actions that respond to a high-stakes challenge" (Rumelt 2011: 6). The features of a good strategy are (1) prescient diagnoses, (2) a guiding policy, and (3) coherent action. These three functions constitute what Rumelt (2011) calls the kernel of strategy.[4] While the actions dictated by the strategy may be visible to rivals and freely imitable, rivals may not perceive it in their interest to do so until it is too late because the underlying diagnosis and policy can be kept secret.

Strategy and dynamic capabilities are analytically distinct concepts, although they are in practice somewhat interrelated (Table 1). For instance, sensing is important to dynamic capabilities but also contains a strong element of diagnosis, which is important to strategy, while seizing and transforming need to be connected to both a guiding policy and coherent action. These correspondences, and the nature of the managerial tasks most associated with each element, are outlined in Table 1.

The causal relationship and connection to value creation and capture by the enterprise is depicted schematically in Fig. 1, which lays out the basic economic logic of the dynamic capabilities framework. In the diagram, organizational capabilities drive enterprise performance. They shape and are undergirded by resources. However, strong dynamic capabilities must also be integrated with good strategy to effectuate strong performance. In short, strong dynamic capabilities, VRIN resources, and good strategy are necessary and sufficient for long-run enterprise financial success.

Firms with weaker capabilities require different strategies from firms with stronger capabilities. For example, in a company that has a culture resistant to change, the new capabilities required to implement a strategy may need to be acquired rather than developed in-house, and the resulting acquisitions managed separately in a way that minimizes conflicts with incumbent units.

[4]According to Rumelt (2011: 7), a guiding policy specifies the approach to dealing with the obstacles called out in the diagnosis. Coherent actions are feasible coordinated actions diagnosed to carry out the guiding policy.

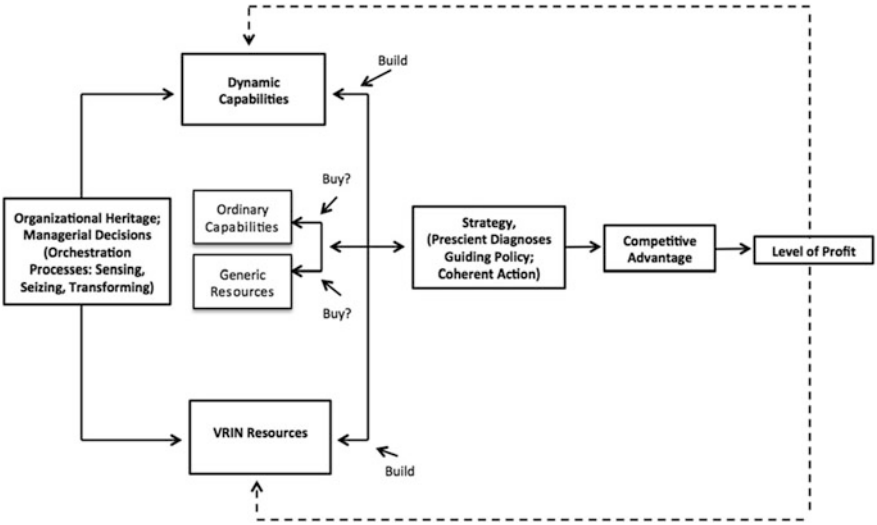

Fig. 1 Logical structure of the dynamic capabilities paradigm. *Source*: Teece (2014b)

Ultimately, high performance requires strong dynamic capabilities to sense, seize, and transform in conjunction with a good strategy. It is important to emphasize the corollary: effectiveness of dynamic capabilities can be compromised by poor strategy. The greater the diversity and rate of change in business environments, and the greater the importance of intangible (including relationship) assets, the more critical good strategy and strong dynamic capabilities become for the firm's growth and financial performance.

5 Non-Routine Capabilities: Managers and Experts

At their core, dynamic capabilities require organizational routines that are beyond best practices, and entrepreneurial leadership (Augier and Teece 2009; Teece 2007, 2012). While capabilities are often embedded throughout the organization, an established routine can adapt only to certain types of contextual change (Turner and Fern 2012). In times of turbulence, routines can be a source of dangerous organizational inertia, and the most prominent feature of the firm's dynamic capabilities become the ability of a CEO and the top management team to recognize a key development or trend, delineate a response, and guide the firm forward.

Many of the entrepreneurial actions taken by managers to seize opportunities and respond to threats simply cannot be routinized. Hence, astute entrepreneurial managers can command a premium (Castanias and Helfat 1991, 2001). Another class of activities that can only be routinized up to a point is the work of experts. Both are discussed below.

5.1 Entrepreneurial Managers

Top management's entrepreneurial and leadership skills around sensing, seizing, and transforming are required to sustain dynamic capabilities. Put differently, an important managerial function—perhaps the most important—is to achieve semi-continuous asset orchestration and renewal. Periodic, if not continuous, asset orchestration is necessary to minimize internal conflict and to maximize complementarities inside and outside the enterprise. This requirement is particularly demanding in larger, more complex firms, where top managers with the requisite unique skills are likely to be highly compensated for leading the firm to above-average returns (Castanias and Helfat 2001).

In fast-paced, globally competitive environments, consumer needs, technological opportunities, and competitor activity are constantly in flux. Opportunities open up for both newcomers and incumbents, putting the profit streams of incumbent enterprises at risk. While the path ahead for some emerging marketplace trajectories is easily recognized, others are much less so. For example, 3D printing has the potential to be a disruptive technology across a wide swath of industries, but whether it can realize the promise, and when that will happen, is far from clear.

Entrepreneurial managers, like entrepreneurs, excel at the scanning, learning, creative, and interpretive activities needed to sense (and later seize) new technological and market opportunities.[5] The requisite ability is clearly not uniformly distributed amongst individuals or enterprises. Opportunity discovery (or creation) requires specific knowledge, creative activity, the ability to understand user/customer decision making, and practical wisdom (Nonaka and Toyama 2007). It involves accumulating, interpreting, and seeing patterns in information in whatever form it appears—a chart, a picture, a conversation at a trade show, news of scientific and technological breakthrough, or the angst expressed by a frustrated customer.

Once opportunities are glimpsed, entrepreneurial managers must develop a forecast about how technologies will evolve and how—and how quickly—competitors, suppliers, and customers will respond. Competitors may or may not see the opportunity, and even if they do they may calibrate it differently.

The next is to develop, test, refine, and implement a business model, preferably one that cannot readily be imitated. Getting the timing and the basic elements of the business model right is a critical part of the innovation process (Teece 1986b, 2010).

These functions of the entrepreneurial manager are quite different from those of the ordinary manager. Managers of ordinary activities must oversee the ongoing efficiency of established processes, ensuring that schedules are met and contracts honored. Although there are creative aspects to accomplishing such tasks, they

[5] It is not given to all managers to be entrepreneurial. While this is partly a function of the MNE's culture and values, it is also specific to each individual's nature and experience. Managers differ not only in the types of skill they possess but also in "their level of ability for each type of skill" (Castanias and Helfat 2001: 663.)

are comparatively straightforward. While necessary, high operational efficiency is unlikely to be sufficient for sustained competitive advantage since creative managerial and entrepreneurial acts (e.g., creating new markets) are, by their nature, often non-routine, even visionary, and can change the fortunes of a company. The differentiated entries of Yum's KFC units in Japan and China would appear to bear out the idiosyncratic nature of the required decision making skills. KFC Japan became a success in the 1970s under the leadership of Loy Weston and Shin Ohkawara, who developed the local branch more as a fashion business than as fast food (Bartlett and Rangan 1986). In China, however, KFC's rapid growth during the 1990s is generally attributed to the aggressive store build-out and radical menu localization that took place under the leadership of Sam Su (China Economic Review 2011).

To the extent that top management processes are routinized, they may generate "signature processes" (Gratton and Ghoshal 2005). Signature processes arise from a company's heritage, including its prior management actions, certain irreversible investments, and context-specific learning. Because of their deep, enterprise-specific roots, they are not easily imitated by other firms with different cultures and histories, and they can therefore contribute to sustainable competitive advantage.

Some of the insight of an inspiring leader can, over time, be embedded in corporate culture and organizational routines both formally, through in-house corporate "universities," and less formally, through repeated demonstration and communication. In the case of sensing capabilities, for example, a desirable approach in many cases is to embed scanning and interpretive processes throughout the organization, while ensuring communication channels for feedback to reach top management.

Ideally, the culture of the organization and the collective resilience of its employees provide the foundation for transformative actions initiated by entrepreneurial managers. Texas Instruments (T.I.), for example, successfully evolved over decades under a series of CEOs who were all career-long T.I. employees who well understood what the organization could do. The company transformed itself from a diversified electronics producer in the 1980s to a narrower focus on memory chips in the 1990s, then switched in the 2000s to digital signal processors (DSPs) for cell phones and other devices, and now has refocused yet again on analog chips. Over the turbulent period 1992–2012, the company averaged a net income margin of nearly 12 % of sales despite the downdraft of the dot-com bust and other challenges affecting the company's electronics customers. The strong, shared values of T.I.'s leaders and employees, the firm's technological capabilities, and its ability to use its technology and intellectual property astutely underpin the dynamic capabilities that have helped the firm to thrive in the very turbulent global electronics industry.

In high-velocity environments, the business enterprise will tend to rely more heavily on the sensing and seizing instincts and actions of the CEO and the top management team. To the extent this is so, the capabilities will, to the extent they are not fully embedded in the values and culture of the organization, suffer from a degree of instability. In such a case, their longevity will depend, in part, on the tenure of entrepreneurs/managers/leaders.

5.2 *Human Resources and the Management of Creative and Expert Talent*[6]

In many cases, technological innovation and new product or service development require collaboration among experts. A firm's stock of experts at any one time will typically be a combination of those who have grown up from within the organization and those who have been hired temporarily or permanently specifically for their expertise. The most valuable experts are "T-shaped," with a deep understanding in their primary discipline (the vertical trunk of the T) and "also capable of interacting with and understanding specialists from a wide range of disciplines and functional areas" (IfM and IBM 2008: 11).

Expert talent has become more important than ever for the creation and management of technology in the global economy (Albert and Bradley 1997; Reich 2002). Nevertheless, if a company becomes too dependent on one or a handful of individuals, and especially if they are remunerated inappropriately, the morale of all employees can be undermined. And hiring more experts generally can't save a dysfunctional organization (Pfeffer 2001). Likewise, experts won't salvage a flawed strategy (Huselid et al. 2005), although they may be able to help replace it with a better one.

Just as important as securing access to the right experts is how they are managed. Product and service integration is often required to deliver solutions that customers will prize. As a result, many job categories are becoming so complex and interdependent that managing them in a traditional structured hierarchical format is no longer a realistic option.

The management style required to get the best out of expert employees is fundamentally different from the management style applicable to regular line employees, principally because it requires a much lighter touch. Top talent generally does not need to be told what to do, at least not by a manager with little familiarity with the expert's area of knowledge. In a global business context, the enterprise must accomplish the difficult but essential tasks of keeping its experts financially satisfied, fostering collegiality and collaboration among them, and allowing them the (guided) professional autonomy their work demands while holding them accountable to the enterprise and the challenge at hand.

Traditional notions of management that rely heavily on hierarchy and decisions driven from the top are unlikely to work well when expert work is a large component of the firm's activities. Cooperative activity for many science and engineering projects can even be self-organized to some degree.

The fundamental principle in the management of experts is that it must employ a "light touch." Otherwise, cooperative efforts will be suppressed, and creativity will be compromised. Management is seldom sufficiently informed to second-guess the difficult and granular technical tradeoffs and judgments of experts with respect

[6]This section is adapted from Teece (2011) and Linden and Teece (2014).

to solving the problem at hand. Job evaluations must be based more on results and less on adherence to established processes. Compensation requires an understanding of the contextual value of talent, and the firm must develop ways of appropriately compensating exceptional talent that can command rich remuneration elsewhere without undermining the morale of other employees.

In fast-paced complex environments where there is heterogeneity in customer needs and the focus is on innovation, it is simply impossible to achieve the necessary flexibility and responsiveness with a command-and-control organizational structure. Moreover, with a highly talented workforce, excessive centralization can shut down local initiative.

The above admonitions are not meant to imply that top management should not guide and coordinate innovative activities. Rather, managers of innovative enterprises must learn to lead without relying on the authority that comes from a position in an organizational chart or the 'C' designation in their title. This imposes new challenges for some companies and some individuals, but it is the way of the future in many contexts.[7] The challenge is to connect individual initiatives to the overall corporate strategy without building an expensive and initiative-sapping hierarchy.

While creative activities need to be organized in a decentralized way, there are operational activities involving experts that should not be managed in this way. The accounting, finance, and treasury functions are obvious examples. As noted above, when the goal is to achieve systemic innovation, there may well be activities that require tight integration because the project/technology spans multiple lines of business, or because there are very significant sales benefits to be achieved from a coordinated approach.

Achieving tight integration requires strong leadership. Many expert activities, such as developing new product lines, involve project work that requires collaboration amongst experts. An enterprise can hire the brightest, most creative people, but it is only through successfully fostering the sharing of information, collaboration, and the use of networks inside and outside the enterprise that their creative potential will be released and tight integration accomplished (Subramaniam and Youndt 2005).

To be effective, this collaborative model of knowledge generation must, in turn, be embedded in a knowledge-friendly enterprise. Table 2 contrasts a Knowledge Model (right-hand column) with the characteristics of the archetypal Industrial Model that still characterize too many large organizations.

Collaboration among experts is made necessary by their high degree of specialization. Interaction among people from diverse disciplines and functional groups is almost always required to solve the complex problems that accompany the exercise of a firm's sensing, seizing, and transforming capabilities. In fact, the top management team is just such a collaboration. A well-integrated top management team, in which members share openly and truly work together on strategic issues,

[7] See, for example, Schmidt and Rosenberg (2014).

Table 2 Contrasting management models of the business enterprise

Organizational characteristics	Industrial model	Knowledge model (for expert talent)
Hierarchy	Deep	Shallow
Leadership	Centralized	Distributed
Work	Segmented	Collaborative
People viewed as	Cost	Asset
Basis of control	Authority	Influence and example
Assumptions about individuals	Opportunistic	Honorable
Financial incentives	Base salary + Discretionary bonus	Metrics-based compensation; limited discretion

has been shown to be associated with proactive strategy formation and agile implementation in fast-moving competitive environments (Chen et al. 2010). While a diverse top management team (in terms of functional background, education, and company tenure) can take longer to reach decisions than a more uniform group, the strategies tend to be more innovative and lead to better profit outcomes (Hambrick et al. 1996).

In principle, the outcome from diverse working groups such as a cross-functional team can exceed the capabilities of its best individual members (Larson 2007). However, if not managed properly, the bringing together of specialists from different parts of the organization can increase conflicts and misunderstanding that impede innovation (Ancona and Caldwell 1992). To ensure a net benefit, the team leader must find a way to reap the creativity-stimulating benefits of diverse membership while managing the frictions.

One cannot, however, simply assume that more is better when it comes to collaboration. Consensus and participatory leadership isn't always a good thing, particularly when the issues are complex and there is considerable asymmetry in the distribution of talents on the team. The right voices need to be heard. While authentic dissent ought to be highly valued, unproductive collaboration, riven by personal conflict, can be more dangerous than missed opportunities for quality engagement and collaboration.

Group leaders can avoid suppressing healthy disagreement (based on the issues, not on the people, involved) by not expressing their positions too early in the process (Janis 1972). Openness should be encouraged by not dismissing any idea too quickly.

It is vital to have leadership, at the team level or higher, that can recognize which ideas can be rejected out of hand. Entrepreneurial managers permit experimentation and search in the pursuit of broad goals, but, as ideas emerge, they must support the best paths and close down unpromising ones.

Project teams should be kept small, intimate, yet diverse. Project groups that complete their task or run into "blind alleys" should disband so that the mix of talents can be reconfigured as needed to meet future demands.

It is also important to recognize that not everything is appropriately organized in collaborative teams. Indeed, there is a great deal about traditional teams that involve hidden and unnecessary costs. When team requirements are too heavy, decision cycles lengthen, expenses mount, and the organization adopts an inward focus.

In the multinational context, teams are increasingly spread across organizational boundaries and/or long distances. This is increasingly true for innovation, as large and small companies have begun to tap into pools of science and engineering talent in industrializing economies. Fortunately, the autonomy and trust that are appropriate for managing experts translate easily to the "virtual team" context, where continuous direct leadership may not be possible due to time zone differences. To overcome the social remoteness of distance, special measures, such as a project kick-off meeting that brings everyone to a single location for a few days, must be devised to at least partially formalize the process of fostering knowledge of individual capabilities, mutual support, and shared purpose (Siebdrat et al. 2009).

6 Case Study from the Oil and Gas Industry[8]

Up to this point, we have described dynamic capabilities and the concomitant management of experts in a general way. In this section, we discuss how the exploration division of a "supermajor" oil and gas producer, identified here as "EXP," operationalized dynamic capabilities to improve its management of expert talent to better deal with challenges facing the whole industry.

The oil and gas industry has been undergoing a number of disruptions that affect large and small companies (Shuen et al. 2014). These include the shift of demand growth to developing countries, the rise of national oil and gas companies, the emergence of large independents and vertically integrated service companies, and the increased need to tap non-traditional ("unconventional") sources such as shale formations and oil sands. These and other factors, such as concerns and regulations related to climate change, have added further technical and strategic complexity to an already complicated, global business ecosystem.

The goal of exploration activities is to accurately predict the presence of hydrocarbons within a subsurface geological formation, the extent of these deposits, their recoverability, and the subsurface risks. Geological, structural and seismic investigations are undertaken, the findings from which are then used to make investment decisions. Oil and gas wells are high-risk ventures, which can cost anywhere from $1 million to more than $1 billion, depending on the conditions (onshore/offshore) and technologies (conventional/experimental) involved. Excellence in exploration is the best opportunity to improve the chances of selecting the right opportunities.

[8]This section is based on Feiler and Teece (2014).

A fundamental challenge for all large, global companies is to recruit and deploy enough of the right people when and where they are needed. For EXP, the process is made more difficult by a decreasing and aging geoscience talent pool. Many companies increasingly rely on contractors who may not be well aligned with the company's culture, processes, and strategy. Projects that are profitable on paper may fall short in practice due to weaknesses in the company's human capital base with respect to new exploration and production requirements.

The asset orchestration at the heart of dynamic capabilities is essential for managing human capital across talent pools inside and outside the organization. A focus limited to internal resources alone ignores solutions available in the company's business ecosystem. EXP managers continuously scan for external capabilities in order to quickly identify and access contractors and joint venture partners that can leverage and expand EXP's own skills base.

To better manage its human capital, EXP added capability-awareness to its long-term planning process, which is the output of a formalized form of sensing. In the past, the criteria used by top management to prioritize ventures for funding were based purely on their relative potential economic value. Starting in 2012, an annual strategic capability evaluation process was put into effect. While expected returns remain a key consideration, project analysis was expanded to consider the quality and availability of key personnel, particularly geoscientists. Some projects with high potential return but with specific and unusual knowledge requirements were recognized to be infeasible with the currently available resources. Greater effort was put into hiring, training, and retaining the right people, then managing them with a Human Resources and IT infrastructure that enables the best-qualified people to collaborate on the highest value projects even in remote corners of the world.

Active projects were also analyzed on a company-wide basis to compare the potential value and complexity of projects with the actual deployment of technical and managerial staff and complementary external resources. A spreadsheet-driven assessment was followed by interviews with regional and selected local managers to confirm deficiencies and develop remedial plans. This was made an annual exercise under the supervision of the Chief Global Geoscientist. In dynamic capabilities terms, these changes sharpened EXP's capabilities for seizing and formalize an annual realignment process (transformation).

The changes made to improve EXP's dynamic capabilities were not limited to human capital. Initiatives were engaged to enhance complementary assets, including faster deployment of technical and process innovations into the field, greater decentralization of operational authority to local managers, and heightened emphasis on cross-region communication and collaboration.

As part of decentralization, regional leaders were given the ability to localize the company-wide exploration checklist. In the past, this was a rigidly implemented, audited series of steps that had to be completed from the beginning to the end of the exploration process. However, not all steps were relevant to all situations. Local managers were given the ability to bypass certain steps under conditions agreed upon at the center.

As part of ongoing cross-region collaboration at EXP, its regional leaders developed "informal markets," which are regularly scheduled meetings where regional delegates could propose trades of experts from their respective staffs with needed skills. This worked similarly to trades among baseball teams. Experts chosen for a move are encouraged and incentivized to agree, but were generally free to refuse, apart from cases of urgent, high-priority need.

To sharpen its ability to seize new opportunities, EXP created a dedicated division for Unconventional ventures. This provided dedicated financial, technological, and managerial resources to efforts that often differed in their requirements from the firm's traditional activities. The separation of innovative activities in some way is often necessary to make sure that they are not hampered by political infighting or the general struggle for resources by more established operations.

EXP's changes were enabled by supportive leadership. And the new activities were made accountable through reporting structures and the establishment of clear deliverables. Leadership and accountability are both important underpinnings of dynamic capabilities.

Overall, these changes in the company's exploration division enabled the top management to receive better intelligence for its decision making and to more effectively orchestrate the company's assets.

7 Conclusion

I began this paper by arguing that traditional approaches to modeling the MNE were too narrow to capture a good deal of what is truly critical to the MNE and its financial performance. The dominant economic theories of the MNE, with their roots in the work of Coase (1937), assume away any significant role for managers, capabilities, and other sources of cross-firm heterogeneity. It is not only the Coasian tradition that has shortcomings, but also other economic and financial views of the enterprise, such as Agency Theory (e.g., Jensen and Meckling 1976), which focuses on conflicts of interest between managers, shareholders, and employees.

A firm that pursues efficiency at the expense of capabilities puts itself at risk. The Boeing 787 Dreamliner case is a powerful negative example. In designing a cost-spreading supply chain for its new 787 "Dreamliner" passenger jet, Boeing decided to rely far more than ever before on a global array of suppliers to develop parts and subsystems. Unfortunately, it had also cut back its monitoring capability without having first ensured that all suppliers had the requisite design and production capabilities. Problems with inadequate components led to a very costly delay of more than 3 years (Kesmodel 2011). The problems arose more from the capability deficiencies of Boeing and its suppliers than from opportunism and contract design deficiencies on either side.

Capabilities awareness must also be reflected in the management of key personnel. Strong dynamic capabilities result in managers placing the right people in the right jobs and having the organization doing the right things as well as doing things right.

What makes MNEs distinctive is the fact that they each have separate histories, and they each span jurisdictions and territories where markets, factors of production, firms and technologies, and national and regional infrastructure are likely to be different. Because of this locational diversity, the asset orchestration at the heart of dynamic capabilities is even more vital for the MNE than for a purely domestic firm. Where certain capabilities and markets are absent, they need to be created. To secure profits, entrepreneurial managers in parent companies and subsidiaries must build signature processes, build and deploy distinct resources, and design good business models and strategies. The basic question to be answered by a robust theory of the MNE is not simply one of where to locate in order to minimize production and transaction costs but of where to locate to build or deploy signature processes and create or leverage VRIN resources.

The multidisciplinary framework presented here advances capabilities concepts without ignoring the transaction-level issues that dominate the mainstream economic model of the MNE. A more unified theory of the MNE will require a productive dialogue amongst scholars of different backgrounds. Such an approach is likely to produce theory of relevance to other theorists and to practitioners.

Acknowledgement I wish to thank Dr. Greg Linden for helpful comments.

References

Albert, S., & Bradley, K. (1997). *Managing knowledge: Experts, agencies and organizations*. Cambridge: Cambridge University Press.

Alderman, L. (2012). In Europe, Starbucks adjusts to a café culture. *New York Times*. Retrieved March 30, from http://www.nytimes.com/2012/03/31/business/starbucks-tailors-its-experience-to-fit-to-european-tastes.html?pagewanted=all

Ancona, D. G., & Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science, 3*(3), 321–341.

Antràs, P., & Helpman, E. (2004). Global sourcing. *Journal of Political Economy, 112*(3), 552–580.

Augier, M., & Teece, D. J. (2007). Dynamic capabilities and the multinational enterprise: Penrosean insights and omissions. *Management International Review, 47*(2), 175–192.

Augier, M., & Teece, D. J. (2009). Dynamic capabilities and the role of managers in business strategy and economic performance. *Organization Science, 20*(2), 410–421.

Barney, J. B. (1986). Organizational culture. *Academy of Management Review, 11*(3), 656–665.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120.

Bartlett, C. A., & Ghoshal, S. (1989). *Managing across borders: The transnational solution*. Boston: Harvard Business School Press.

Bartlett, C. A., & Rangan, U. S. (1986). *Kentucky Fried Chicken (Japan) Limited, Case No. 9-387-043*. Boston: Harvard Business School Case Services.

Birkinshaw, J. (2000). *Entrepreneurship in the global firm*. London: Sage.

Brouthers, K. D. (2013). A retrospective on: Institutional, cultural and transaction cost influences on entry mode choice and performance. *Journal of International Business Studies, 44*(1), 14–22.

Buckley, P. J. (2007). The strategy of multinational enterprises in the light of the rise of China. *Scandinavian Journal of Management, 23*(2), 107–126.

Buckley, P. J. (2009). Internalisation thinking: From the multinational enterprise to the global factory. *International Business Review, 18*(3), 224–235.

Buckley, P. J., & Casson, M. C. (1976). *The future of multinational enterprise*. London: Macmillan.

Burkitt, L. (2012). Starbucks plays to local Chinese tastes. *Wall Street Journal*. Retrieved November 26, from http://professional.wsj.com/article/SB10001424127887324784404578142931427720970.html

Burrows, P. (2004). The seed of Apple's innovation. *businessweek.com*. Retrieved October 12, from http://www.businessweek.com/print/bwdaily/dnflash/oct2004/nf20041012_4018_db083

Cantwell, J. A. (1989). *Technological innovation and multinational corporations*. Oxford: Blackwell.

Casson, M. C. (1982). *The entrepreneur—an economic theory*. Oxford: Martin Robertson.

Casson, M. C. (1997). *Information and organization: A new perspective on the theory of the firm*. Oxford: Clarendon.

Casson, M. C. (2000). *Enterprise and leadership: Studies on firms, networks and institutions*. Cheltenham: Edward Elgar.

Casson, M. C. (2005). Entrepreneurship and the theory of the firm. *Journal of Economic Behaviour and Organization, 58*(2), 327–348.

Castanias, R. P., & Helfat, C. E. (1991). Managerial resources and rents. *Journal of Management, 17*(1), 155–171.

Castanias, R. P., & Helfat, C. E. (2001). The managerial rents model: Theory and empirical analysis. *Journal of Management, 27*(6), 661–678.

Chandler, A. (1977). *The visible hand*. Cambridge, MA: The Belknap Press of Harvard University Press.

Chen, M.-J., Lin, H.-C., & Michel, J. G. (2010). Navigating in a hypercompetitive environment: The roles of action aggressiveness and TMT integration. *Strategic Management Journal, 31*(13), 1410–1430.

Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox Corporation's Technology. *Industrial and Corporate Change, 11*(3), 529–555.

China Economic Review. (2011). Original recipe: KFC remains far ahead of McDonald's in China. *chinaeconomicreview.com*. Retrieved November 1, from http://www.chinaeconomicreview.com/content/original-recipe

Clark, D., & Tibken, S. (2011). Cisco to reduce its bureaucracy. *WSJ.com*. Retrieved May 6, from http://online.wsj.com/article/SB10001424052748703859304576304890449176956.html

Coase, R. (1937). The nature of the firm. *Economica, 4*(16), 386–405.

Dunning, J. (1981). Alternative channels and modes of international resource transmission. In R. W. Moxon & H. V. Perlmutter (Eds.), *Controlling international technology transfer: Issues, perspectives and implications* (pp. 3–27). New York: Pergamon Press.

Dunning, J. H., & Lundan, S. M. (2008). *Multinational enterprises and the global economy*. Cheltenham: Edward Elgar.

Dunning, J. H., & Lundan, S. M. (2010). The institutional origins of dynamic capabilities in multinational enterprise. *Industrial and Corporate Change, 19*(4), 1225–1246.

Ernst, D., & Guerrieri, P. (1998). International production networks and changing trade patterns in East Asia: The case of the electronics industry. *Oxford Development Studies, 26*(2), 191–212.

Feiler, P., & Teece, D. (2014). Case study, dynamic capabilities and upstream strategy: Supermajor EXP. *Energy Strategy Review, 3*, 14–20.

Gans, J., & Stern, S. (2010). Is there a market for ideas? *Industrial and Corporate Change, 19*(3), 805–837.

Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy, 12*(1), 78–104.
Ghemawat, P. (2003). Semiglobalization and international business strategy. *Journal of International Business Studies, 34*(2), 138–152.
Gratton, L., & Ghoshal, S. (2005). Beyond best practice. *MIT Sloan Management Review, 46*(3), 49–57.
Grossman, S. J., & Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy, 94*, 691–719.
Grossman, G. M., & Helpman, E. (2004). Managerial incentives and the international organization of production. *Journal of International Economics, 63*(2), 237–262.
Hambrick, D. C., Cho, T. S., & Chen, M.-J. (1996). The influence of top management team heterogeneity on firms' competitive moves. *Administrative Science Quarterly, 41*(4), 659–684.
Helfat, C. E., Finkelstein, S., Mitchell, W., Peteraf, M. A., Singh, H., Teece, D. J., et al. (2007). *Dynamic capabilities: Understanding strategic change in organizations*. Oxford: Blackwell.
Hennart, J. F. (2009). Down with MNE-centric theories! Market entry and expansion as the bundling of MNE and local assets. *Journal of International Business Studies, 40*, 1432–1454.
Hennart, J. F. (2010). Transaction cost theory and international business. *Journal of Retailing, 86*(3), 257–269.
Huselid, M. A., Beatty, R. W., & Becker, B. E. (2005). A players or a positions? *Harvard Business Review, 83*(12), 110–117.
IfM, & IBM. (2008). *Succeeding through service innovation*. Cambridge: University of Cambridge Institute for Manufacturing. http://www.ifm.eng.cam.ac.uk/dlregister/ccdf17e234a153587c7c88c027429367/resources service succeeding-through-service-innovation/
Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Boston: Houghton, Mifflin.
Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360.
Kesmodel, D. (2011). Boeing examines supply chain for weak links. *WSJ.com*. Retrieved December 30, from. http://online.wsj.com/article/SB10001424052970204058404577111091095438300.html
Kogut, B., & Zander, U. (1993). Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies, 24*, 625–645.
Larson, J. R., Jr. (2007). Deep diversity and strong synergy: Modeling the impact of variability in members' problem-solving strategies on group problem-solving performance. *Small Group Research, 38*(3), 413–436.
Linden, G., & Teece, D. J. (2014). Managing expert talent. In P. Sparrow, H. Scullion, & I. Tarique (Eds.), *Strategic Talent Management: Contemporary Issues in International Context* (pp. 87–116). Cambridge: Cambridge University Press.
McGirt, E. (2008). Revolution in San Jose. *Fast Company, 131*, 90–93.
Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725.
Nonaka, I., & Toyama, R. (2007). Strategy as distributed practical wisdom (Phronesis). *Industrial and Corporate Change, 16*(3), 371–394.
Penrose, E. G. (1959). *The theory of the growth of the firm*. New York: Wiley.
Pfeffer, J. (2001). Fighting the war for talent is Hazardous to your organization's health. *Organizational Dynamics, 29*(4), 248–259.
Pisano, G., & Teece, D. J. (2007). How to capture value from innovation: Shaping intellectual property and industry architecture. *California Management Review, 50*(1), 278–296.
Pitelis, C. N. (2004). Edith Penrose and resource—based view of (international) business strategy. *International Business Review, 13*(4), 523–532.
Pitelis, C. N. (2007). Edith Penrose and a learning—based perspective on the MNE and OLI. *Management International Review, 47*(2), 207–220.

Pitelis, C. N., & Teece, D. J. (2010). Cross-border market co-creation, dynamic capabilities and the entrepreneurial theory of the multinational enterprise. *Industrial and Corporate Change, 19*(4), 1247–1270.

Reich, R. (2002). *The future of success: Working and living in the new economy*. New York: Vintage.

Rugman, A. M. (1981). *Inside the multinationals: The economics of internal markets*. New York: Columbia University Press.

Rugman, A. M., & Verbeke, A. (2003). Extending the theory of the multinational enterprise: Internalization and strategic management perspectives. *Journal of International Business Studies, 34*(2), 125–137.

Rumelt, R. (2011). *Good strategy/bad strategy: The difference and why it matters*. New York: Crown Business.

Santos, J., Spector, B., & Van der Heyden, L. (2009). Toward a theory of business model innovation within incumbent firms. Faculty & Research Working Paper 2009/16/EFE/ST/TOM. *INSEAD, Fontainebleau, France*. http://flora.insead.edu/fichiersti_wp/inseadwp2009/2009-16.pdf

Schmidt, E., & Rosenberg, J. (2014). How Google works. *Business Plus*.

Shuen, A., Feiler, P., & Teece, D. (2014). Dynamic capabilities in the upstream oil and gas sector: Managing next generation competition. *Energy Strategy Review, 3*, 5–13.

Siebdrat, F., Hoegl, M., & Ernst, H. (2009). How to manage virtual teams. *MIT Sloan Management Review, 50*(4), 63–68.

Spencer, B. J. (2005). International outsourcing and incomplete contracts. *Canadian Journal of Economics, 38*(4), 1107–1135.

Sturgeon, T. J. (2002). Modular production networks: A new American model of industrial organization. *Industrial and corporate change, 11*(3), 451–496.

Subramaniam, M., & Youndt, M. A. (2005). The influence of intellectual capital on the types of innovative capabilities. *Academy of Management Journal, 48*(3), 450–463.

Teece, D. J. (1975). *The multinational corporation and the resource cost of international technology transfer*. Philadelphia: Economics Department, University of Pennsylvania.

Teece, D. J. (1981a). The multinational enterprise: Market failure and market power considerations. *Sloan Management Review, 22*(3), 3–17.

Teece, D. J. (1981b). The market for know-how and the efficient international transfer of technology. Annals of the Academy of Political and Social Science, *458*(November), 81–96. In R. G. Hawkins & A. J. Prasad (Eds.), *Technology transfer and economic development*. Greenwich, CT: JAI Press.

Teece, D. J. (1986a). Transaction cost economics and the multinational enterprise: An assessment. *Journal of Economic Behavior and Organization, 7*(1), 21–45.

Teece, D. J. (1986b). Profiting from technological innovation. *Research Policy, 15*(6), 285–305.

Teece, D. J. (2000). *Managing intellectual capital: Organizational, strategic, and policy dimensions*. Oxford: Oxford University Press.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350.

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning, 43*(2–3), 172–194.

Teece, D. J. (2011). Human capital, capabilities and the firm: Literati, Numerati, and entrepreneurs in the 21st-century enterprise. In A. Burton-Jones & J.-C. Spender (Eds.), *The Oxford handbook of human capital* (pp. 527–562). Oxford: Oxford University Press.

Teece, D. J. (2012). Dynamic capabilities: Routines versus entrepreneurial action. *Journal of Management Studies, 49*(8), 1395–1401.

Teece, D. J. (2014a). A dynamic capabilities-based entrepreneurial theory of the multinational enterprise. *Journal of International Business Studies, 45*(1), 8–37.

Teece, D. J. (2014b, March 25). The foundations of enterprise performance: Dynamic and ordinary capabilities in an (economic) theory of firms. *Academy of Management Perspectives.* doi:10.5465/amp.2013.0116.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533.

Turner, S. F., & Fern, M. J. (2012). Examining the stability and variability of routine performances: The effects of experience and context change. *Journal of Management Studies, 49*(8), 1407–1434.

Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics, 80*(2), 190–207.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*(2), 171–180.

Williamson, O. E. (1981). The modern corporation: Origins, evolution, attributes. *Journal of Economic Literature, 19*(4), 1537–1568.

Williamson, O. E. (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. New York: Free Press.

Zahra, S. A., Ireland, R. D., & Hitt, M. A. (2000). International expansion by new venture firms: international diversity, mode of market entry, technological learning, and performance. *Academy of Management Journal, 43*(5), 925–950.

Determinants of the Firm's Foreign Internalization Decision

Susanna Thede

Abstract This paper empirically investigates predictions of recently developed property-rights theories on the firm's foreign internalization decision (Antràs and Helpman, *Journal of Political Economy*, *112*(3), 552–580, 2004; Antràs and Helpman, *The Organisation of Firms in a Global Economy.* Harvard University Press, 2008). Using census data on Swedish manufacturing firms for the 1997–2007 time period, we provide novel and complementary evidence on the topic. In particular, our data enables us to investigate the firm's foreign internalization decision directly in a sample that is not subject to sample selection bias. Our estimation results provide stark evidence in favor of the predicted impact of firm heterogeneity and foreign contracting enforcement that lies at the heart of the reference theories. Firms in industries with larger within-industry firm variation, as depicted by the variation in TFP levels, are more prone to integrate offshored production. We also provide novel evidence on the interplay of firm heterogeneity and property-rights factors at the firm level. In particular, the estimation results confirm the Antràs and Helpman (2008) prediction that foreign legislative institutions in support of headquarter provision favor the firm's foreign internalization decision.

1 Introduction

The purpose of this paper is to empirically examine determinants of the firm's foreign internalization decision based on recently developed trade theories incorporating within-industry firm heterogeneity.[1] Our particular focus is placed on the Antràs and Helpman (2004) model, which relies on a property-rights approach to explain the boundary of the firm, and the subsequent Antràs and Helpman

[1]See Antràs and Rossi-Hansberg (2009), Helpman (2006) and Spencer (2005) for surveys of models explaining the firm's foreign internalization decision.

S. Thede (✉)
Institute for European Studies, University of Malta, Msida, Malta
e-mail: Susanna.Thede@um.edu.mt

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_6

(2008) model, which extends the prior model to a more complex contracting environment. Despite the fact that several empirical studies examine predictions from these models (Bernard et al. 2010; Corcos et al. 2013; Defever and Toubal 2007; Jabbour 2012; Nunn and Trefler 2008; Yeaple 2006),[2] discrepancies in the chosen approaches and data availability of these contributions leave a scattered picture of the models' performance.[3]

This paper provides novel and complementary evidence on the topic with the five main contributions being as follows. First, we identify model concordances with observed firm behavior to supplement the examination of model predictions on offshoring firms' integration patterns. Second, our empirical investigation is performed using direct evidence on the firm's foreign internalization decision. Third, we disentangle behavioral effects stemming from the property-rights approach to examine whether foreign contracting implications of the Antràs and Helpman (2004, 2008) models can be validated by the data. Fourth, the estimations are based on a census of manufacturing firms that offshore production from the home country (Sweden) and can supplement survey and otherwise restricted data evidence that may be subject to sample selection bias. Fifth, our longitudinal data set allows us to control for time-specific effects in the estimations.

The economic study of the firm's internalization decision was initiated by Williamson (1975, 1985). It has recently been revived by several trade theorists as Melitz (2003) and others have introduced new modeling techniques that enable the study of within-industry firm heterogeneity. The tendency for firms in the same industry to self-select into various organizational forms is generally observed in the data, and since a firm's sales and profits are influenced by its choice of organizational form, this self-selection mechanism contribute to shaping the market structure and performance of industries. In this context, identifying determinants of the firm's foreign internalization decision can contribute to properly predict the international fragmentation of production. A better understanding of the firm choice to integrate its offshored production can thereby have important domestic and foreign welfare implications.

We provide a thorough description of predicted effects of determinants on firms' organizational patterns, disentangling their individual impacts to underpin the empirical investigation. We combine our empirical analysis of model predictions with a detailed investigation of whether observed firm behavior concurs with underlying model assumptions. In particular, we identify whether more productive firms choose to integrate their input provision in the source country. The Antràs and Helpman (2004, 2008) models are founded on this assumption. Corcos et al. (2013) attributes discrepant empirical evidence based on French survey data to sample selection problems, revealing that the underlying productivity ranking of firms is upheld when accounting for sample selection in the estimations. Our access

[2] Antràs (2014) provides a recent survey of empirical contributions in the research field.

[3] In addition, Federico (2010) and Kohler and Smolka (2009) identifies productivity premia of foreign internalization decisions with reference to Antràs and Helpman's (2004) model.

to (Swedish) census data enables us to establish the sorting behavior of firms in a sample that includes small and purely domestic offshoring firms omitted in prior empirical studies on the topic. From this perspective, we use more representative data on firms' offshoring behavior that enables us to provide supplementary evidence on the issue. To preview the results, our findings support the predicted sorting pattern.

We directly investigate the firm's foreign internalization decision formalised in the Antràs and Helpman (2004, 2008) models. A predominant part of prior research in the area has used the alternative strategy of analysing intra-firm trade shares (Bernard et al. 2010; Corcos et al. 2013; Defever and Toubal 2007; Nunn and Trefler 2008; Yeaple 2006), which is a perfectly consistent way to capture the firms engaged in offshoring-FDI in theory. As noted by Bernard et al. (2010) and Corcos et al. (2013) in performing Heckman estimations on intra-firm trade data, the empirical determination of the choice to engage in offshoring-FDI may however differ from that of offshoring-FDI levels. It is therefore motivated to re-examine whether the results identified in prior research based on intra-firm trade shares are empirically supported by the direct study of the firm's internalization decision. In coherence with prior empirical research evidence, our results largely support predictions of the Antràs and Helpman (2004, 2008) models.

Our reference models depict the firm's foreign internalization decision as a two-dimensional organizational choice of ownership structure (integration/outsourcing) and source location (North/South).[4] Our access to detailed data on firm behavior in each source market allows us to identify the direct impact of locational factors. Our reference models describe how contracting limitations affect the firm's incentive to acquire control over assembly production. Due to incomplete contracting and relationship-specific investments in headquarter and assembly production, there is a hold-up problem between parties in charge of different production stages. The firm takes this into account in making its ownership decision, facing a starker integration incentive under better foreign contracting enforcement as more is regained from assembly production in case of a contractual breach (Antràs and Helpman 2004) and/or because lower foreign investment risks stimulate assembly production and raises the importance of incentivizing headquarter provision (Antràs and Helpman 2008).

We directly examine whether this model implication is validated by identifying whether firms' integration productivity threshold is decreasing in the foreign legislative support of assembly provision. This way, we provide the novel contribution of disentangling the effect from the transaction cost effect through which better foreign contracting enforcement triggers offshoring-outsourcing of low-productivity firms (Antràs and Helpman 2008). Our firm estimations provide mixed evidence

[4]It can be argued that a cautious empirical interpretation of these two-country models should include firms' domestic integration and outsourcing decisions. Due to data limitations, the paper follows closely related research in the area and excludes this internalization aspect from the investigation.

of this prediction. We instead yield stark support of the prediction that foreign legislative institutions working to protect headquarter provision intensifies the need to incentivize relationship-specific investments in assembly production, raising firms' integration productivity threshold (Antràs and Helpman 2008). Our findings also reveal that firms' integration productivity threshold is decreasing in the foreign development level, which may reflect a self-selection effect triggered as firms to which headquarter activities are more important in the production process are less sensitive to low-wage cost incentives (Antràs 2005a).

Our empirical investigation is performed based on census data of all manufacturing firms active in Sweden over the 1997–2007 time period. In the data set, which has been provided by the Swedish Agency for Growth Policy Analysis, firms are categorized by the standard Swedish industry (SNI) classification at the 4-digit level of aggregation.[5] Our sample is restricted to including Swedish-owned firms (with a Swedish ownership of at least 50 %) to capture firms with headquarter activities located in the home country. Firms in our sample offshore production to a total of 254 source countries. Logit estimations are performed to investigate the offshoring firm's integration decision in a particular source location while Poisson Pseudo Maximum Likelihood (PPML) and Heckman estimations are used as alternative methods to examine the within-industry share of offshoring firms that integrate in a particular source location. The results of our industry estimations conform to model predictions, though these receive starker support using the Heckman method.

The rest of this paper is organized as follows. In Sect. 2, the Antràs and Helpman (2004, 2008) model determinants of the firm's foreign internalization decision are introduced. Section 3 contains a description of our empirical approach. The estimation results are presented in Sect. 4. The last section of the paper concludes.

2 Determinants of the Firm's Foreign Internalization Decision

This section introduces determinants of the firm's foreign internalization decision in the Antràs and Helpman (2004, 2008) models. Factors that directly influence this decision are the firm's productivity level, the industry's fixed organizational costs of production, the industry's dependency on headquarter activities in production and the wage gap between the home and source country. Due to the property-rights model foundation, contracting limitations indirectly influence firm gains from foreign integration and/or outsourcing if the offshoring production relationship fails (Antràs and Helpman 2004) and/or succeeds (Antràs and Helpman 2008). The main predictions of the Antràs and Helpman (2004) model are transferred to their subsequent model, which depicts additional (multifaceted) effects of foreign

[5]Our investigation is based on the 2002 version of this classification. Earlier data (up to 2002) is updated using the concordance key provided by statistics Sweden.

contracting enforcement. As a consequence, the theory description in this section is primarily focused on generally applicable implications of the former model.

In the Antràs and Helpman (2004, 2008) models, within-industry productivity variation determines heterogeneous firm behavior. Each firm receives its productivity level upon (home) market entry. The self-selection pattern of firms into organizational form can be identified through a set of productivity thresholds that are jointly determined by the remaining set of determinants. Combined with information on the industry dispersion of firm productivity levels, these thresholds determine the integrating share of offshoring firms in a particular industry and source country. Our empirical investigation relies on the model scenario in which firms of all four organizational forms coexist in the same industry.[6]

The productivity mapping of firms into organizational categories depends on the fixed costs required to establish alternative organizational structures. The fixed organizational cost of integration exceeds that of outsourcing so that only the most productive firms can profit from acquiring ownership of assembly production. Antràs and Helpman (2004) introduce this assumption with the argument that managerial overload exceeds economies of scope in management. Due to data limitations, we are unable to investigate the direct impact of fixed organizational costs and instead infer whether firm productivity sorting into offshoring-outsourcing and offshoring-FDI is consistent with the underlying cost ranking. Our reference models also predict that fixed organizational costs are lower in the North than in the South. This cost discrepancy is attributed to the fact that higher research, monitoring and communication costs are encountered with production offshoring to the South.

The firm's incentive to integrate hinges on the industry's dependency on headquarter activities in production, the so-called headquarter intensity of production. As integration incentivizes the provision of headquarter activities while disincentivizing the provision of assembly activities under the property-rights approach, a firm can only gain from integration if headquarter activities are sufficiently important in its production process. Because the integration incentive is stronger for firms' using a more headquarter-intensive technology, our reference models predict that the share of offshoring firms acquiring control over assembly production is increasing in the industry's headquarter-intensity of production. Following Yeaple (2006), we measure this factor through the research intensity of production while controlling for lower transaction costs to integration in capital-intensive production (Antràs 2003).

The firm's foreign internalization decision is affected by the wage reduction that can be attained by offshoring production to the South. A positive wage gap vis-à-vis the source country introduces a foreign (variable) cost advantage that increases is attractiveness as offshoring destination, reducing the offshoring-outsourcing productivity threshold in the location. The wage effect favors the offshoring firm's choice of foreign market entry in outsourcing mode, reducing the integrating share of offshoring firms. As firms with higher headquarter intensity are likely to

[6]Even if fewer organizational forms coexist in some of our disaggregated manufacturing industries, we rely on the model implication that the ranking of existing forms remain intact.

respond less to low-wage cost incentives under the property-rights approach (Antràs 2005a), a self-selection effect may also be triggered through which the productivity integration threshold decreases in the foreign wage. We control for this additional effect in the firm estimations. As we do not access comparable international wage data, the foreign wage is proxied by the source country's level of development.

Our reference models provide partially overlapping foreign contracting implications for offshoring firm behavior. In the Antràs and Helpman (2004) model, firm gains from integration depend on the quality of foreign legislative institutions. Specifically, the integrating firm's outside option (in case the offshoring relationship fails) is improved with better foreign contracting enforcement. This effect works to reduce the productivity threshold at which offshoring firms integrate assembly production. In the Antràs and Helpman (2008) model, improved foreign contracting enforcement can stimulate relationship-specific assembly investments that makes it more profitable for the firm to incentivize the provision of headquarter activities through integration. However, Antràs and Helpman (2008) also highlight that improved foreign contracting enforcement may reduce headquarter investment risks in the relationship-specific production chain. This way, the Antràs and Helpman (2008) model depicts contrasting effects of foreign legislative institutions working through the provision of assembly and headquarter activities under the property-rights approach.

The model also depicts that better foreign contracting enforcement gives rise to a transaction cost effect making the host country more attractive as an offshoring destination. Since offshoring is more profitable upon foreign market entry in the outsourcing category, this effect works to reduce the offshoring-outsourcing productivity threshold. The effect may overturn the property-rights induced effect through which better foreign legislative support of assembly provision raises the share of offshoring firms that integrate in the destination. In our empirical investigation, we measure the impact of foreign contracting enforcement on assembly/headquarter provision through interaction variables that jointly capture the industry's relationship-specificity with respect to assembly/headquarter provision and the legislative quality in the offshoring destination. The assembly contracting measure has previously been used by Nunn and Trefler (2008).

Only the most productive firms that offshore production to a source country gain from entering the market in integration mode because these firms yield a larger production surplus from integrating assembly production under the property-rights approach. At given organizational productivity thresholds, a more dispersed productivity distribution therefore raises the integrating share of offshoring firms.[7]

[7]Firm productivity levels are assumed to be Pareto distributed (as in Melitz (2003)).

3 Empirical Approach

We investigate model assumptions on firm behavior before examining industry model implications. In analyzing assumptions underlying the firm's foreign internalization decision, particular focus is placed on productivity effects. Then, predicted industry patterns comprising the combined impact of within-industry firm heterogeneity, property-rights and transaction cost effects are put to data.

To study the firm's foreign internalization decision, we empirically examine the offshoring firm's choice to integrate in the source country. The probability that firm i, which offshores production in industry j to country k at time t, integrates in the source location is determined by a logit model:

$$\begin{aligned}\Pr\left(INT_{ijkt}=1\right) = \Lambda\Big(&\alpha_j D_j+\alpha_k D_k+\alpha_t D_t+\beta\cdot PROD_{ijt}+\chi\cdot RINT_{ijt}\\&+\delta\cdot CINT_{ijt}+\phi_1\cdot PROD_{ijt}\cdot ASCON_{jkt}\\&+\phi_2\cdot PROD_{ijt}\cdot HQCON_{jkt}+\phi_3\cdot PROD_{ijt}\cdot DEV_{kt}\Big)\end{aligned} \tag{1}$$

where *INT* is a dichotomous variable taking the value one under integration, *D* denote dummy control variables capturing industry, location and time effects, *PROD* is the productivity level, *RINT* and *CINT* are research and capital intensities of production, *ASCON* and *HQCON* are variables capturing assembly and headquarter provision effects of foreign contracting enforcement and *DEV* is the foreign development level. The interaction variables are included to examine the property-rights implications that a favorable foreign legislative impact on assembly (headquarter) provision reduces (raises) the productivity threshold of integrating firms, controlling for the possible impact of firm self-selection into low-wage locations.

In our reference models, industry patterns of firms' foreign internalization decisions are determined through joint variable effects. Our industry investigation is therefore performed for a multiplicative specification, which is evaluated using alternative estimation methods. Amongst firms that offshore production in industry j to country k at time t, the share that integrates in the source location is determined by:

$$INTS_{jkt}=PDISP_{jt}^{\varphi}\cdot ARINT_{jt}^{\gamma}\cdot ACINT_{jt}^{\eta}\cdot ASCON_{jkt}^{\kappa_1}\cdot HQCON_{jkt}^{\kappa_2}\cdot DEV_{kt}^{\lambda}\cdot e^{\mu_k D_k+\mu_t D_t} \tag{2}$$

where *INTS* is the integrating share of offshoring firms, *PDISP* is the firm productivity dispersion, and *ARINT* and *ACINT* are (industry averages of) research and capital intensities of production. To avoid misspecification due to sample selection bias, we use Poisson Pseudo Maximum Likelihood (PPML) and Heckman methods to estimate the industry specification (directly or in its log-linear form). This way, we account for industry-location observations for which the *INTS* variable takes zero values.

We use the standard offshoring measurement introduced by Feenstra and Hanson (1996, 1999), which depicts offshoring in terms of imports. An offshoring firm is thereby identified through imports from the source country. As the firm's imports of intermediate inputs can be argued to provide a more accurate representation of assembly offshoring, we use this as an alternative measurement. In the dependent variables construction, offshoring measurement through imports and imports of intermediate inputs is denoted by *I* and *II*. The *INT* variable takes the value one if the offshoring firm has affiliate production in the source country, as inferred from foreign employment of personnel. The *INTS* variable equals the number of firms that have affiliate production in the source country out of the total number of firms offshoring production to the location.

The *PROD* variable is measured using Olley and Pakes' (1996) and Aw et al.'s (2003) alternative TFP calculating methods. As Aw et al. (2003) measures the TFP deviation from a hypothetical firm with input and output levels at industry means, and we want results to be directly comparable across estimations, we construct a corresponding measure based on Olley and Pakes productivity estimates using the TFP deviation from the industry productivity mean. TFP levels are estimated using data on the firm's value added, number of employees, capital stock and capital investment. To minimize accounting error while maintaining a large number of observations, the Olley and Pakes TFP estimates are based on a 3-digit level of industry aggregation. Value added, capital stock and capital investment figures are adjusted by yearly (2-digit level) producer price indices from Statistics Sweden.[8] The *PDISP* variable is measured using the standard deviation of log productivity levels (measured by Olley and Pakes' or Aw et al's method) amongst all Swedish firms in the industry. The *RINT* variable is measured by the employment share of researchers. The *CINT* variable is measured by the log capital to labor ratio, which is constructed from inflation-adjusted data on capital stocks and employment data. The *ARINT* and *ACINT* variables are average research and capital intensities of all Swedish firms in the industry. All firm-level data has been provided by the Swedish Agency for Growth Policy Analysis.

The *ASCON* and *HQCON* variables are measured by composite indicators of the industry's relationship-specificity and the source country's legislative quality. Based on the argument that production of heterogeneous products requires relationship-specific investments, we follow Nunn and Trefler (2008) in measuring relationship-specificity in assembly provision through the share of heterogeneous input products in the industry. The variable construction is based on Rauch's (1999) conservative industry classification, which categorizes goods sold on an organized exchange and reference-priced goods as homogenous and differentiated products as heterogeneous. The share of homogenous input products comes from Nunn's (2007) corresponding 1997 indicators, which are aggregated up to 4-digit

[8]For three industries that lack producer price indices at the 2-digit level, we approximate index values using simple averages across industries for which price indices are reported in the same 1-digit level industry category.

averages and matched using SNI and NAICS industry concordances interlinked via European NACE Rev 1 and UN ISIC Rev 3 correspondence tables.[9,10] We measure relationship-specificity in headquarter provision through the heterogeneity of output products in the industry, constructed directly from the Rauch (1999) classification using SNI and SITC Rev 2 industry concordances interlinked via European NACE Rev 1 and UN ISIC Rev 2 industry correspondence tables. Since a SNI industry category may correspond to several SITC industry categories, this measure can take any value between 0 and 1.

Our measure of foreign legislative quality is based on the World Bank Rule of Law index, which captures various quality aspects of a country's legislative institutions (ranging from the business use of security payments to the consistency of court decisions).[11] The index, which is constructed from a distribution ranging between −2.5 and 2.5, is recalculated to lie in the 0–1 interval. The legislative quality of the few outlier observations that have malfunctioning institutions (with index values below −2.5) such as Somalia, receive the corresponding negative values in our index transformation. For the years 1999 and 2001, interpolated indicator values are used to replace missing values. To capture the impact of foreign legislative institutions on assembly/headquarter provision, each composite variable equals the product of the inverse measure of relationship-specificity at the designated party's production stage and our legislative quality measure.

Lastly, the *DEV* variable is measured by log GDP per capita levels calculated from PPP-adjusted GDP and population figures provided in the World Bank's World Development Indicator data base. Descriptive statistics for variables entering our firm and industry samples are provided in Tables 3 and 4 in the appendix.

4 Empirical Results

In Table 1, logit estimation results based on Eq. (1) are reported for firm specifications incorporating alternative *INT* and *PROD* measures. All estimations include industry, location and time dummies (that are unreported for brevity). The results are largely consistent across estimations, though results obtained for the *ASCON* and *PROD* interaction variable are sensitive to productivity measurement. The *PROD* parameter estimate is positive and mostly significant from zero at the one percent level, validating that more productive offshoring firms self-select into integration mode. The result reveals that the modeled productivity ranking of outsourcing and

[9]We follow Nunn (2007) in letting the input use in US industries approximate that of industries outside the country, which is consistent with the setup of our reference models.

[10]In the few cases when product homogeneity measures are lacking at the 4-digit level, these are approximated by simple averages across corresponding industries at higher levels of aggregation.

[11]A detailed description of the basic index is provided by Kaufmann et al. (1999). The index version used in this paper is described in Kaufmann et al. (2009).

Table 1 Estimation results, firm specification

Indvar\Depvar	*INT(I)*	*INT(I)*	*INT(II)*	*INT(II)*
PROD(OP)	0.264		0.808***	
	(0.173)		(0.276)	
PROD(Aw)		1.977***		1.997***
		(0.076)		(0.103)
RINT	1.305***	1.094***	1.853***	1.614***
	(0.147)	(0.151)	(0.263)	(0.030)
CINT	0.308***	1.654***	0.322***	1.606***
	(0.009)	(0.025)	(0.011)	(0.260)
PROD(OP)·ASCON	−0.255***		−0.062***	
	(0.085)		(0.122)	
PROD(Aw)·ASCON		0.558***		0.350**
		(0.101)		(0.140)
PROD(OP)·HQCON	0.910***		0.744***	
	(0.067)		(0.090)	
PROD(Aw)·HQCON		0.295***		0.377***
		(0.070)		(0.090)
PROD(OP)·DEV	−0.071***		−0.124***	
	(0.021)		(0.032)	
PROD(Aw)·DEV		−0.234***		−0.221***
		(0.004)		(0.005)
Nobs	241,354	241,354	146,920	146,920
Loglikelihood value	−84,746	−81,176	−56,541	−54,291
Pseudo R2	0.133	0.169	0.138	0.172

Robust standard errors are reported within parenthesis
*, ** and *** denote parameter estimates different from zero at the 10, 5 and 1 % significance level, respectively

integrating offshoring firms can be validated in the absence of sample selection problems. The *RINT* and *CINT* parameter estimates are always positive and significantly different from zero at the 1 % level. These results provide stark additional support of the property-rights implication that a higher headquarter intensity of production favors the firm decision to integrate and of Antràs' (2003) conjecture that integration transaction costs are lower in capital-intensive production.

As previously described, the results are mixed for the property-rights prediction that the integration productivity threshold is decreasing in the foreign legislative support of assembly provision. While this does not rule out that foreign contracting enforcement working in favor of assembly provision raises the offshoring firm's integration incentive, it gives no firm evidence to validate the prediction. In contrast, parameter estimates for the *HQCON* and *PROD* interaction variable are always positive and significantly different from zero at the one percent level, yielding strong support to the property-rights prediction that offshoring firms' integration threshold

is increasing in the foreign legislative support of headquarter provision. The result strengthens prior firm-level evidence provided by Corcos et al. (2013).

The parameter estimate of the *PROD* and *DEV* interaction variable is always negative and significantly different from zero at the 1 % level, which is coherent with a negative foreign wage impact on the integration productivity threshold. The result provides stark evidence in support of firm self-selection into low-wage locations (stemming from Antràs (2005a)), highlighting that the foreign wage can reduce the offshoring-FDI productivity threshold under the property-rights approach.

In Table 2, PPML and Heckman estimation results based on Eq. (2) are presented for industry specifications incorporating alternative *INT* and *PROD* measures. While parameter signs are consistent across estimations, the *ARINT* variable only receives support in those using the Heckman method. The parameter estimates of all other variables are significantly different from zero at the 1 % level, yielding overall support of predicted industry patterns. The positive *PDISP* parameter estimate is consistent with the model prediction that a larger share of offshoring firms choose to integrate assembly production under starker domestic firm heterogeneity in the industry. Indeed, the productivity dispersion can be shown to affect firms' integration decisions even under the specification corresponding to the firm estimation for which no individual productivity effect could be identified. This discrepancy highlights that the reference models' industry focus can provide a well-attuned description of how firm productivity differences feed onto offshoring firms' integration decisions.

As previously described, empirical support of the impact of the *ARINT* variable is only obtained with the Heckman estimation method. This discrepancy may be due to method differences in the account of sample selection effects, in particular if the industry's headquarter intensity of production impose unpredicted nonlinear effects on firms' offshoring and integration decisions. This form of nonlinearity could occur if assembly production is standardized over time, yielding a product-cycle scenario where research-intensive industries at a later stage of the product cycle are more inclined to outsource assembly provision (Antràs 2005b). The *ACINT* parameter estimates are positive, confirming Antràs' (2003) prediction that firms are more prone to integrate offshored production in capital-intensive industries.

The *ASCON* parameter estimates are positive, supporting the prediction that offshoring firms' integration decisions are favored by legislative institutions in support of assembly provision. The result is consistent across estimations, showing that property-right implications of starker legislative support of assembly contracting are borne out by data at the industry level.[12] Moreover, while there may be a transaction-cost effect reducing the offshoring-outsourcing productivity threshold in the designated source market, it is exceeded by the property-rights effect. This

[12]The discrepant results obtained for some firm specifications may reflect that the firm variation in relationship-specificity in assembly provision differs across industries, yielding results that are sensitive to individual firm productivity measurement.

Table 2 Estimation results, industry specification

Method	Poisson				Heckman			
Indvar\Depvar	*INTS(I)*	*INTS(I)*	*INTS(II)*	*INTS(II)*	*INTS(I)*	*INTS(I)*	*INTS(II)*	*INTS(II)*
PDISP(OP)	0.011***		0.013***		0.140***		0.219***	
	(0.001)		(0.001)		(0.045)		(0.052)	
PDISP(Aw)		0.010***		0.011***		0.229***		0.210***
		(0.001)		(0.001)		(0.046)		(0.053)
ARINT	0.184	0.132	0.794	0.743	0.035***	0.032***	0.067***	0.060***
	(0.659)	(0.660)	(0.707)	(0.709)	(0.005)	(0.005)	(0.007)	(0.006)
ACINT	0.684***	0.704***	0.688***	0.710***	0.967***	1.050***	1.028***	0.999***
	(0.019)	(0.019)	(0.021)	(0.021)	(0.048)	(0.046)	(0.062)	(0.060)
ASCON	1.808***	1.614***	1.660***	1.502***	0.773***	0.548***	1.031***	0.707***
	(0.285)	(0.292)	(0.315)	(0.322)	(0.090)	(0.048)	(0.139)	(0.085)
HQCON	−0.676***	−0.708***	−0.726***	−0.767***	−0.363***	−0.338***	−0.305***	−0.284***
	(0.103)	(0.103)	(0.109)	(0.109)	(0.024)	(0.023)	(0.021)	(0.020)
DEV	0.596***	0.606***	0.704***	0.707***	0.798***	0.833***	1.282***	1.140***
	(0.155)	(0.155)	(0.175)	(0.175)	(0.117)	(0.113)	(0.158)	(0.145)
Inv. Mills					8.046***	8.118***	9.112***	7.952***
					(0.556)	(0.457)	(0.811)	(0.671)
Nobs	51,083	51,083	37,552	37,552	11,656	11,707	9211	9236
Loglik. value	−17,271	−17,268	−14,588	−14,586				
Pseudo R2	0.061	0.061	0.060	0.060				
R2					0.451	0.463	0.429	0.435

Robust standard errors are reported within parenthesis
*, ** and *** denote parameter estimates different from zero at the 10, 5 and 1 % significance level, respectively

conclusion is consistent with evidence from prior studies investigating the combined impact of these effects (Nunn and Trefler 2008; Corcos et al 2013).

The *HQCON* parameter estimates are negative, thereby validating the prediction that foreign legislative institutions in support of headquarter contracting reduce firm incentives to integrate assembly production in the source country. The result reinforces those obtained in firm estimations, providing stark support of the Antràs and Helpman (2008) model's property-rights foundation. The positive *DEV* parameter estimate confirms the prediction that firms are less prone to integrate production in less developed, low-wage, source destinations. The result is consistent with the reference models' implication that the outsourcing-offshoring productivity threshold is increasing in the foreign wage and our prior result that the integration-offshoring productivity threshold is decreasing in the foreign wage.

5 Conclusions

This paper provides a thorough investigation of predictions from recent theories using a property-rights approach to explain the firm's foreign internalization decision (Antràs and Helpman 2004, 2008). We investigate determinants of the offshoring firm's two-dimensional choice of ownership and location using direct data on firm behavior. Since prior empirical work by Bernard et al. (2010) and Corcos et al. (2013) suggests that the firm's foreign internalization decision may not be properly proxied by its intra-firm trade shares, our empirical research is needed to verify whether prior results based on intra-firm trade can be validated. Furthermore, our access to census data on (Swedish) manufacturing firms allows us to contribute to prior empirical research on the topic by providing evidence that is not subject to sample selection effects. In conclusion, our estimations provide general support of the reference models' predictions in line with prior evidence on the topic.

Our estimation results provide stark evidence in favor of the predicted impact of firm heterogeneity and foreign contracting enforcement that lies at the heart of the reference theories. Firms in industries with larger within-industry firm variation, as depicted by the variation in TFP levels, are more prone to integrate offshored production. Also, a larger (smaller) share of offshoring firms in the industry integrate under better foreign legislative support of assembly (headquarter) provision. That firm heterogeneity creates winners in foreign markets aligns our results to a large literature relying on Melitz' (2003) analysis of firm behavior. In addition, our results on foreign contracting enforcement validate the property-rights foundation of our reference models.

We also directly investigate individual firm behavior to identify if underlying model assumptions can be validated by the data. In particular, our evidence confirms that more productive offshoring firms integrate in the source country. Since we access census data, the result extend the support of the models' productivity ranking beyond that provided in prior research studies based on survey and/or size-restricted firm data. In addition, we provide the novel contribution of investigating

if productivity threshold implications of foreign contracting enforcement can be verified. The results provide strong support of the Antràs and Helpman (2008) prediction that foreign legislative institutions in support of headquarter provision reduces offshoring firms' integration productivity threshold. This evidence confirms that relationship-specificity in headquarter production makes the provision of these activities sensitive to foreign legislative institutions.

Our findings also reveal that firms' integration productivity threshold is lower in more developed offshoring destinations, which could reflect the Antràs (2005a) argument that the location decision of headquarter-intensive firms is less sensitive to low-wage cost incentives. That our evidence is in line with this self-selection effect gives additional support of that property-rights effects play a fundamental role for the firm's foreign internalization decision. This conjecture is strengthened by additional results from our firm estimations showing that firms with larger research and capital intensities of production are more prone to integrate, which is consistent with the implication that firm gains from integration are starker in headquarter-intensive production and the Antràs' (2003) prediction that lower integration transaction costs are incurred in capital-intensive production.

Acknowledgements The author thanks Peter Vikström at Growth Policy Analysis for enabling access to confidential firm-level data and Markus Lindvert at Growth Policy Analysis for assistance in compiling the firm-level data set. Special thanks goes to two anonymous referees for useful comments, Nils-Åke Gustafson for general data assistance and Fredrik Wilhelmsson for empirical advice. Research in this paper is part of the author's ongoing work on the topic of firm behavior and has in this respect benefited from the close cooperation with colleagues at Department of Economics, Lund University.

Appendix

Table 3 Descriptive statistics, firm-location sample

Variable	Nobs	Min	Max	Mean	STD
INT(I)	311,761	0.000	1.000	0.131	0.337
INT(II)	184,715	0.000	1.000	0.155	0.361
PROD(OP)	306,933	−9.426	4.501	0.018	0.494
PROD(Aw)	307,178	−11.118	8.437	0.149	0.648
RINT	307,689	0.000	1.000	0.005	0.030
CINT	311,510	−3.745	15.317	6.436	0.791
ASCON	268,927	−0.028	0.921	0.733	0.147
HQCON	291,094	−0.028	0.923	0.628	0.212

Table 4 Descriptive statistics, industry-location sample. All variables are transformed with the log-linear specification used in Heckman estimations except for the *ACINT* and *DEV* variables (that already are in log form)

Variable	Nobs	Min	Max	Mean	STD
INT S (I)	64,248	0.000	1.000	0.119	0.255
INTS (II)	46,396	0.000	1.000	0.145	0.286
PDISP (OP)	118,021	0.000	43.306	10.383	7.086
PDISP (Aw)	118,021	0.000	57.329	14.435	9.350
ARINT	118,018	0.000	0.445	0.005	0.013
ACINT	118,010	4.115	9.236	6.400	0.452
ASCON	96,604	−0.015	0.460	0.292	0.095
HQCON	108,652	−0.015	0.462	0.253	0.112
DEV	111,472	5.222	11.292	9.398	1.069

References

Antràs, P. (2003). Firms, contracts and the product cycle. *Quarterly Journal of Economics, 118*(4), 1375–1418.

Antràs, P. (2005a). Property rights and the international organization of production. *American Economic Review, 95*(2), 25–32.

Antràs, P. (2005b). Incomplete contracts and the product cycle. *American Economic Review, 95*(4), 1054–1073.

Antràs, P. (2014). Grossman-Hart (1986) goes global: Incomplete contracts, property rights, and the international organization of production. *Journal of Law, Economics and Organization, 30*, i118–i175.

Antràs, P., & Helpman, E. (2004). Global sourcing. *Journal of Political Economy, 112*(3), 552–580.

Antràs, P., & Helpman, E. (2008). Contractual frictions and global sourcing. In E. Helpman, D. Marin, & T. Verdier (Eds.), *The organisation of firms in a global economy*. Harvard: Harvard University Press.

Antràs, P., & Rossi-Hansberg, E. (2009). Organizations and trade. *Annual Review of Economics, 1*, 43–64.

Aw, B. Y., Chung, S., & Roberts, M. J. (2003). Productivity, output, and failure: A comparison of Taiwanese and Korean manufacturers. *Economic Journal, 113*, 485–510.

Bernard, A. B., Jensen, J. B., Redding, S., & Schott, P. K. (2010). Intrafirm trade and product contractibility. *American Economic Review, 100*(5), 444–448.

Corcos, G., Irac, D. M., Mion, G., & Verdier, T. (2013). The determinants of intra-firm trade: Evidence from French firms. *Review of Economics and Statistics, 95*(3), 825–838.

Defever, F., & Toubal, F. (2007). *Productivity and the sourcing modes of multinational firms: Evidence from French firm-level data*. CEP Discussion Paper No. 842.

Federico, S. (2010). Outsourcing versus integration at home and abroad and firm heterogeneity. *Empirica, 37*, 47–63.

Feenstra, R. C., & Hanson, G. H. (1996). Globalization, outsourcing, and wage inequality. *American Economic Review, 86*, 240–245.

Feenstra, R. C., & Hanson, G. H. (1999). The impact of outsourcing and high technology capital on wages: Estimates for the United States, 1979–1990. *Quarterly Journal of Economics, 114*(3), 907–940.

Helpman, E. (2006). Trade, FDI and the organisation of firms. *Journal of Economic Literature, 44*(3), 589–630.

Jabbour, L. (2012). 'Slicing the value chain' internationally: Empirical evidence on the offshoring strategy by French firms. *World Economy, 35*(11), 1417–1447.

Kaufmann, D., Kraay, A., & Zoido-Lobaton, P. (1999). *Aggregating governance indicators*. World Bank Policy Research Working Paper Series No. 2195.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). *Governance matters VIII: Aggregate and individual governance indicators, 1996-2008*. World Bank Policy Research Working Paper Series No. 4978.

Kohler, W., & Smolka, M. (2009). *Global sourcing decisions and firm productivity: Evidence from Spain*. CESinfo Working Paper No. 2903.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *Quarterly Journal of Economics, 122*(2), 569–600.

Nunn, N., & Trefler, D. (2008). The boundaries of the multinational firm: An empirical analysis. In E. Helpman, D. Marin, & T. Verdier (Eds.), *The organisation of firms in a global economy*. Harvard: Harvard University Press.

Olley, S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*, 1263–1297.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics, 48*(1), 7–35.

Spencer, B. J. (2005, November). International outsourcing and incomplete contracts. *Canadian Journal of Economics*, *38*(4), 1107–1135.

Yeaple, S. R. (2006). Offshoring, foreign direct investment, and the structure of U.S. trade. *Journal of the European Economic Association, 4*, 602–611.

Williamson, O. E. (1975). *Markets, hierarchies: Analysis, antitrust implications*. New York: Free Press.

Williamson, O. E. (1985). *The economic institutions of capitalism*. New York: Free Press.

Multinationals and Corruption: Business as Usual?

Ayça Tekin-Koru

Abstract This paper disentangles the effects of corruption on entry mode decision by carrying out an empirical analysis with rich, firm-level data on the activities of Swedish MNCs around the globe in manufacturing sectors from 1987 to 1998. A number of propositions emerge from a simple theoretical framework. The panorama of the results from the empirical part supports most of these propositions: Entry mode decision of an MNC is a complex one and there are many asymmetries involved when it comes to the impact of corruption on this decision. First, greenfield investments are always discouraged by higher levels of corruption. This is more so for firms with high levels of mobile skills. Second, M&As are encouraged by moderate levels of corruption. For firms with high levels of non-mobile skills this effect is stronger. However, when corruption levels are beyond a certain threshold, M&As are deterred as well. Third, firms with a wider network of foreign affiliates are more immune to the effects of corruption, whereas small, single affiliate firms are severely affected. These results confirm the findings of the recent literature and add to it by testing a number of extensions of this view.

1 Introduction

While foreign direct investment (FDI) flows have the potential to make significant contributions to economic and social development, there exist widespread perception and anecdotal evidence that these flows are often restricted by corrupt practices of local or national government officials in different countries around the globe. In effect corruption acts as a barrier to entry by multinational corporations (MNCs) into new markets.

Corruption is often defined as the misuse of public power for private benefit which includes bribing of the public officials, kickbacks in public procurement and the misappropriation of public funds. Corruption need not involve money changing hands; it may be observed in the form of "trading influences" or granting favors.

A. Tekin-Koru (✉)
Department of Business Administration, TED University, Ziya Gokalp Bulvari, No: 48, Cankaya, Ankara, Turkey
e-mail: ayca.tekinkoru@tedu.edu.tr

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_7

The *level of corruption* has two dimensions: The frequency of corrupt undertakings and the total value of bribes paid—or the magnitude of influences traded—which go hand in hand, i.e. in countries where bribery is the rule of the game, more than a trifling proportion of firm revenues tends to represent the bribes paid.

In 2012, Transparency International conducted a survey on 105 MNCs which are worth more than $11 trillion. These firms touch the lives of millions of people across the globe. Of the 105 companies surveyed in the TI report, 50 do not disclose revenue/sales in any country of foreign operations, 85 do not disclose income tax in any country of foreign operations and 39 do not disclose any financial data (tax, revenue, sales, pre-tax income, capital investment, community contributions) in their countries of operation. Under these circumstances, it becomes absolutely necessary to rethink the effects of corruption not only on the national firms, but also on the FDI flows channeled through different modes of entry with mode-specific consequences for the countries hosting considerable amounts of FDI.

MNCs undertake foreign direct investment in different ways: Cross-border M&As, greenfield investments, joint ventures, partial acquisitions, and different forms of low-equity commitment such as sales offices, licensing, research centers, etc.[1] Except for a few recent studies, trade literature has considered the effects of corruption mainly in the context of whether or not to produce overseas, but has not differentiated specific types of FDI and therefore treated the different entry modes as perfect substitutes. However, for most firms seeking foreign market access, cross-border acquisitions and greenfield investments represent unlikely candidates for perfect substitution. While acquisitions provide rapid access to a foreign market with increased market power and a means of exploiting synergies—derived from the non-mobile skills such as knowledge of the local conditions—between buyer and seller firms, greenfield investments offer the most profitable internal utilization of mobile firm-specific assets (R&D, marketing expenditures, scientific and technical workers, product newness and complexity and product differentiation) for reasons including moral hazard and technology diffusion.

There has been a systematic work on the effects of corruption on FDI flows using aggregate data as explained in the next section. However, exploration of corruption's effect on FDI using firm level data is much rarer due to the lack of systematic data on corruption and scarcity and confidentiality of firm-level data on FDI until recently.

The objective of this paper is to add to this newly developing literature by offering a theoretical analysis that takes into consideration not just how corruption may restrict FDI flows, but also how corruption may facilitate these flows and then testing the findings of the theory on-site in Sweden by employing rich, firm-level data on the activities of Swedish MNCs around the globe in manufacturing sectors from 1987 to 1998.

First, I offer a simple theoretical framework to motivate the empirical analysis. Six propositions are generated. The first two are related to the direct (primary) effect

[1] Due to lack of data, the middle ground between wholly owned operations and no entry could not be included in the analysis in this paper.

of corruption on the mode of foreign expansion: (1) Higher levels of corruption discourage greenfield investments and (2) under certain circumstances encourage cross-border M&As. The last four propositions and their corollaries address the indirect (secondary) effects: (3) MNCs with higher endowments of mobile skills are more likely to invest in the host country. This likelihood is dampened the higher the degree of corruption in the host country. The dampening effect is stronger for M&As than greenfield investments. (4) MNCs with higher endowments of non-mobile skills are more likely to invest in the host country. This likelihood is magnified the higher the degree of corruption in the host country. The magnification is stronger for M&As than greenfield investments.

Main innovations present in the empirical part are as follows: First, I include both foreign access strategies (cross-border M&As and greenfield investments) in the analysis, which differs from many studies that include one of the strategies at a time. Second, I consider not only the direct effects of corruption on the modes of foreign expansion but also the indirect effects by allowing the corruption tolerance of a MNC vary with its skill set. Third, I apply the bivariate probit model to account for the correlation between different entry strategies, which reduces the inconsistency of the estimators significantly.

Results of the empirical analysis show that corruption reduces the likelihood of foreign entry as conjectured by recent studies. Entry mode decision of an MNC is a complex one and there are many asymmetries involved when it comes to the impact of corruption on this decision. First, greenfield investments are always discouraged by higher levels of corruption. This is more so for firms with high levels of mobile skills. Second, M&As are encouraged by moderate levels of corruption. For firms with high levels of non-mobile skills this effect is stronger. However, when corruption levels are beyond a certain threshold, M&As are deterred as well. Third, firms with a wider network of foreign affiliates are more immune to the effects of corruption, whereas small, single affiliate firms are severely affected. These results confirm the findings of the recent literature and add to it by testing a number of extensions of this view.

The paper continues as follows: In the next section, I present the related recent literature. In Sect. 3, I lay out a simple model and present the testable propositions generated from it. In Sect. 4, I discuss the econometric analysis. Sections 5 reports the results and I conclude in Sect. 6.

2 The Recent Literature

There are two strands of literature that are relevant to the analysis of corruption and entry modes of multinational corporations. The first one is the literature on foreign direct investment, too vast to be addressed here at length (Markusen 2002 and citations therein). The second line of literature relevant for this paper is voluminous as well and considers the causes and consequences of corruption in general. The work spans many different areas such as the impact of corruption on regulatory

discretion, existence of rents and opportunities for rent-seeking, and civil service wage policy. There are also studies that consider the impact of corruption on various aspects of economic systems such as growth, military expenditure and procurement, delivery of public services, and inequality. Excellent surveys of this literature are provided by Jain (2001) and Aidt (2003).

There is a fast growing literature on corruption and FDI connection. The existing work is mostly empirical and can be summarized under two headings: (1) Corruption acting as the "grabbing hand" in FDI: Corruption in a host country introduces additional direct or indirect costs for the foreign investor and therefore makes FDI less likely. This argument finds support in the works of Hines (1995), Wei (2000), Hellman et al. (2002), Habib and Zurawicki (2002), Busse and Hefeker (2007), Hakkala et al. (2008) and Javorcik and Wei (2009). (2) Corruption acting as the "helping hand" in FDI: By greasing squeaky wheels of an inefficient bureaucracy and softening rigid regulations, corruption can reduce the obstacles in front of FDI and thus will not necessarily discourage it. The works of Lui (1985), Wheeler and Mody (1992), Saha (2001), Egger and Winner (2005), Bjorvatn and Soreide (2005), Wu (2006), Tekin-Koru (2006) and Barassi and Zhou (2012) are examples of this line of study.

The current paper builds on this earlier empirical work examining the effects of corruption on FDI. Many of these studies use aggregate cross-country data. The fewer ones using firm-level data usually take steps beyond investigating the effect of corruption on FDI. Some concentrate of the ownership structure of the firm and some on the different types of affiliate activity in host country and how they are affected by corruption in the host country or by the corruption distance between the parent and host countries. Two such recent studies which are more relevant for the current paper are contributions made by Javorcik and Wei (2009) and Hakkala et al. (2008).

Javorcik and Wei (2009) investigate how the volume of FDI and its ownership structure may be affected by the extent of corruption. They use firm-level data from 22 transition economies and find that corruption adversely affects the probability of foreign investments taking place in the host country. However, conditional on FDI taking place, their results suggest that joint ventures are more likely in corrupt environments unless the technological sophistication of the foreign firm is high. Different from Javorcik and Wei (2009) who bundle acquisitions and greenfield investments together as sole ownership, in the current paper I treat them differently since the motives for undertaking an M&A and a greenfield investment are not the same. Moreover, in the current paper data on global operations of Swedish multinational firms are used which provides a broader coverage of host countries.

The most relevant work for the current paper is by Hakkala et al. (2008) who consider the impact of corruption on FDI in two dimensions: (1) corruption may influence the probability that a firm chooses to invest in a foreign market but not the size of the affiliate activities once the investment is undertaken, and (2) corruption may have different effects on different types of affiliate activities such as horizontal, vertical or export platform sales. Using Swedish firm-level data, Hakkala et al. find that Swedish MNCs are less likely to invest in corrupt countries. In their detailed

analysis, horizontal investments are deterred by high levels of corruption, however, there is no robust effect of corruption on vertical or export platform sales. This asymmetry is explained by greater costs incurred by the firm in case of production for local sales rather than production for exporting to other markets. They also find that R&D intensive or large firms weather corruption better compared to less technical or smaller firms. The current paper complements (Hakkala et al., 2008) by using the same Swedish multinational firm data to examine another aspect of FDI, namely the effect of corruption on M&As and greenfield investments.

3 Corruption and Foreign Entry: A Theoretical Framework

This section presents a model of the effects of corruption in countries on the choice of FDI, the mode-selection of FDI, and on the levels of FDI. Section 3.1 lays out the assumptions, followed by a sequential entry game involving Nash-Bargaining in Sect. 3.2. Finally, the complex effects of host country corruption levels on FDI decisions of the MNCs are explored in Sect. 3.3.

3.1 Assumptions

A MNC from the parent country (P) considers entering the host country market, (H). It can choose between building its own establishment (greenfield investments, g) or to acquire an already existing indigenous firm (mergers and acquisition (M&A), m). The outside alternative for the MNC is not to enter at all (n). Thus, $\Sigma = \{m, g, n\}$ represents the set of possible entry strategies (s). I assume, for the sake of simplicity, that exporting to H is not a feasible option due to transport cost reasons.[2]

Preferences. There exist two final goods sectors; X (increasing returns, imperfect competition) and Y (constant returns, perfect competition). Good Y is produced from a single factor L (Labor), where one unit of L produces one unit of Y. Good X, on the other hand, is produced using firm specific assets and factor L, both in fixed proportions. The linear demand functions are derived from the quasi-linear utility function maximized subject to a budget constraint. Income is derived from labor and profits.

$$\max U = \alpha X - \left(\frac{\beta}{2}\right) X^2 + Y \tag{1}$$

$$\text{subject to } L + \Pi = Y + pX$$

[2]The model could be broadened in such a way that the firm's choice extends to serving country H by exporting, which does not alter the principal insights. See Tekin-Koru (2012) for a model with exporting as an alternate way of serving the potential host country.

where L and Y are numéraires. The inverse demand function for good X is as follows:

$$P = \alpha - \beta X \tag{2}$$

Firms. I assume that there are two (potential) firms producing X in country H, one from parent country p and one indigenous to host country h.

Firms differ in their skills. As in Nocke and Yeaple (2004) there are two types of firm-specific skills: mobile (μ) and non-mobile (ν). The efficiency of a firm's production technology is assumed to travel internationally at little to no cost. On the other hand, the degree of familiarity with the local business conditions is assumed immobile. These non-mobile skills, including but not limited to the degree of influence on political process and of the strength of ties with local bureaucracy, marketing strategies geared towards the expectations of the host country and greater access to distribution channels affect both marginal production costs and the fixed entry costs as does the level of corruption (ζ) in country H. Variable production costs of firms p and h are given by c_p and c_h:

$$c_p(s) = \begin{cases} c(\mu_p, \nu_p, \zeta) & \text{if } s = g \\ c(\mu_p, \nu_h, \zeta) & \text{if } s = m \end{cases} \tag{3}$$

$$c_h(s) = c(\mu_h, \nu_h, \zeta) \qquad \text{for all } s \tag{4}$$

Due to investment in R&D and long term diverse experience in managerial practice, which are internationally mobile, firm p is endowed with an alternative technology which allows it to have greater cost advantages in producing good x compared to firm h in cases of both M&As and greenfield investments. On the other hand, if greenfield investment is the chosen mode of entry, then firm h enjoys greater cost savings due to being better acquainted with the local business conditions. Since firm p has access to the non-mobile skills of the indigenous firm (ν_h) in an M&A regime, the size of the cost savings is as big as the ones enjoyed by firm h in greenfield and no entry regimes. Higher levels of corruption (ζ) in country H will increase the variable costs of production at an increasing rate. In countries with widespread corruption, for example, the variable costs may reach prohibitively high levels.[3]

I also consider the interactions of firm-specific skills with the level of corruption in country H. I assume that high degrees of mobile skills make the MNCs less corruption tolerant. Thursby and Thursby (2006) in their study of more than 200 multinational companies across 15 industries, mostly headquartered in the

[3]The variable productions costs are concave in both mobile (μ) and immobile skills (ν) and convex in the level of corruption (ζ): $\frac{\partial c}{\partial \mu_p} < \frac{\partial c}{\partial \mu_h} < 0$, $\frac{\partial c}{\partial \nu_h} < \frac{\partial c}{\partial \nu_p} < 0$, $\frac{\partial c}{\partial \zeta} > 0$ and $\frac{\partial^2 c}{\partial \mu_p^2} = \frac{\partial^2 c}{\partial \mu_h^2} < 0$, $\frac{\partial^2 c}{\partial \nu_h^2} = \frac{\partial^2 c}{\partial \nu_p^2} < 0$, $\frac{\partial^2 c}{\partial \zeta^2} > 0$
Moreover, the variable production costs are assumed to be submodular in their arguments.

United States and Western Europe, find that only 22 % of the R&D effort in emerging countries is for new science. In other words, companies keep their most cutting-edge research in developed countries where intellectual property protection is the strongest. Therefore, one can argue that mobile skills get less beneficial for the firm as corruption increases.[4]

On the other hand, firms may adopt a "when in Rome, do as Romans do" strategy in their foreign operations. This can translate into non-mobile skills, such as local connections, becoming much more valuable in the existence of corruption. There are many anecdotes to the effect. When Candy, the Italian home appliances manufacturer, decided to expand into Russia in 2005 and open the company's first plant there, the company had to deal with extreme difficulties for the company's trucks to even make it across the border. After struggling with the problem for weeks, Candy decided to set up a division of the company in Moscow that would have the sole responsibility of dealing with customs clearance. The problem was solved in days. Another high-profile example is the recent scandal of GlaxoSmithKline, the British drug giant, which involves using bribes, kickbacks and other fraudulent means by the Chinese executives of the company to bolster drug sales in China over 2007–2013 period. Therefore, in the light of these anecdotal evidence, it may not be wrong to assume that high degrees of non-mobile skills make the MNCs more corruption tolerant.[5]

In addition to the variable costs, the multinational firm incurs fixed entry costs as well. First, there is a fixed greenfield establishment cost ($F(g)$). Second, there is an M&A cost ($F(m)$) if the indigenous firm is acquired.

Last but not least, there is a bureaucratic cost of entry ($F^b(s) = F^b(v_i, \zeta)$ for $i = \{h, p\}$) in the host country, H. This cost, F^b essentially measures the procedures, time, cost and paid-in minimum capital required for a firm to start-up and formally operate in the host country. The regulation of entry enables the regulators to collect bribes from the potential entrants and serves no social purpose. Therefore, in the model, it is assumed that F^b increases in corruption. More extensive regulation should be associated with socially inferior outcomes, particularly corruption. Djankov et al. (2002) finds evidence supporting the public choice view that entry regulation benefits politicians and bureaucrats. Since it is possible to avoid some if not all of these barriers, a higher degree of familiarity with the local business/governance conditions will help reducing these costs by itself and will do even more so in more corrupt environments.[6]

[4]The cross derivatives of cost functions with respect to mobile skills and the level of corruption are as follows: $\frac{\partial^2 c}{\partial \mu_p \partial \zeta} > 0$, $\frac{\partial^2 c}{\partial \mu_h \partial \zeta} > 0$.

[5]The cross derivatives of cost functions with respect to non-mobile skills and the level of corruption are as follows: $\frac{\partial^2 c}{\partial v_h \partial \zeta} < 0$, $\frac{\partial^2 c}{\partial v_p \partial \zeta} < 0$.

[6]The official costs of entry are concave in immobile skills (v) and convex in the level of corruption (ζ): $\frac{\partial F^b}{\partial v_h} < \frac{\partial F^b}{\partial v_p} < 0$, $\frac{\partial F^b}{\partial \zeta} > 0$ and $\frac{\partial^2 F^b}{\partial v_h^2} = \frac{\partial^2 F^b}{\partial v_p^2} > 0$, $\frac{\partial^2 F^b}{\partial \zeta^2} > 0$. The cross derivatives $\frac{\partial^2 F^b}{\partial v_h \partial \zeta}$ and $\frac{\partial^2 F^b}{\partial v_p \partial \zeta}$ are negative by the same logic used in variable costs.

3.2 The Game

Firm p maximizes its profits in the host country through its choice of entry mode and the quantity supplied.[7] In the first stage, the MNC chooses its entry mode and in the second stage makes its quantity decision in a usual Cournot setting by taking the entry mode from the previous stage as given.

A strategy for firm p has two elements: (1) the firm's entry mode choice, $s \in \Sigma$ where $\Sigma = \{m, g, n\}$ is the set of all possible entry nodes and (2) the firm's quantity choice, $x^p(s)$ where $x^p(s) > 0$ indicates that firm p is active in the host country; $x^p(s) = 0$ indicates that firm p chooses n and thus not to produce in country H.

Aggregate supply to the consumers by firms $i = \{h, p\}$ in the host country given the entry mode choice s, is:

$$X(s) = \sum_i x_i(s), \quad i = \{h, p\} \tag{5}$$

and the aggregate profits generated under each entry strategy s for firms p and h in country H are given by

$$\begin{aligned} \Pi_p(s, x_p(s)) &= [(P - c_p(s))x_p(s)] - F(s) - F^b(s) = \pi_p(s, x_p(s)) - F(s) - F^b(s) \\ \Pi_h(s, x_h(s)) &= [(P - c_h(s))x_h(s)] = \pi_h(s, x_h(s)) \end{aligned} \tag{6}$$

where $x_i(s)$, $c_i(s)$, $F(s)$ and $F^b(s)$ are the quantity choice, variable cost of production for each firm i, fixed costs associated with each entry mode s and bureaucratic costs of entry. π_i signifies the operating profits of firm $i = \{h, p\}$. The M&A cost $F(m)$ is endogenized through a simple bilateral Nash bargaining process.

Bargaining In this game, the acquiring firm (firm p) and the target firm (firm h) seek to split a total value $\pi_p(m) - F^b(m)$ which they can achieve if and only if they agree on a specific division. If there is no agreement between firm h and firm p, the latter opts for the next best alternative among no entry and greenfield investment. The following payoffs, thus, can be called backstop payoffs and be signified by a tilda:

$$\begin{aligned} \tilde{\Pi}_p &= \max\left\{\Pi_p(g), \Pi_p(n)\right\} \\ \tilde{\Pi}_h &= \max\left\{\Pi_h(g)\,|_{\tilde{\Pi}_p = \Pi_p(g)}, \Pi_h(n)\,|_{\tilde{\Pi}_p = \Pi_p(n)}\right\} \end{aligned} \tag{7}$$

For the solution of this bargaining procedure, it should be assumed that there is a positive surplus ($\pi_p(m) - F^b(m) - \tilde{\Pi}_p - \tilde{\Pi}_h > 0$) from agreement. If this were not

[7] This paper explores the profit maximization of firm p in the host country in isolation and does not take into account the multinational's profit maximization neither in the parent country and nor in its broader worldwide network due to the much needed simplicity in this highly non-linear model.

the case, the whole bargaining process would be unlikely because each side would just take up its outside opportunity and receives its backstop payoff. Next, consider the following rule coming from the solution of bilateral Nash-bargaining process.

Criterion 1 *Given* $\theta \in [0, 1]$ *each party is to be given its backstop payoff plus a share of the surplus, a fraction* θ *for firm p and a fraction* $(1-\theta)$ *for firm h.*

All bargaining scenarios have two things in common: There is always a surplus after reaching an agreement and it is not a zero-sum game. This is true for the multinational and the local firm as well. First, if the total payoff from an M&A is not greater than the status quo neither the MNC nor the local firm will engage in such fruitless negotiations; ergo there is a surplus. Second, even though both the MNC and the local firm try to get more for themselves and leave less for the other party, if no agreement is reached no one will get any surplus at all, therefore, both sides win when an agreement is reached. The above stated criterion formalizes these two issues to have a solution to the Nash's cooperative game approach to bargaining.

Writing $\Pi_p(m)$ and $\Pi_h(m)$ for the amounts that firm p and firm h receive, the above stated bargaining criterion can be translated as

$$\begin{aligned} \Pi_p(m) &= \tilde{\Pi}_p + \theta(\pi_p(m) - F^b(m) - \tilde{\Pi}_p - \tilde{\Pi}_h) = \pi_p(m) - F^b(m) - F(m) \\ \Pi_h(m) &= \tilde{\Pi}_h + (1-\theta)(\pi_p(m) - F^b(m) - \tilde{\Pi}_p - \tilde{\Pi}_h) = F(m) \end{aligned} \tag{8}$$

Next, define the reservation price of the buying party as $R_p = \pi_p(m) - F^b(m) - \tilde{\Pi}_p$ and that of the selling party as $R_h = \tilde{\Pi}_h$. The reservation price of the multinational is the profits of the multinational after the M&A minus its next best alternative, i.e. the maximum of its greenfield or no entry. The reservation price of the local firm is the local firm's profits if no agreement is reached. Then, one can arrive at the cost of M&A by solving the equations in (7) for $F(m)$:

$$F(m) = (1-\theta)R_p + \theta R_h \tag{9}$$

When $\theta = 1$, firm p has all the bargaining power implying that $F(m) = R_h$. When $\theta = 0$, on the other hand firm p has no bargaining power and thus the cross-border M&A price is the same as its reservation price, i.e. $F(m) = R_p$.

Equilibrium The game is solved in the usual logic of backward induction. I seek the subgame perfect equilibrium of this game. The second stage of the game involves the product market where firms compete á la Cournot. The equilibrium output levels and total profits of all firms are reported in the Appendix. Production and sales take place with firms moving simultaneously. The game is solved for Nash equilibria in pure strategies. Each equilibrium point is assumed to have equal probability.

Criterion 2 *Denote by* $X_p(s)$ *the set of possible quantity choices for form p in the host country market given entry mode choice s. The Nash equilibrium for the*

second-stage quantity sub-game for any s is the quantity choice $x_p^*(s)$ *such that:*

$$\Pi_p(s, x_p^*(s)) \geq \Pi_h(s, x_p^*(s), x_p^{*-}(s)) \quad \text{for all } x_p(s) \in X_p(s) \tag{10}$$

Denote by $\Pi_p^*(s^*)$ *the profit to firm p from the Nash equilibrium quantity choice corresponding to the entry mode choice s. An equilibrium for the first-stage entry game is an entry strategy* s^* *such that:*

$$\Pi_p^*(s^*) \geq \Pi_p^*(s, s^{*-}) \quad \text{for all } s \in \Sigma \tag{11}$$

The MNC makes its entry decision in the following way. In the case where the best alternative to a negotiated agreement is no entry, firm p chooses cross-border M&A over no entry if the payoff from cross-border M&A is higher than zero. On the other hand, in the case where the best alternative is greenfield investment, firm p chooses cross-border M&A if the payoff from M&A is higher than that from greenfield investment. The final decision in regard to entry mode, henceforth, will be the outcome of the relative magnitudes of and the interrelations among country size, level of corruption in the country, firms' relative endowments of mobile and immobile skills, and the relative bargaining power of the parties.

Considering the current setup of the model, it would be natural to expect a negative impact of host country corruption on the FDI modes of entry. In other words, as the corruption level of the host country increases both the variable production costs and initial start-up costs will get higher and thus discourage the MNC from investing in the host country all together. However, this argument ignores both the changes in the acquisition price—which is endogenized in this model—with respect to changes in corruption and the importance of the MNC's relative endowment of mobile and non-mobile skills. The *corruption tolerance*—how well the firm copes with the potential negative impact of corruption on profitability—of a MNC with high levels of mobile skills will be quiet low compared to a MNC endowed with not to so strong mobile skills but with impressive levels of non-mobile skills. Therefore, to formalize this discussion I use comparative statics in the next section.

3.3 Comparative Statics

In this section, I analyze the effects of corruption on the entry mode decision of a multinational firm to generate testable hypotheses.

3.3.1 Corruption

What is the impact of host country corruption on the equilibrium patterns of greenfield investments, M&As and no entry at all? To answer this question, I

compare the effects of corruption level (ζ) in country H on the payoffs of firm p from different entry strategies.

For notational convenience I will henceforth use $\Pi_p(s)$ for $\Pi_p(s, x_p^*(s))$. Before the total derivative of $\Pi_p(s)$ for $\forall s \in \Sigma$ with respect to ζ is calculated two cases should be differentiated: The case where greenfield investment is the next best alternative to M&A and the case where no entry is the next best. The latter is trivial as both variable production costs and bureaucratic costs of entry will be higher for higher corruption levels. Therefore, I continue with the more complex case where greenfield investment is the next best alternative to M&A.

Proposition 1 *More corruption in the host country reduces the likelihood of greenfield investments.*

Proof The payoff from greenfield investment to firm p is

$$\Pi_p(g) = \pi_p(m) - F^b(g) - F(g) \tag{12}$$

First, take the total derivative of $\Pi_p(g)$ with respect to ζ and then substitute the explicit forms of the payoff functions given in Appendix in the general form equations. Finally by applying the assumption that $\frac{\partial c_p}{\partial \zeta} = \frac{\partial c_h}{\partial \zeta}$, arrive at

$$\frac{d\Pi_p(g)}{d\zeta} = -\frac{2}{3\beta}\frac{\partial c_p}{\partial \zeta} - \frac{\partial F^b(g)}{\partial \zeta} \tag{13}$$

Given that $\frac{\partial F^b(s)}{\partial \zeta} > 0$ and $\frac{\partial c_p}{\partial \zeta} > 0$ for $s \in \{m, g\}$ the above derivative is negative. ■

In this model, greenfield investments are affected by host country corruption through operating costs and bureaucratic costs of entry. As the level of corruption in the host country increases, MNC's profits from a greenfield investment project decline due to rising operating costs (e.g., costs of dealing with red-tape in day-to-day activities) and rising start-up costs (e.g., long waiting times for permits). This in turn reduces the payoff and eventually the likelihood of greenfield investments in corrupt host countries. This result is in line with the existing literature on the effects of corruption on FDI. Next, consider the M&A case.

Proposition 2 *More corruption in the host country (i) reduces the likelihood of M&As when the degree of corruption is low; (ii) increases the likelihood of M&As when the degree of corruption is high.*

Proof The payoff from M&A to firm p is given by

$$\begin{aligned} \Pi_p(m) &= \pi_p(m) - F^b(m) - F(m) \\ \text{where } F(m) &= (1-\theta)[\pi_p(m) - F^b(m) - \Pi_p(g)] + \theta\,\Pi_h(g) \end{aligned} \tag{14}$$

Notice that the cost of acquisition is determined by the other parameters of the model. Take the total derivative of $\Pi_p(m)$ with respect to ζ

$$\frac{d\Pi_p(m)}{d\zeta} = \theta \frac{d\pi_p(m)}{d\zeta} + (1-\theta)\frac{d\Pi_p(g)}{d\zeta} - \theta \frac{d\Pi_h(g)}{d\zeta} - \theta \frac{\partial F^b(m)}{\partial \zeta} \tag{15}$$

Given the Cournot profits from the Appendix, it is trivial to find that $\frac{d\pi_p(m)}{d\zeta} = -\frac{x_p(m)}{\beta}\frac{\partial c_p}{\partial \zeta} < 0$ since $\frac{\partial c_p(m)}{\partial \zeta} > 0$. The second term, $\frac{d\Pi_p(g)}{d\zeta} < 0$ as proven above. The third term, $\frac{d\Pi_h(g)}{d\zeta} = -\frac{2x_h(m)}{3\beta}\frac{\partial c_h}{\partial \zeta} < 0$ since $\frac{\partial c_h(g)}{\partial \zeta} > 0$. The last term, $\frac{\partial F^b(m)}{\partial \zeta} > 0$.

(i) If $\left|\frac{d\Pi_h(g)}{d\zeta}\right|$ is sufficiently small, then the derivative in Eq. (15) becomes negative.

(ii) If $\left|\frac{d\Pi_h(g)}{d\zeta}\right|$ is sufficiently large, then the derivative in Eq. (15) becomes positive. ■

While there is a monotonically decreasing relationship between corruption in the host country and the likelihood of greenfield investments, this relationship becomes non-linear (U-shaped) for M&As. This stems from the assumption that the indigenous firm is also hurt by corruption and it works to the advantage of the MNC by reducing the acquisition price. In relatively less corrupt environments, the existing local firms are hurt less by corruption and this causes a discount in the M&A prices at a lesser degree. However, in highly corrupt environments, the damage caused by corruption on the profits of the local firm is high and this translates into a large price reduction in M&A negotiations. This price reduction then becomes the dominant factor and corruption increases the likelihood of M&As compared to the greenfield investments.

3.3.2 Corruption and Mobile Skills

I now turn to the issue that how the corruption tolerance of firm p changes as the level of mobile skills (μ_p) change. Specifically, how does the equilibrium pattern changes when the interaction of the mobile skills with the level of corruption is considered? As stated earlier, mobile skills provide production cost advantages to the MNC. However, the marginal impact of these skills gets lower in the existence of rising corruption levels for reasons such as poor intellectual property right (IPR) protection.

Proposition 3 *MNCs with higher endowments of mobile skills are more likely to conduct greenfield investments. This likelihood is dampened the higher the degree of corruption in the host country.*

Proof From the assumption that $\frac{\partial c}{\partial \mu_p} < 0$, the proof of the first part of the proposition is trivial.

For the second statement, consider the cross derivative of $\Pi_p(g)$ given in Eq. (12) with respect to μ_p and ζ :

$$\frac{d^2\Pi_p(g)}{d\mu_p d\zeta} = \frac{4}{9\beta^2}\frac{\partial c_p}{\partial \mu_p}\frac{\partial c_p}{\partial \zeta} - \frac{4}{3\beta}x_p(g)\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} \tag{16}$$

Given that $\frac{\partial c_p}{\partial \mu_p} < 0$ and $\frac{\partial c_p}{\partial \zeta} > 0$ and $\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} > 0$, the above derivative is negative. ■

Higher endowment of mobile skills give the MNC an advantage over local firms and thus guarantees a higher market share. Therefore, independent of the degree of corruption, mobile skills such as technological know-how or managerial ability increase the profits of the MNC in the host country. However, as the degree of corruption in the host country increases the usefulness of such skills deteriorate due to poor IPR protection or bureaucratic red-tape. Therefore, it is likely that MNCs with high levels of mobile skills may avoid investing in highly corrupt environments in the form of greenfield investments.

Proposition 4 *MNCs with higher endowments of mobile skills are more likely to conduct M&As. This likelihood is dampened the higher the degree of corruption in the host country.*

Proof From the assumption that $\frac{\partial c}{\partial \mu_p} < \frac{\partial c}{\partial \mu_h} < 0$, the proof of the first part of the proposition is trivial.

For the second statement, consider the derivative of $\frac{d\Pi_p(m)}{d\zeta}$ given in Eq. (15) with respect to μ_p. To derive an expression for it, consider the components of Eq. (15) one by one. The derivative of the first component $\frac{d\pi_p(m)}{d\zeta}$ with respect to μ_p is

$$\frac{d^2\pi_p(m)}{d\mu_p d\zeta} = \frac{2}{3\beta^2}\frac{\partial c_p}{\partial \mu_p}\frac{\partial c_p}{\partial \zeta} - \frac{1}{\beta}x_p(m)\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} \tag{17}$$

which is negative as $\frac{\partial c_p}{\partial \mu_p} < 0$, $\frac{\partial c_p}{\partial \zeta} > 0$ and $\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} > 0$. The derivative of the second component is already given in Eq. (16) which is proven to be negative. Next, consider the derivative of $\frac{d\Pi_h(g)}{d\zeta}$ with respect to μ_p

$$\frac{d^2\Pi_h(g)}{d\mu_p d\zeta} = -\frac{2}{9\beta^2}\frac{\partial c_p}{\partial \mu_p}\frac{\partial c_p}{\partial \zeta} + \frac{2}{3\beta}x_h(g)\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} \tag{18}$$

which is positive as $\frac{\partial c_p}{\partial \mu_p} < 0$, $\frac{\partial c_p}{\partial \zeta} > 0$ and $\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} > 0$. When all of the components are put together

$$\frac{d^2\Pi_p(m)}{d\mu_p d\zeta} = \frac{4}{9\beta^2}\frac{\partial c_p}{\partial \mu_p}\frac{\partial c_p}{\partial \zeta}(1+\theta) - \frac{1}{\beta}\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta}\left[\theta x_p(m) + \frac{2\theta}{3}x_h(g) + (1-\theta)x_p(g)\right] \tag{19}$$

Given that $\frac{\partial c_p}{\partial \mu_p} < 0$ and $\frac{\partial c_p}{\partial \zeta} > 0$ and $\frac{\partial^2 c_p}{\partial \mu_p \partial \zeta} > 0$, the derivative in Eq. (19) is negative. ■

Corollary 1 *The dampening effect is stronger for M&As than greenfield investments as long as the MNC has less bargaining power.*

Using the payoff functions in the Appendix, one can easily prove that $\frac{d^2\Pi_p(m)}{d\mu_p d\zeta} < \frac{d^2\Pi_p(g)}{d\mu_p d\zeta}$ if $\theta < \frac{1}{2}$ and $x_i(s)$ where $i = (p, h\}$.

In more corrupt environments, increases in both greenfield and M&A profits due to higher endowments of mobile skills will be lower than what they would be for lesser amounts of corruption. While mobile skills provide variable cost savings for firm p, increasing levels of corruption dampens the effectiveness of these cost savings. This in turn implies an advantage for the indigenous firm, which will exploit it at the negotiation table by demanding a higher acquisition price. Therefore, in case of an M&A, firm p will be hurt due to reduced effectiveness of its mobile skills coupled with an increase in the M&A entry cost. Moreover, if the MNC has a lower bargaining strength than the local firm, then this effect gets amplified.

3.3.3 Corruption and Non-mobile Skills

The analysis so far has highlighted the importance of increasing levels of corruption and the cross effects with mobile skills. I now investigate how the degree of non-mobile skills endowment with increasing levels of corruption affect the equilibrium mode of entry.

Proposition 5 *MNCs with higher endowments of non-mobile skills are more likely to conduct greenfield investments. This likelihood is magnified the higher the degree of corruption in the host country.*

Proof From the assumption that $\frac{\partial c}{\partial \mu_p} < 0$, the proof of the first part of the proposition is trivial.

For the second statement, consider the cross derivative of $\Pi_p(g)$ given in Eq. (12) with respect to ν_p and ζ :

$$\frac{d^2\Pi_p(g)}{d\nu_p d\zeta} = \frac{4}{9\beta^2}\frac{\partial c_p}{\partial \nu_p}\frac{\partial c_p}{\partial \zeta} - \frac{4}{3\beta}x_p(g)\frac{\partial^2 c_p}{\partial \nu_p \partial \zeta} - \frac{\partial^2 F^b(g)}{\partial \nu_p \partial \zeta} \tag{20}$$

Note that $\frac{\partial c_p}{\partial \nu_p} < 0$ and $\frac{\partial c_p}{\partial \zeta} > 0$ and $\frac{\partial^2 c_p}{\partial \nu_p \partial \zeta} < 0$ and $\frac{\partial^2 F^b(g)}{\partial \nu_p \partial \zeta} < 0$. Since c_p is assumed to be submodular in its arguments, $\left|\frac{\partial^2 c_p}{\partial \nu_p \partial \zeta}\right| > \left|\frac{\partial c_p}{\partial \nu_p}\frac{\partial c_p}{\partial \zeta}\right|$, the above derivative is positive. ■

Submodularity here means that the beneficial effect of non-mobile skills in corrupt environments is magnified compared to the direct positive effect of non-mobile skills or direct negative impact of corruption on firm profits. Higher endowment of non-mobile skills, in other words familiarity with local business conditions, increases the profits of the MNC in the host country. However, as the degree of corruption in the host country increases, such skills become more valuable because firms adopt a "when in Rome, do as Romans do" strategy in their foreign operations. Therefore, it is likely that MNCs with high levels of non-mobile skills may avoid investing in highly corrupt environments in the form of greenfield investments.

Proposition 6 *MNCs with higher endowments of non-mobile skills are more likely to conduct M&As. This likelihood is magnified the higher the degree of corruption in the host country.*

Proof From the assumption that $\frac{\partial c}{\partial \mu_p} < \frac{\partial c}{\partial \mu_h} < 0$, the proof of the first part of the proposition is trivial.

For the second statement, consider the derivative of $\frac{d\Pi_p(m)}{d\zeta}$ given in Eq. (15) with respect to $\nu_p = \nu_h$. Since firm p buys firm h, it adopts firm h's superior non-mobile skills. To derive an expression for it, examine the components of Eq. (15) one by one. The derivative of the first component $\frac{d\pi_p(m)}{d\zeta}$ with respect to ν_p is

$$\left.\frac{d^2\pi_p(m)}{d\nu_p d\zeta}\right|_{\nu_p=\nu_h} = \frac{2}{3\beta^2}\frac{\partial c_p}{\partial \nu_h}\frac{\partial c_p}{\partial \zeta} - \frac{1}{\beta}x_p(m)\frac{\partial^2 c_p}{\partial \nu_h \partial \zeta} \tag{21}$$

As long as c_p in its arguments the above derivative is positive. The derivative of the second component is already given in Eq. (20) which is proven to be negative. Next, consider the derivative of $\frac{d\Pi_h(g)}{d\zeta}$

$$\left.\frac{d^2\Pi_h(g)}{d\nu_p d\zeta}\right|_{\nu_p=\nu_h} = -\frac{2}{9\beta^2}\frac{\partial c_p}{\partial \nu_p}\frac{\partial c_p}{\partial \zeta} + \frac{2}{3\beta}x_h(g)\frac{\partial^2 c_p}{\partial \nu_p \partial \zeta} \tag{22}$$

As long as c_p in its arguments the above derivative is positive. When all of the components are put together

$$\left.\frac{d^2\Pi_p(m)}{d\nu_p d\zeta}\right|_{\nu_p=\nu_h} = \frac{4}{9\beta^2}\frac{\partial c_p}{\partial \nu_h}\frac{\partial c_p}{\partial \zeta}(1+\theta) - \frac{1}{\beta}\frac{\partial^2 c_p}{\partial \nu_h \partial \zeta}\left[\theta x_p(m) + \frac{2\theta}{3}x_h(g) + (1-\theta)x_p(g)\right] \tag{23}$$

Note that $\frac{\partial c_p}{\partial \nu_h} < 0$ and $\frac{\partial c_p}{\partial \zeta} > 0$ and $\frac{\partial^2 c_p}{\partial \nu_h \partial \zeta} < 0$. Since c_p is assumed to be submodular in its arguments, $\left|\frac{\partial^2 c_p}{\partial \nu_p \partial \zeta}\right| > \left|\frac{\partial c_p}{\partial \nu_p}\frac{\partial c_p}{\partial \zeta}\right|$, the derivative in Eq. (23) is positive. ■

Corollary 2 *The magnification is stronger for M&As than greenfield investments as long as the MNC has more bargaining power.*

Using the payoff functions in the Appendix, one can easily prove that $\frac{d^2\Pi_p(m)}{d\nu_h d\zeta} > \frac{d^2\Pi_p(g)}{d\nu_h d\zeta}$ if $\theta > \frac{1}{2}$ and $x_i(s)$ where $i = (p, h\}$.

In an M&A what firm p buys is the non-mobile skills (ν) of the indigenous firm as well. These skills affect both marginal production costs and the fixed entry costs as does the level of corruption (ζ) in country H. These skills become more valuable in corrupt environments. If the discrepancy between the non-mobile skills of the MNC and that of the indigenous firm is very high, then the acquisition price will be higher, too. Therefore, if the bargaining strength of the MNC is high, then the multinational can negotiate a price lower than the reservation price of the indigenous firm and at the same time can make higher profits by using the newly earned superior non-mobile skills in an M&A scenario.

In a way, it is the relative rates of corruption tolerance between these firms which determines the equilibrium mode of entry. For instance, if the corruption tolerance of firm p is very high due to superior mobile skills and if it couples with low levels of non-mobile skill endowments, then the indigenous firm can enjoy a greater advantage. It might as well be the case that the indigenous firm has all the "right" contacts with the local bureaucracy and it is going to be costly for the MNC to buy these non-mobile skills. If the indigenous firm happen to have a high bargaining strength then it might prevent the entry of the MNC all together and stay as the national monopoly.

In summary, as corruption increases multinational entry gets discouraged. However, passed a certain threshold, corruption may induce more M&As. Moreover, when the importance of skill endowments is considered, multinational firms with rich mobile and non-mobile skills would prefer greenfield investments in corrupt environments. MNCs with low levels of mobile and non-mobile skills would prefer no entry. Multinationals with high levels of mobile, but low levels of non-mobile skills would prefer M&A if the bargaining strength is high. Otherwise, they would go with the greenfield choice or no entry.

The results of this section lend themselves to empirical testing and I now turn to a discussion of the empirical analysis and the data set.

4 Econometric Analysis

The theoretical framework presented in the previous section suggests that corruption in a host country can have asymmetric effects on different ways of serving a foreign market. The following econometric analysis provides the impact of corruption on foreign entry modes by using a sample of Swedish multinational firms.

4.1 Econometric Model

Propositions 1 and 2 in the previous section state that corruption in the host country has asymmetric effects on a multinational's mode of foreign expansion. While greenfield investments decline with higher levels of corruption, cross-border M&As can be encouraged under certain circumstances. I use the following specification to test these predictions:

$$y_{ikt;s} = \beta_{0;s} + \beta_{1;s}\zeta_{kt} + \boldsymbol{\beta}'_{2;s}\mathbf{x}_{it} + \boldsymbol{\beta}'_{3;s}\mathbf{x}_{kt} + \varepsilon_{ikt;s} \tag{24}$$

where $y_{ikt;s}$ is a binary indicator if firm i's entry into country k during time period t in the form of $s \in \{m, g, n\}$, ζ_{kt} denotes corruption, $\mathbf{x}_{it}$ is a vector of firm-specific variables (including mobile skills μ_{it}; non-mobile skills ν_{it}; bargaining strength θ_{it}) and $\mathbf{x}_{kt}$ is a vector of country-specific variables (including variable production costs c_{kt}; market size α_{kt}). I also include time and industry fixed effects in all specifications to account for the effect of unobservables. Due to data limitations of industry-specific variables for different countries, the regressions have no such variables.

Propositions 3 through 6 and their corollaries involve more complex, secondary effects of corruption on mode of foreign entry decision which act through mobile and non-mobile skill endowments of the multinational firm. To test the predictions of the model I use marginal effects calculations, which will be discussed in detail below.

The most appropriate econometric method to use would be the nested logit model since the MNC first figures out the next best alternative to a negotiated agreement and then enters. However, due to lack of choice specific attributes in the data, the nested logit model becomes useless. Therefore, the empirical part of the paper adopts the most general setting where the firm decides if and how to enter.

In this paper, the bivariate probit model is the main method used. This model is useful in providing the marginal effects for each entry strategy. First, I estimate effects of corruption on FDI (M&A and greenfield together) versus no entry, because it would provide a useful comparison to some of the existing literature that does not take different entry modes into account. When the bivariate probit is used for the choice between FDI and no entry, there are two equations (one for FDI and one for no entry) and two binary dependent variables, $y_{ikt;fdi}$ (1 if there is FDI and 0 otherwise) and $y_{ikt;n}$ (1 if there is no entry and 0 otherwise). If the MNC chooses FDI, then $y_{ikt;fdi} = 1$ and $y_{ikt;n} = 0$. If the MNC chooses not to enter the host market, then $y_{ikt;fdi} = 0$ and $y_{ikt;n} = 1$.

Then, bivariate probit estimates of effects of corruption on new entry by Swedish multinationals through cross-border M&As and greenfield investments are estimated. Once more, there are two equations (one for M&As and one for greenfield investments) and two binary dependent variables, $y_{ikt;m}$ (1 if there is an M&A and 0 otherwise) and $y_{ikt;g}$ (1 if there is a greenfield investment and 0 otherwise). If the MNC chooses M&A, then $y_{ikt;m} = 1$ and $y_{ikt;g} = 0$. If the MNC chooses greenfield investment, then $y_{ikt;m} = 0$ and $y_{ikt;g} = 1$.

Error terms $\varepsilon_{ikt;s}$ are distributed as bivariate normal, each with a mean of zero, and variance-covariance matrix V, where V has values of 1 on the leading diagonal and correlations ρ as off-diagonal elements. The bivariate probit model shows structural similarities to a seemingly unrelated regression model, except that the dependent variables are binary indicators.

The LR test is used to test the independence of residuals to explore the existence of nesting possibilities if any.

4.2 The Dependent Variable

This section provides detailed information on the dependent variable. The data set used in the paper covers information on the cross-border activities of Swedish MNCs in 42 countries during three distinct time periods: 1987–1990, 1991–1994 and 1995–1998. The country coverage is determined by the availability of the corruption measure and control variables.

The firm-level data used in this paper is the product of a questionnaire sent to Swedish MNCs by the Research Institute of Industrial Economics (RIIE) in Stockholm, Sweden about every fourth year since 1960s. The data include all Swedish MNCs in manufacturing industry and contain detailed information on employment, production, R&D and entry modes of each majority owned foreign manufacturing affiliate. Only the period between 1987 and 1998 is used in this paper due to pronounced changes in the survey questions over time and the low response rate in the 2003 survey.

The degree of multinationality varies significantly in the data. More than half of the firms are single affiliate multinationals. An overwhelming majority of firms have foreign operations in just a few countries. When a new opportunity to serve a host country arises, this chance may come to a multinational active in another market.

The definitions of cross-border M&As and greenfield investments are taken from the RIIE survey. The RIIE asks the following four questions to each foreign affiliate: (1) From what year has the affiliate been a production company of the group? (2) Was the affiliate a sales company of the group before the year mentioned above? (3) Did the affiliate operate as a production company of another group before the year mentioned above? (4) Was the affiliate a state-owned company before the year mentioned above? If the answers to last three questions are all negative, then the investment is classified as a greenfield investment. If the answer to question 3 is affirmative, then the mode of entry is a cross-border M&A. The frequency of new affiliates transformed from sales companies of the group and the state-owned enterprise acquisitions is low.

Table 1 summarizes the foreign expansion transactions by Swedish MNCs between 1987 and 1998. The numbers of cross-border M&As and greenfield investments as well as the location of these investments in broad regional categories are reported. When examining this table, several remarks can be made. First, as can be observed in the bottom half of Table 1, in each time period foreign entry is small

Table 1 Entry characteristics of Swedish MNCs by regions

	1987–1990		1991–1994		1995–1998		All periods	
	m	*g*	*m*	*g*	*m*	*g*	*m*	*g*
Western Europe	107	21	63	16	42	7	212	44
Major non-European OECD	18	5	9	3	10	2	37	10
Eastern Europe and Russia	0	0	8	8	2	5	10	13
South and Central America	3	0	2	1	6	2	11	3
Asia/Africa	0	0	2	3	8	6	10	9
Cross-border M&As	1987–1990		1991–1994		1995–1998		All periods	
	128		84		68		280	
Greenfield investments	26		31		22		79	
No entry	4676		5387		3690		13,753	
Number of firms	115		131		90		330	
Number of countries	42		42		42		42	

when compared to no entry, which is true for an overwhelming majority of MNCs around the globe. However, among the two entry modes the total number of M&As is substantially higher than that of greenfield investment in all three time periods.

Second, observe the top half of Table 1. An overwhelming majority of investments are in Western Europe followed by major non-European OECD countries. Both M&As and greenfield investments in these two regions are higher than all the other regions together. The common denominator of all these countries is their level of development. FDI goes predominantly to advanced countries where corruption is relatively low, even though the share of developing countries has been rising. Apart from lower corruption levels, developed countries offer a large and growing demand coupled with ease of finding sub-contractors and distribution channels all of which favor entry.

Third and last, developed countries supply a higher number of high quality acquisition targets. Table 1 shows that Swedish MNCs have considerably higher M&As in Western Europe and major non-European OECD countries. The preferred mode of entry in developing countries is not as clear, however. The share of greenfield investments in all entry modes (calculated by using the last two columns of the top half of Table 1) in developing countries is 45 %, whereas it is only 18 % in developed countries.

4.3 Measuring Corruption

In this paper, I use corruption indices constructed from survey responses. Whether perceptions of corruption as enunciated by survey responses indeed reflect the reality is a commonly discussed issue. This paper is partial to the idea that

although perceptions may deviate from reality at the margin, there will not be wide divergences.

There is a plethora of corruption indices made available by different institutions through surveys conducted. In this paper, I use two different corruption indices: One is the Worldwide Governance Indicators (WGI), a long-standing World Bank research project to develop cross-country indicators of governance and the other one is the Corruption Perception Index (CPI) annually published by Transparency International (TI). Both are essentially polls of polls. I rescaled the values between 1 and 100 for comparison purposes, where higher values indicate higher levels of corruption in the host country.

The more widely known of the two is the CPI which collates results of up to 12 individual surveys conducted by the World Bank (World Business Environment Survey), the European Intelligence Unit, the World Economic Forum (Global Competitiveness Report), the Institute of Management Development, Political and Economic Risk Constancy in Hong Kong, etc. Many of the same sources used by the WGI are used by the CPI as well, and thus, not surprisingly, the WGI and the CPI are highly correlated.

The WGI consist of six composite indicators of broad dimensions of governance covering over 200 countries: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. The advantage of the WGI measure over others is its lesser susceptibility to poll—or question—specific idiosyncrasies due to its breadth of coverage and the variety of sources employed in compiling the index. The main sources for the WGI are polls conducted by various sources such as Standard and Poor's DRI (in conjunction with McGraw-Hill), the Economist Intelligence Unit, Political Risk Services (International Country Risk Guide), and the World Bank (in conjunction with the University of Basel). There are of course subtle differences between the questions asked by these sources. Country coverage is not exactly the same either. However, the survey respondents are divided between two groups: (1) business people and/or residents of a country, and (2) experts (who are asked to rank countries on various dimensions). A composite index for each dimension of governance is constructed using these individual surveys through an unobserved components model.

In this paper, I use the Control of Corruption, *CC* from the WGI as the main corruption indicator. It captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. The *CPI* is also used as a robustness check. In the following robustness exercises, I also consider another very relevant dimension of governance, namely the Rule of Law, *RL*. It captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.

Table 2 lists all countries included in the sample, the Control of Corruption Index in 1998, the Corruption Perception Index in 1998, the number of firms producing there in 1998, and the sum of all Swedish M&As and greenfield investments in

Table 2 The sample of countries, 1987–1998

Country	*CC* 0–100 1998	*CPI* 0–100 1998	No. of firms 1998	No. of *m*	No. of *g*
Germany	6.7	21	28	42	11
UK	5.4	13	26	28	4
USA	19.0	25	26	29	6
Denmark	2.8	1	25	29	3
Poland	36.7	54	21	4	10
France	21.9	33	20	16	6
Finland	2.6	4	18	16	7
Netherlands	4.6	10	16	12	0
Spain	22.5	39	15	9	1
Italy	39.6	54	15	20	4
Norway	3.6	10	14	16	3
Belgium	23.6	46	14	8	1
Brazil	50.0	60	12	6	2
Canada	5.2	8	8	4	3
Austria	8.7	25	8	8	3
China	55.0	65	8	3	5
India	55.7	71	7	3	1
Mexico	57.6	67	6	5	1
Australia	14.7	13	4	3	0
Hungary	36.9	50	4	4	1
Russia	68.7	76	4	2	2
Malaysia	38.9	47	4	0	1
Japan	31.4	42	4	1	1
Czech Republic	39.1	52	4	0	0
Greece	28.9	51	1–3	1	0
Portugal	23.3	35	1–3	5	1
Korea	43.3	58	1–3	2	0
South Africa	37.0	48	1–3	1	1
Philippines	52.9	67	1–3	0	0
Ireland	18.4	18	1–3	2	0
Argentina	53.7	70	1–3	0	0
Thailand	50.0	70	1–3	0	0
Turkey	61.7	66	1–3	0	1
Colombia	59.0	78	1–3	0	0
Taiwan	37.3	47	1–3	0	0
Indonesia	71.6	80	1–3	1	0
Slovenia	24.0	48	0	0	0
New Zealand	3.7	6	0	0	0
Chile	22.8	32	0	0	0
Venezuela	69.1	77	0	0	0
Iceland	9.7	7	0	0	0
Israel	24.1	29	0	0	0

the sample period. Table 2 does not reveal much about the relationship between corruption and form of FDI. The bottom of table shows many countries with very high corruption levels and low levels of Swedish entry. The top part shows low corruption levels coupled with high degrees of M&As and greenfield investments. However, this may simply reflect that Swedish multinationals mainly invest in developed European countries which also have lower corruption levels.

4.4 Firm Characteristics

The model presented in Sect. 3 is a highly stylized one written to provide a framework for the empirical analysis. The controls used in the regressions hereafter are inspired both from this simple model and the broader FDI literature.

Firm-specific skills. As Markusen (2002) points out, multinationals arise from the use of knowledge capital, a broad term that includes human capital of employees, patents, blueprints and procedures, which are called firm specific skills.

Multinationals can reduce their production costs through extensive use of these skills some of which can be provided to additional plants without reducing their value in existing plants. I use R&D intensity as a proxy for mobile-skills. *Mobile* is the MNC's total R&D expenditures divided by total sales at the end of each time period. High-tech firms are more dependent on their own technology creation and production technology, and as a result are more likely to enter by greenfield investments. Thus, I expect R&D to affect greenfield investments positively—pointed out by the theory in Sect. 3 as well.

Some skills, on the other hand, are location specific and cannot travel across borders. I proxy these non-mobile skills by previous experience in the host country. *Non-mobile* is the number of the previous affiliates of the MNC in the host country. *Non-mobile* carries information about the local knowledge of the firm that is specific to the host country, such as distribution networks, connections to local bureaucracy, and knowledge of local business culture. Note that *Non-mobile* may also represent competitive effects of the bargaining strength. If the MNC already has affiliates in the host country, it may not want to hurt itself by increasing the competition through a new venture and thus may incline more towards M&As which eliminate rivals. There is a well-established international business literature drawing attention to the differential impact of this variable on entry modes. Previous experience increases the local knowledge and connections of the MNC and thus may foster greenfield investments over cross-border M&As. On the other hand, it may also promote M&As because experienced MNCs are able to monitor their partners more effectively. Therefore, the expected sign is positive for both entry strategies yet the strength of this effect on each entry mode is ambiguous.

Bargaining strength. Market share of the firm is the most widely used bargaining power measure in the empirical industrial organization literature. There is a lack of data with broad industry and country coverage for the market share of a multinational in industry j in country k in time t. The next best alternative is using

the market concentration in industry *j* in country *k* in time *t*. OECD STAN database offers concentration measures for a limited number of countries and sectors from 1980 to 2000. I used these in my early regressions without much success due to many missing observations and small sample sizes.

Starting back with Anderson and Gatignon (1986), in the international business and management strategy literatures, international experience has been cited as an indicator of low levels of internal uncertainty and greater confidence in business dealings and thus stronger bargaining positions around the negotiation table. Therefore, in this paper, I assume that multinationals with more international experience are stronger bargainers. *Affworld* is the number of the previous affiliates of the MNC all around the world and represents a broad international experience that fosters FDI by MNCs (Caves, 2007). The expected sign for this variable for both entry modes is positive. However, I expect a stronger positive for cross-border M&As since international experience is anticipated to boost the bargaining strength and thus the probability of M&As. I also use firm size measured by total employment or sales of the firm as an indicator of the bargaining strength (results not reported in the paper due to brevity but available upon request), since larger firms with deep pockets are considered to be more experienced and stronger bargainers (See Caves, 2007).

4.5 Country Characteristics

Guidance from Blonigen and Piger (2011) is followed to motivate the use of country level controls that are not directly suggested by the model. I include almost all variables with high inclusion probabilities (above 50 %) for cross-border M&As reported in their worldwide sample.[8]

Market size (measured by *GDP*), infrastructure (measured by telephone mainlines per one million people,*Tel*), skill level of the labor force in the host country (measured by the share of university graduates in the population, *Skill*), trade openness of the host country (share of trade volume in GDP, *Open*) and distance (measured by using the great circle formula that calculates the minimum distance along the surface of the earth between Sweden and the host country, *Distance*) are widely used determinants of entry and are expected to favor both kinds of entry (Brainard 1997; Carr et al. 2001). In addition, host country tax rate (measured by corporate tax rate, *Tax*), double taxation treaties (measured by bilateral tax treaties in effect between Sweden and the host country, *DTT*), bilateral investment treaties

[8]Parent level variables such as parent GDP, parent GDP per capita and parent education level are not included since Sweden is the only parent country and there is no cross-sectional variation in these variables. In essence, they act as time fixed effects and drop out of regressions. Common official language, colonial relationships and contiguous border variables do not work in the regressions due to little variation, as well.

Table 3 Descriptive statistics

	Units	Mean	Standard deviation	Minimum	Maximum
CC	Number	36.5	21.4	1.72	82
CPI	Number	44.4	25.3	0	89
RL	Number	35.7	19.3	5.09	80
Mobile	Number	0.021	0.034	0	0.262
Non-mobile	Number	0.118	0.646	0	14
Affworld	Number	6.14	15.2	1	125
Time	Days	32.3	28.50	2	128
GDP	In trillions of USD	0.741	1.44	0.008	8.79
GDP/capita	In thousands of USD	16.3	11.1	0.426	39.0
Open	Number	0.611	0.383	0.110	2.93
Tel	Per one million people	0.378	0.189	0.022	0.684
Distance	In thousands of kms	4.66	4.42	0.4	17.0
Skill	Percentage	2.91	1.25	0.437	6.33
Tax	Percentage	32.4	6.71	9.8	50
DTT	Number	0.757	0.429	0	1
BIT	Number	0.230	0.421	1	1

(denoted as *BIT*) and regional trade agreement (*RTA*) dummies are included as suggested by Blonigen and Piger (2011).

GDP per capita is used to account for the availability of acquisition targets in the host country because it is a broad measure of general level of development. Even though it is easier to find sub-contractors and distribution channels in developed countries, which in fact favors entry, another important issue is that a developed country supplies a bigger number of more high quality acquisition targets. It is harder to find suitable acquisition targets in less developed countries. Therefore, acquisitions are expected to be more favorable in countries with high *GDP/capita*.

Direct costs of entry into the host country are not available in the RIIE data set. I use the official time it takes to start-up a new firm in the host country as presented in Djankov et al. (2002), *Time*, as proxy for fixed entry costs.

The country-level data are collected from the International Financial Statistics of IMF and the World Development Indicators Database of the World Bank. More information about variables is provided in Table 3.

5 Results

This section presents the results of econometric analysis in three subsections. Due to reasons explained in the previous section, rather than a nested qualitative choice model, I adopt the most general setting where the firm decides if and how to enter a host country market. Considering the structure of the theoretical model the next

best econometric model is a multivariate probit because it allows a flexible pattern of conditional covariance among the latent utilities of alternatives.

The greatest shortcoming of the multivariate probit model, however, is that the computation of marginal effects is overly complex and do not always yield meaningful estimates as stated in Greene and Hensher (2009). The next alternative is using the bivariate probit model. Cappellari and Jenkins (2003) present a comparison of bivariate probit (maximum likelihood estimation) to their multivariate probit (simulated maximum likelihood estimation) analysis and come to a conclusion that as long as the number of random draws and the sample size are large enough, the two methods yield very similar predictions. Since these two conditions are satisfied in the estimations in this paper, I use bivariate probit estimation to be able to discuss economic size of the estimates.

5.1 FDI Decision Alone

Other than a few exceptions, an overwhelming majority of the existing work on the effects of corruption on foreign direct investment makes no distinction between the modes of foreign entry. Therefore, in this subsection I begin with the bivariate probit estimates of the effects of corruption on both types of FDI by the Swedish multinational corporations to put the results in perspective with the existing literature.

The first two columns in Table 4 present the coefficient estimates while the last two columns report the marginal effects of explanatory variables on the success probability of each strategy. Marginal effects from the bivariate probit model are obtained using the method proposed by Greene (1996) where both direct and indirect partial effects (as common in models with multiple equations such as this one) are calculated and added at the means of the explanatory variables. Since the sample size is sufficiently large, marginal effects at means (MEM) and average marginal effects (AME) are similar. Therefore, only MEMs are reported.[9]

All regressions in Table 4 include a constant as well as time, industry and RTA fixed effects. Wald χ^2 is 1462 indicating a good fit. Correlation coefficient ρ is significant revealing that *fdi* and *n* are not independent from each other as foreign expansion strategies.

Corruption proxied by the Control of Corruption measure from the WGI database is significantly negative in equation *fdi* (column 1) and positive and significant in equation *n* (column 2), revealing that higher levels of corruption in a host country discourage FDI by Swedish multinationals. This is in line with Hakkala et al. (2008) and Javorcik and Wei (2009) as well as the previous literature where researchers

[9]No standard error or significance indicators are given for marginal effects because Greene (2010) argues that the process of statistical testing about partial effects produces mostly uninformative and sometimes contradictory and even misleading results.

Table 4 FDI versus no entry

	Bivariate probit			
	Estimates		MEM	
Entry mode	*fdi*	*n*	*fdi*	*n*
Corruption	−1.06**	1.89***	−0.092	0.412
	(0.542)	(0.364)		
Mobile	0.163*	−9.314***	0.015	−0.694
	(0.976)	(1.122)		
Non-mobile	0.093**	0.032	0.011	0.022
	(0.044)	(0.056)		
Affworld	0.023***	0.004	0.002	0.003
	(0.0007)	(0.035)		
Time	−0.745*	0.527	0.019	0.018
	(0.438)	(0.671)		
GDP	0.072***	−0.065***	0.006	−0.021
	(0.026)	(0.017)		
GDP/capita	0.004	0.042***	0.003	0.015
	(0.008)	(0.006)		
Open	−0.069	−0.154	−0.006	−0.011
	(0.156)	(0.102)		
Tel	0.862**	0.939	0.052	0.415
	(0.414)	(0.749)		
Distance	−0.125**	0.158**	−0.134	0.021
	(0.054)	(0.079)		
Skill	0.165***	−0.218***	0.018	−0.082
	(0.041)	(0.027)		
Tax	0.008	0.002	0.021	0.034
	(0.022)	(0.043)		
DTT	0.562***	−0.128***	0.091	−0.074
	(0.187)	(0.045)		
BIT	0.236**	−0.078**	0.084	−0.093
	(0.127)	(0.042)		
Observations	13,258			
Wald χ^2	1462			
ρ	−0.858			
LR test of indep. of eq.	552.5 (0.000)			

Standard errors are in parentheses; ***, **, * denote significance at the 1, 5, 10 % level, respectively; all regressions include a constant as well as time, industry and RTA fixed effects

generally have found a significant negative effect of corruption on multinational entry without differentiating between different entry modes using aggregate data.

Turning to economic size of the estimated parameters, calculations of marginal effects show that a small increase in *Corruption* reduce the probability of FDI by

9.2 %. Although this is not large in absolute magnitude, compared to the probability evaluated at the sample mean of 2.7 % (the success probability of FDI in the sample), this is nevertheless economically meaningful.

Mobile skills of the multinational increase the likelihood of *fdi* with a small marginal effect but reduces the odds for *n* with a marginal effect of −69.4 %. Experience in the host country (*Non-mobile*) measured as the number of previous affiliates in the host country has no effect on probability of no entry, however, it increases the likelihood of FDI. The marginal effect is rather small.

International experience (*Affworld*), infrastructure (*Tel*), market size (*GDP*), labor skill in the host country (*Skill*) and double taxation and bilateral investment treaties (*DTT* and *BIT*) increase the likelihood of FDI as expected while FDI declines in distance (*Distance*). Trade openness of the host country *(Open*), time that its takes to start a new business (*Time*) and corporate tax rate in the host country (*Tax*) are not significant.

5.2 M&As Versus Greenfield Investments

I now turn to the bivariate probit estimates of effects of corruption on new entry by Swedish multinationals. The first two columns in Table 5 present the coefficient estimates whereas the last two columns report the marginal effects of explanatory variables on the success probability of M&As and greenfield investments. All regressions include a constant as well as time, industry and RTA fixed effects. Wald χ^2 is 473 indicating a good fit. Correlation coefficient ρ is significant revealing that A and G are not independent from each other as strategies.

Corruption is positive and significant at 10 % in equation *m* (column 1) and negative and highly significant in equation *g* (column 2), revealing that higher levels of corruption in the host country may grease the squeaky wheels in case of M&As while discouraging greenfield investments conducted by Swedish multinationals.

When compared with the recent literature—apart from the fact that none of the recent studies concentrates on M&As at the firm level—Javorcik and Wei (2009) find that in highly corrupt environments, conditional on entry, joint ventures will be the chosen mode of entry rather than wholly owned subsidiaries. The initial result here is consistent with their finding.

Calculating the marginal effects shows that an infinitesimal increase in *Corruption* increases the probability of an M&A by 9.6 %. Again, although this seems small in absolute magnitude, compared to the probability evaluated at the sample mean of 2 % (the success probability of M&As in the sample), this is economically meaningful. The same marginal effect for a greenfield investment is −16.8 %. In other words, a small increase in corruption reduces the likelihood of a greenfield project by about 17 %.

Proposition 1 from the theoretical model presented in Sect. 3 predicts an inverse relationship between corruption and greenfield investments and the empirical

Table 5 MAs versus greenfield investments

	Bivariate probit			
	Estimates		MEM	
Entry mode	*m*	*g*	*m*	*g*
Corruption	1.48*	−2.14***	0.096	−0.168
	(0.813)	(0.728)		
Mobile	−1.66	5.96***	−0.013	0.108
	(1.58)	(1.53)		
Non-mobile	0.106**	−0.138	0.211	−0.003
	(0.043)	(0.095)		
Affworld	0.018***	0.029***	0.001	0.0004
	(0.004)	(0.011)		
Time	−0.344	−0.521	−0.028	0.012
	(0.367)	(0.507)		
GDP	0.064***	0.041*	0.003	0.0005
	(0.029)	(0.022)		
GDP/capita	0.019**	0.007	0.004	0.0002
	(0.009)	(0.016)		
Open	−0.165	0.421*	−0.005	0.007
	0.187	(0.230)		
Tel	1.29*	0.039	0.048	−0.0003
	(0.751)	(1.28)		
Distance	−0.087*	−0.192**	−0.068	−0.093
	(0.046)	(0.088)		
Skill	0.241***	0.192	0.028	0.009
	(0.073)	(0.164)		
Tax	0.008	0.002	0.018	0.012
	(0.035)	(0.064)		
DTT	0.104	0.092	0.005	0.004
	(0.164)	(0.243)		
BIT	0.562**	0.036	0.123	0.002
	(0.206)	(0.032)		
Observations	13,258			
Wald χ^2	473			
ρ	−0.592			
LR test of indep. of eq.	8.79 (0.01)			

Standard errors are in parentheses; ***, **, * denote significance at the 1, 5, 10 % level, respectively; all regressions include a constant as well as time, industry and RTA fixed effects

finding here confirms it. Proposition 2 points to a nonlinear relationship between corruption and M&As. The positive and weakly significant corruption effect on M&As reported in Table 5 is a new result and further investigated in the next subsection.

Swedish MNCs with high *Mobile* skills favor greenfield investments with a marginal effect of 10.8 %. On the other hand, *Non-mobile* skills always favors cross-border M&As with a large marginal effect of only 21.1 % and reduces to odds against greenfield investments, which suggests that Swedish MNCs endowed with stronger connections to local bureaucracy or knowledge of local business culture prefer cross-border M&As to greenfield FDI. This may also be interpreted as Swedish MNCs with more bargaining power derived from their previous experience in the host market acquire local firms rather than establishing wholly owned subsidiaries.

Turning to other coefficient estimates in the first two columns of Table 5, international experience (*Affworld*) and market size (*GDP*) increase the likelihood of both kinds of entry. The host country GDP per capita, the skill level and bilateral investment treaties, respectively proxied by *GDP/capita, Skill* and *BIT* increase the odds in favor of M&As only. All of these have relatively small marginal effects on the mode of entry except for *BIT*. The existence of a bilateral investment treaty between Sweden and the host country increases the probability of having an M&A by 12.3 %. Trade openness of the host country *(Open*), host country infrastructure (*Tel*), corporate tax rate (*Tax*), double taxation treaties (*DTT*) time that its takes to start a new business (*Time*) are mostly insignificant for both types of entry.

5.3 More Marginal Effects

In the previous section, the effect of corruption on greenfield investments is significant and negative as expected by the FDI literature. The same variable has quite a different effect on cross-border M&As; it is significant only at 10 % nonetheless positive. On the one hand, it is highly preferable to overtake a local firm with all its knowledge about the host country conditions particularly in countries with high levels of corruption. On the other hand, if the MNC is endowed with high levels of mobile skills such as technological sophistication then a local partner may open the door for leakage of these valuable mobile skills when corruption is high.

The theoretical model in this paper weighs heavily on such nonlinear interactions. Interaction terms are used extensively in applied econometrics to account for such nonlinearities. However, in nonlinear models such as the bivariate probit used here, the magnitude of the interaction effect does not equal the marginal effect of the interaction term (Ai and Norton, 2003). Therefore, to test Propositions 2–6 in the model, I estimate changes in marginal effects for ranges of values (10th, 25th, 50th, 75th and 90th percentiles) for *Corruption* and evaluate how the marginal effects differ across these ranges.

Table 6 shows the marginal effect of *Corruption* as the corruption level in the host country changes. The marginal effect estimation is executed using the Greene (1996) method. For brevity only the corruption terms are reported. The top part of the table shows the marginal effects at means (MEM) from Table 5 for easier

Table 6 The marginal effect of corruption

	Bivariate probit	
	Mergers and acquisitions	Greenfield investment
	$y_{ikt;m} = 1$ and $y_{ikt;g} = 0$	$y_{ikt;m} = 0$ and $y_{ikt;g} = 1$
MEM		
Corruption	0.096	−0.168
Representative values		
Corruption level		
10th percentile	−0.023	−0.043
25th percentile	−0.052	−0.078
50th percentile	0.084	−0.171
75th percentile	0.121	−0.457
90th percentile	−0.073	−0.896

comparison. The bottom part reports how the marginal effect changes across five different ranges reported in the table.

Adding strength to Proposition 1, as corruption in the host country increases the partial effect of corruption on the probability of greenfield investments decline. In other words, in more corrupt environments Swedish multinationals reduce their greenfield investments at increasing rates. In the least corrupt host countries (10th percentile) the marginal effect of corruption on greenfield probability is −4.3 % while it reaches almost −90 % in the most corrupt host countries (90th percentile).

The results for M&As are broadly consistent with the predictions of Proposition 2. The marginal effect is negative in the 10th and 25th percentile ranges while it turns to positive in the 50th and 75th percentile ranges. In other words, in less corrupt host countries as corruption increases the likelihood of M&As decline. However, in more corrupt environments, as corruption increases more M&As are likely. That is to say, in corrupt environments, a multinational with high non-mobile skills chooses to enter the host country through an M&A because that way it can employ its own knowledge of that market—if any and internalize the local knowledge of the acquired local firm as well. The result obtained for the 90th percentile is noteworthy because in the most corrupt countries increasing corruption levels reduce the M&A probability. In other words, as corruption levels pass a certain threshold the multinational's likelihood of both kinds of entry declines. This result is in line with Javorcik and Wei (2009).

Next, I investigate the change in the marginal effect of *Mobile* skills for varying corruption levels and report the results in Table 7. Propositions 3 and 4 state that multinationals that are endowed with higher levels of mobile skills are more likely to engage in foreign direct investment. However, this likelihood is dampened in more corrupt host countries, particularly for M&As.

The bottom part of Table 7 shows that the negative impact of corruption on greenfield investments gets stronger with higher levels of mobile skills. In other words, a multinational with sophisticated technology or novel marketing ideas is

Table 7 The marginal effect of mobile skills

	Bivariate probit	
	Mergers and acquisitions	Greenfield investment
	$y_{ikt;m} = 1$ and $y_{ikt;g} = 0$	$y_{ikt;m} = 0$ and $y_{ikt;g} = 1$
MEM		
Corruption	0.096	−0.168
Mobile	−0.013	0.108
Representative values		
Corruption level		
10th percentile	−0.021	0.216
25th percentile	−0.016	0.147
50th percentile	−0.017	0.113
75th percentile	−0.012	0.082
90th percentile	−0.027	0.031

Table 8 The marginal effect of non-mobile skills

	Bivariate probit	
	Mergers and acquisitions	Greenfield investment
	$y_{ikt;m} = 1$ and $y_{ikt;g} = 0$	$y_{ikt;m} = 0$ and $y_{ikt;g} = 1$
MEM		
Corruption	0.096	−0.168
Non-mobile	0.211	−0.003
Representative values		
Corruption level		
10th percentile	0.171	−0.007
25th percentile	0.184	−0.004
50th percentile	0.352	0.001
75th percentile	0.582	0.005
90th percentile	0.758	0.009

affected severely by corruption. While higher levels of mobile skills increases the likelihood of greenfield investment by 21.6 % in the least corrupt countries, this rate declines to 3.1 % in the most corrupt countries. This result is in line with Proposition 3. The results in Table 7 do not show strong support for Proposition 4. Mobile skills are not strongly relevant for M&As. Neither the direct nor the indirect effects are significant economically.

Finally, I investigate how the usefulness of *Non-mobile* skills change as level of corruption changes. Table 8 reports the marginal effect results for non-mobile skills at different percentiles of *Corruption*. Propositions 5 and 6 state that multinationals that are endowed with higher levels of non-mobile skills are more likely to engage in foreign direct investment. However, this likelihood is higher in more corrupt host countries, particularly for M&As.

The bottom part of Table 8 presents that *Non-mobile* skills are almost irrelevant for the probability of conducting greenfield investments, showing little support for Proposition 5. On the other hand, Swedish multinationals increase their M&As by 21.1 % on average as their non-mobile skills improve. In the least countries the impact of these skills on M&A probability is 17.1 % while it reaches 75.8 % in the most corrupt countries. In other words, if the multinational has rich knowledge of local conditions through its previous affiliates in the host country then its non-mobile skills get even more useful for this firm. This result is in line with Proposition 6.

5.4 Robustness

Table 9 reports the marginal effects of *Mobile* and *Non-mobile* at different *Corruption* levels using different measures of corruption. The first four columns report the results with widely used CPI of Transparency International. The next four columns present the results with the Rule of Law measure again from WGI. Results are very similar to the ones in Tables 7 and 8.

Next, I turn my attention to non-OECD countries as these countries are distinctly different from the bulk of countries that Swedish MNCs mainly invest.[10]

Table 9 Robustness, different corruption measures

	Bivariate probit							
	CPI				Rule of Law			
	m	*g*	*m*	*g*	*m*	*g*	*m*	*g*
MEM								
Corruption	0.078	−0.147	0.078	−0.147	0.098	−0.171	0.098	−0.171
Mobile	−0.012	0.095			−0.013	0.108		
Non-mobile			0.205	−0.002			0.211	−0.003
Representative values *Corruption level*	ME of mobile skills		ME of non-mobile skills		ME of mobile skills		ME of non-mobile skills	
10th percentile	−0.019	0.204	0.167	−0.006	−0.020	0.217	0.174	−0.008
25th percentile	−0.015	0.138	0.174	−0.004	−0.017	0.149	0.191	−0.004
50th percentile	−0.017	0.112	0.327	0.002	−0.019	0.117	0.363	0.002
75th percentile	−0.011	0.076	0.538	0.006	−0.015	0.089	0.588	0.005
90th percentile	−0.025	0.029	0.701	0.008	−0.032	0.036	0.763	0.013
Observation	12,719				13,258			

[10]Most of the countries that Swedish MNCs invest are developed countries which also have lower corruption levels than average country. Swedish MNCs invest in nearby developed countries because they have lots of potential M&A targets, and these countries just happen to have low levels of corruption cross-sectionally. Even though there are country-level regressors to control for level of development of a country in previous estimations, a more compelling experiment is to restrict

Table 10 Robustness, different subsamples

	Bivariate probit							
	Non-OECD				Large firms			
	m	*g*	*m*	*g*	*m*	*g*	*m*	*g*
MEM								
Corruption	−0.025	−0.382	0.045	−0.382	0.045	−0.062	0.045	−0.062
Mobile	−0.038	0.051			−0.001	0.284		
Non-mobile			0.251	0.032			0.391	−0.005
Representative values *Corruption level*	ME of mobile skills		ME of non-mobile skills		ME of mobile skills		ME of non-mobile skills	
10th percentile	−0.051	0.216	0.171	0.009	−0.001	0.473	0.249	−0.007
25th percentile	−0.045	0.147	0.184	0.012	−0.002	0.365	0.328	−0.002
50th percentile	−0.039	0.113	0.352	0.029	−0.001	0.276	0.371	0.002
75th percentile	−0.019	0.082	0.582	0.045	−0.005	0.079	0.565	0.009
90th percentile	−0.032	0.031	0.758	0.062	−0.006	0.021	0.674	0.012
Observation	5434				1689			

In Table 10, direct effects of corruption are observed for all entry strategies while secondary effects realized through mobile and non-mobile skills remain important. One very notable change in this set of marginal effects is the flip in the sign of *Corruption* in the *m* equation (Column 1) which suggests that higher levels of corruption reduce the likelihood of M&As in the non-OECD countries. In other words, as corruption levels get very high, the tolerance of the MNCs even with high levels of non-mobile skills diminishes.

The very final exercise is restricting the sample to large firms only, although this limits a lot of the time series variation in the sample. For the purposes of this exercise, a large firm is defined as a firm with 10 or more affiliates around the globe.[11] The results are reported in the last four columns of Table 10. The direct effect of corruption is smaller in size. Secondary effects are still strong. Multi-affiliate MNCs have better and wider distribution networks around the globe and most importantly more international experience. Therefore, the M&As and greenfield investments conducted by these firms might be less prone to changes in corruption. In short, endowment of mobile and non-mobile skills as well as the degree of multinationality matter for how profound the effect of corruption will be on the mode of entry.

the sample to these developed countries only to avoid potentially spurious results. The results are very similar to overall regressions.

[11] Other thresholds (2 or more and 5 or more) are used in the estimations and the results are similar.

6 Conclusion

In the last decade, corruption has become an eminent item on the agenda of the international institutions. The UN Convention against Corruption, adopted in Mexico in December 2003 is the first global instrument embracing a comprehensive range of anti-corruption measures to be taken at the national level. According to the OECD Convention of Combating Bribery of Foreign Public Officials in International Business Transactions, which was signed in 1997, and went into effect in 1999, bribery of foreign officials by firms from member countries is a crime. The 15th International Anti-Corruption Conference (IACC) was completed in November 2012 with record participation from 140 countries and the main message of the conference was ending impunity.

In this paper, I attempt to disentangle the effects of corruption on entry mode decision by carrying out an empirical analysis with rich, firm-level data on the activities of Swedish MNCs around the globe in manufacturing sectors from 1987 to 1998. A number of propositions emerge from the theoretical framework. As corruption increases multinational entry gets discouraged. Corruption reduces the likelihood of greenfield investments while increasing the odds in favor of cross-border M&As. MNCs with higher endowments of mobile skills are more likely to invest in the host country. This likelihood is dampened the higher the degree of corruption in the host country. On the other hand, MNCs with higher endowments of non-mobile skills are more likely to invest in the host country. This likelihood is magnified the higher the degree of corruption in the host country.

The panorama of the results presented in the previous section shows the following: Firstly, corruption has a direct negative impact on greenfield investments and a weak positive impact on M&As. Secondly, there are complex, asymmetric, secondary effects of corruption on the mode of entry. Thirdly, international experience dampens the effect of corruption on the mode of entry. Lastly, the results are robust to differences in measures of corruption.

Acknowledgements Financial support from Oregon State University Research Office and Valley Library is gratefully acknowledged.

Appendix

Aggregate profit to firms p and h from sales in the host country for entry mode s and quantity choice $x_i(s)$ can be expressed respectively as follows:

$$\Pi_p(s, x_p(s)) = [(\alpha - \beta X(s) - c_p(s))x_p(s)] - F(s) - F^b(s) \tag{25}$$

$$\Pi_h(s, x_h(s)) = [(\alpha - \beta X(s) - c_h(s))x_h(s)] \tag{26}$$

where $X(s) = x_p(s) + x_h(s)$. When $s = m$, $x_h(s) = 0$ and when $s = n$, $x_p(s) = 0$. Maximizing (25) and (26) with respect to $x_p(s)$ and $x_h(s)$ in that order and solving for $x_p(s)$ and $x_h(s)$ in the first order conditions gives the equilibrium profit levels for each firm as

$$\Pi_p(s, x_p(s)) = \beta[x_p(s)]^2 - F(s) - F^b(s) \tag{27}$$

$$\Pi_h(s, x_h(s)) = \beta[x_h(s)]^2 \tag{28}$$

where

$$x_p(s) = \frac{\alpha - 2c_p(s) + c_h(s)}{3\beta} \tag{29}$$

$$x_h(s) = \frac{\alpha - 2c_h(s) + c_p(s)}{3\beta} \quad \text{if 2 firms are active} \tag{30}$$

or

$$x_i(s) = \frac{\alpha - 2c_i(s)}{2\beta} \quad \text{if only 1 firm is active} \tag{31}$$

where $i = (p, h\}$.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123–129.

Aidt, T. (2003). Economic analysis of corruption. *Economic Journal, 113*(491), 632–652.

Anderson, E., & Gatignon, H. (1986). Modes of foreign entry: A transaction cost analysis and propositions. *Journal of International Business Studies, 17*(3), 1–26.

Barassi, M. R., & Zhou, Y. (2012). The impact of corruption on FDI: A parametric and non parametric analysis. *European Journal of Political Economy, 28*(3), 302–312.

Bjorvatn, K., & Soreide, T. (2005). Corruption and privatization. *European Journal of Political Economy, 21*(4), 903–914.

Blonigen, B. A., & Piger, J. (2011). *Determinants of Foreign Direct Investment*. National Bureau of Economic Research: 16704.

Brainard, L. (1997). An empirical assessment of the proximity-concentration trade-off between multi-national sales and trade. *American Economic Review, 87*, 520–545.

Busse, M., & Hefeker, C. (2007). Political risk institutions and foreign direct investment. *European Journal of Political Economy, 23*(2), 397–415.

Cappellari, L., & Jenkins, S. (2003). Multivariate probit regression using simulated maximum likelihood. *Stata Journal, 3*(3), 278–294.

Carr, D., Markusen, J., & Maskus, K. (2001). Estimating the knowledge-capital model of the multinational enterprise. *American Economic Review, 91*, 693–708.

Caves, R. (2007). *Multinational enterprise and economic analysis*. Cambridge: Cambridge University Press.

Djankov, S., La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. (2002). The regulation of entry. *The Quarterly Journal of Economics, 117*(1), 1–37.

Egger, P. H., & Winner, H. (2005). How corruption influences foreign direct investment: A panel data study. *Economic Development and Cultural Change, 54*(2), 459–486.

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters, 107*(2), 291–296.

Greene, W. H. (1996). *Marginal Effects in the Bivariate Probit Model*. Leonard N. Stern School of Business, Department of Economics, New York University.

Greene, W. H., & Hensher, D. A. (2009). Modeling ordered choices. Unpublished manuscript, 1–181.

Habib, M., & Zurawicki, L. (2002). Corruption and foreign direct investment. *Journal of International Business Studies, 33*, 291–308.

Hakkala, K., Norbäck, P.-J., & Svaleryd, H. (2008). Asymmetric effects of corruption on FDI: Evidence from Swedish multinational firms. *The Review of Economics and Statistics, 90*(4), 627–642.

Hellman, J. S., Jones, G., & Kaufmann, D. (2002). *Far From Home: Do Foreign Investors Import Higher Standards of Governance in Transition Economies?* Mimeo.

Hines, J. (1995). *Forbidden Payment: Foreign Bribery and American Business After 1977*. National Bureau of Economic Research Working Paper 5266.

Jain, A. K. (2001). Corruption: A review. *Journal of Economic Surveys, 15*(1), 71–121.

Javorcik, B., & Wei, S.-J. (2009). Corruption and cross-border investment in emerging markets: Firm-level evidence. *Journal of International Money and Finance, 28*(4), 605–624.

Lui, F. (1985). An equilibrium queuing model of bribery. *Journal of Political Economy, 93*(4), 760–781.

Markusen, J. (2002). *Multinational firms and the theory of international trade*. Cambridge, MA: MIT Press.

Nocke, V., & Yeaple, S. (2004). *An Assignment Theory of Foreign Direct Investment*. National Bureau of Economic Research Working Paper: 11003.

Saha, B. (2001). *Foreign Investment, Financial Markets and Industry: India in the 1990s*. Mimeo.

Tekin-Koru, A. (2006). Corruption and the ownership composition of the multinational firm at the time of entry: Evidence from Turkey. *Journal of Economics and Finance, 30*(2), 251–269.

Tekin-Koru, A. (2012). Asymmetric effects of trade costs on entry modes: Firm level evidence. *European Economic Review, 56*(2), 277–294.

Thursby, J. G., & Thursby, M. C. (2006). Here or there? A survey of factors in multinational R&D location: Report to the Government/University/Industry Research Roundtable. Kauffman Foundation Large Research Projects Research.

Wei, S. (2000). How taxing is corruption on international investors? *Review of Economics and Statistics, 82*(1), 1–11.

Wheeler, D., & Mody, A. (1992). International investment location decisions: The case of U.S. firms. *Journal of International Economics, 33*, 57–76.

Wu, S. (2006). Corruption and cross-border investment by multinational firms. *Journal of Comparative Economics, 34*, 839–856.

Export Market Exit and Firm Survival: Theory and First Evidence

Sanne Hiller, Philipp J.H. Schröder, and Allan Sørensen

Abstract This paper deploys a dynamic extension of the Melitz (Econometrica 71(6):1695–1725, 2003) model to generate predictions on export market exit and firm survival in a setting where firms endogenously make exit decisions. In contrast, previous work typically assumes simple random death. The central driver of the model dynamics—creating the endogenous exit decision of firms—is the inclusion of exogenous economy wide technological progress. Thus our framework contains vintage capital properties and accordingly creative destruction of old firms by new firms. The model predicts—inter alia—that a higher relative productivity increases not only the likelihood of exporting, but also the chances of firm survival and continued export market engagements. We relate the model predictions to the empirical stylized facts of export market exit and firm survival based on Danish firm-level data. Overall, our empirical results support the central predictions from the model. First, productivity is a good predictor for survival. Secondly, firms experience a decline in market share prior to death and export market exit. Thirdly, firms that die or quit their exporting activity are small. Fourthly, small firms constitute the largest share of exits in a given cohort. Fifthly, the productivity distribution of survivors first-order dominates the one of exiting firms on both domestic and export markets. Finally, we confirm that a large share of firms which stop exporting continue to exist as pure domestic firms.

S. Hiller
Ruhr University, Bochum, Germany
e-mail: sanne.hiller@rub.de

P.J.H. Schröder (✉) • A. Sørensen
Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark
e-mail: psc@econ.au.dk; allans@econ.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_8

1 Introduction

Industry dynamics, and the associated reallocation patterns, have moved center stage in the theoretical and empirical international economics literature. Yet, while the main thrust of the literature has been on entry and market serving modes, much less attention has been paid to the patterns of firm death, and more importantly export market exit. In fact, the workhorse model of heterogenous firms trade (Melitz, 2003) abstracts from exit decision altogether. Instead it introduces random death, such that exit probabilities are unrelated to firms' characteristics. While this is a reasonable assumption for many applications, it gives little guidance for designing empirical investigations into export market exit patterns and firms survival—which at least since the introduction of Stigler's (1958) survivor principle are central performance characteristics of firms within the field of industrial organization. Moreover, data on firm survival and export market exit is widely available in micro data sets (see e.g. Eaton et al. 2008 or Jäkel 2014).

The present chapter attempts to close this gap. We use the highly trackable dynamic extension of the Melitz (2003) model provided in Schröder and Sørensen (2012) to replicate several theoretical insights on firm survival and export market exit dynamics that apply in heterogeneous firms trade settings. From the model we find, inter alia, that higher productivity firms survive longer, and that low productivity is associated with high probabilities of export market exit. Furthermore, among the exiting and dying firms young firms account for the lions share, similarly the model predicts that export market exit and/or firm death are preceded by shrinking sales (i.e. shrinking market shares).

Subsequently, we contrast these theoretical findings to the empirical regularities. Our empirical work corroborates the general predictions from theory. First, we find that the market share of firms declines prior to firm death and export market exit. Secondly, we observe that exiting firms or exiting exporters are small. Thirdly, our results suggest that small firms constitute the largest share of exits in a given cohort. Fourth, our empirical findings suggest the productivity distribution of survivors first-order dominates the one of exits on both domestic and export markets. Somewhat in contrast to theoretical predictions we find that empirically fewer than half of all firms that stop their export market activities continue their domestic activity. Finally, the theoretical prediction on the initial size distributions between stayers and exits is not mirrored in the data, which highlights that the clear link between size and productivity that is presumed in the theoretical model is, at least in our data, not present.

The key model mechanism at the center of the theoretical model by Schröder and Sørensen (2012) that we apply in this chapter is the inclusion of exogenous technological progress into the standard Melitz (2003) model augmented with Pareto distributed productivities. This extension generates an analytically solvable dynamic framework that preserves all the established novelties of Melitz (2003) while adding endogenous export market exit and closure decisions of firms. Instead of random death, the competition forces are the transmitting channel between technological progress and firm export market exit and survival. In the presence of technological

progress and vintage capital properties, the arrival of new (younger) competing producers, that draw their productivity from an improved distribution, squeezes the market share of existing producers such that incumbent firms eventually will exit the various markets.

There is a number of other theoretical models that explicitly address issues of dynamics, including firm death and export market exit, in heterogeneous firms trade settings. Rich exit dynamics are generated by the inclusion of some firm-specific random process affecting productivity, e.g. Impullitti et al. (2013) and Arkolakis (2016). In Felbermayr and Spiegel (2014) firms differ with respect to the expected economic lifetime of their innovations, a mechanism also at work in our current set-up, yet their focus is on the resulting heterogeneity in the cost of finance for firms. Alternatively, firms may uncover firm-specific innovation advances that secure firm growth or the absence of which causes firm exit once a good innovation draw is lacking, see for example Atkeson and Burstein (2010) and Costantini and Melitz (2008) who deal with the joint innovation and export decision of firms. Yet, while these works focus on several dimensions of industry dynamics, they all derive dynamics from an intra-firm process, such as a random productivity development, that is typically paired with a constant exogenous firm death probability, as evoked in Melitz (2003). In contrast, the model of Schröder and Sørensen (2012) that we apply in the present chapter highlights a different—but no doubt central—channel of firm exit dynamics initially suggested by Dasgupta and Stiglitz (1981): technological progress, paired with vintage capital, competition and creative destruction. Survival and export market exit is thus exogenous to the firm, deterministic but specific to the economy. Most importantly, the model generates rich predictions that can be directly contrasted to the empirical regularities.

Our empirical approach follows the standard methodologies used in the empirical international trade literature. Our data is based on detailed firm-level data provided by Statistics Denmark. The data covers the time period from 1996 to 2007, allowing us to analyze four different exit cohorts. We focus on a manufacturing panel, which is based on 3080 Danish firms.[1] Given the rich Danish data on firms, we are able to estimate TFP using the method suggested by Levinsohn and Petrin (2003) and to conduct non-parametric Kolmogorov-Smirnov tests comparing exit and survivor cohorts. Previous empirical work has established a clear link between productivity and firms survival. For example, empirically, high productivity firms survive longer, most firm closures are young firms, larger size entrants have lower exit probabilities, more productive exporters are more likely to continue to export compared to less productive exporters, and firms are regularly observed to withdraw from some markets while staying active on others (see for example Jovanovic 1982; Dunne et al. 1988; Caves 1998; Klepper 2002; Farinas and Ruano 2005; Wagner 2009, 2013; and Eaton et al. 2008). Our empirical findings are broadly in line with these previous works.

[1] See Jäkel (2014) for an instructive industry-level analysis of firm exit from the home market due to international competition from imports as well as export-pull, based on Danish firm-level data.

The next two sections introduce several testable theoretical predictions of Schröder and Sørensen (2012) and provide sketches of the central modeling items and proofs. Section 4 introduces our data and empirical research design. Section 5 contains our empirical findings. Section 6 discusses extensions and Sect. 7 concludes.

2 The Theoretical Predictions

We draw upon the Schröder and Sørensen (2012) extension of Melitz (2003), which includes dynamic firm-level behavior of exit in a situation of continuous technological progress. Progress is modeled as exogenous and continuous improvements in the productivity distribution that is available to firms prior to entering the market. With this simple assumption future firms will have expected higher productivity and larger market shares on all markets that they enter. Accordingly, potential new entrants realize prior to starting production of their drawn variety that they eventually will be ousted first from the foreign market (if they should choose to export) and later the domestic market. They take this expected time of exit into account when deciding whether to enter the industry or not.

As is shown in Schröder and Sørensen (2012), the resulting dynamic model, where exit and death of firms is model endogenously determined and not depending on an exogenous death probability, has a number of testable properties concerning its cross-sectional and dynamic predictions for export market exit and firm closure. Most central for the purpose of the present chapter and our subsequent empirical investigation are:

Prediction 1 *Surviving firms are more productive than exiting firms (this applies both for export market exit and for firm death).*

Prediction 2 *Entry productivity and size of firms on a given market are positively linked to the duration of serving this market.*

Prediction 3 *For a given exit wave, younger (older) age cohorts account for a larger (smaller) share of the exiting firms.*

Prediction 4 *Only small firms leave a market.*

Prediction 5 *Firms observe a declining market share before market exit (this applies both for export market exit and for firm death).*

Prediction 6 *Exporters that cease to export continue to exist as pure domestic firms for a while.*

The next section introduces the central modeling elements and sketches of the proofs.

3 The Central Modeling Elements

In this section, we present a summary of the continuous-time-dynamic extension of the symmetric n-country Melitz (2003) model provided by Schröder and Sørensen (2012). Throughout the exposition the well-established Melitz (2003) notation and conventions such as the assumption of Pareto-distributed productivity draws are applied. We deal throughout the theoretical exposition with marginal productivity. For additional details the reader should refer to Schröder and Sørensen (2012).

3.1 Technological Progress

Technological progress is introduced as exogenous and continuous, thus following Dasgupta and Stiglitz's (1981) work on the dynamics of oligopolistic industries and innovation. In particular, the model introduces exogenous and continuous improvements in the distribution of productivities available to entering firms. At entry each firm pays sunk innovation costs of f_e labor units and draws a firm-specific marginal productivity φ that it maintains throughout its endogenous life cycle. Due to selection at the production cut-off, the average productivity of the incumbents of today will be higher than the average productivity draw of tomorrow, i.e. at any point in time some of the drawn blueprints are not worth bringing to the market. The central implication of continuously improving productivity draws among new entrants paired with constant firm-specific productivities is that incumbent firms experience declining relative productivity and thus falling market shares over time. Eventually individual market shares decline to levels, such that firms cannot cover fixed costs and endogenously shut down.

Turning to the specifics, a firm born at time t of type ω has productivity $\varphi_t(\omega)$ which is a realization from the Pareto

$$G_t\left(\varphi_t(\omega)\right) = 1 - \left(\frac{\varphi_t(\omega)}{\bar{\varphi}_t}\right)^{-k} \text{ for } \varphi_t(\omega) \geq \bar{\varphi}_t, \tag{1}$$

where $\bar{\varphi}_t$ determines the location and k determines the shape of the distribution. The location of the distribution improves exogenously and continuously at rate $\beta > 0$, i.e. $\bar{\varphi}_t = \bar{\varphi}_0 e^{\beta t}$.

We can decompose each productivity draw into a deterministic increasing state of the technology component $\bar{\varphi}_t = \bar{\varphi}_0 e^{\beta t}$ and a stochastic *lottery* component $\Psi(\omega, t) = \Psi(\omega) = \frac{\varphi_t(\omega)}{\bar{\varphi}_t}$ which is time-invariant and Pareto-distributed with shape parameter

k and location parameter 1.[2] It follows that we can write productivity draws at any time t as

$$\varphi_t(\omega) = \varphi_0(\omega)\, e^{\beta t}, \tag{2}$$

where $\varphi_0(\omega)$ is a draw from the Pareto $G\left(\varphi_0(\omega)\right) = 1 - \left(\frac{\varphi_0(\omega)}{\bar{\varphi}_0}\right)^{-k}$ for $\varphi_0(\omega) \geq \bar{\varphi}_0$. To see this, note that

$$\begin{aligned}
\Pr\left(\varphi_0(\omega)\, e^{\beta t} < x\right) &= \Pr\left(\varphi_0(\omega) < x e^{-\beta t}\right) \\
&= 1 - \left(\frac{x e^{-\beta t}}{\bar{\varphi}_0}\right)^{-k} = 1 - \left(\frac{x}{\bar{\varphi}_0 e^{\beta t}}\right)^{-k} \\
&= 1 - \left(\frac{x}{\bar{\varphi}_t}\right)^{-k} = \Pr\left(\varphi_t(\omega) < x\right).
\end{aligned}$$

3.2 Households

The representative household supplies exogenously L units of labor and chooses a consumption path $\{C_s\}_{s=t}^{\infty}$ to maximize utility[3] $U = \int_t^{\infty} \ln(C_s)\, ds$ subject to a budget constraint. The optimal expenditure path has $E_t = E = L$ for all t which equals labor income (because wages are normalized to unity). Expenditures in any period are spread over the set of available varieties, Ω_t, to maximize $C_t = \left[\int_{\omega\in\Omega_t} [c_t(\omega)]^{\frac{\sigma-1}{\sigma}}\, d\omega\right]^{\frac{\sigma}{\sigma-1}}$ implying a demand of

$$c_t(\omega) = \frac{E}{P_t}\left(\frac{p_t(\omega)}{P_t}\right)^{-\sigma} \text{ for all } \omega \in \Omega_t, \tag{3}$$

where $p_t(\omega)$ is the price of variety ω and $P_t = \left[\int_{\omega\in\Omega_t} [p_t(\omega)]^{1-\sigma}\, d\omega\right]^{\frac{1}{1-\sigma}}$ is the price index.

[2] To see this, consider the probability of drawing a *lottery* component below $\psi \geq 1$:

$$\begin{aligned}
&\Pr\left(\Psi(\omega,t) < \psi\right) = \Pr\left(\frac{\varphi(\omega,t)}{\bar{\varphi}_t} < \psi\right) \\
&= \Pr\left(\varphi(\omega,t) < \psi\bar{\varphi}_t\right) = G_t\left(\psi\bar{\varphi}_t\right) = 1 - \psi^{-k}
\end{aligned}$$

Hence Ψ is Pareto with shape parameter k and location parameter 1.

[3] For simplicity, we follow Melitz (2003) by imposing the assumption of no time discounting.

3.3 Firms

After the payment of the sunk innovation costs f_e and the realization of the marginal cost drawn from the Pareto distribution (1), a firm's labor requirement conditional on production is $l_t(\omega) = f + \frac{q_t(\omega)}{\varphi(\omega)}$, where f is fixed production costs and $q_t(\omega)$ is output. A firm has to pay fixed costs of f_x for each export market it serves, and moreover exports are subject to iceberg trade costs, $\tau \geq 1$. It is straightforward to include further types of trade barriers resulting in richer policy implications (see Schröder and Sørensen, 2014), however, such extensions are ignored here. Upon entry and subsequently at each point in time, firms decide conditional on productivity and industry structure which markets to serve.

Given the constant elasticity of demand, prices are set as a constant mark-up, $\frac{\sigma}{\sigma-1}$, on marginal costs. Flow profits at time t on the domestic market and on export markets for a firm with *lottery* component φ_0 and age m are given by

$$\pi_{t,m}(\varphi_0) = B_t e^{\beta(\sigma-1)(t-m)} \varphi_0^{\sigma-1} - f \tag{4}$$

$$\pi^x_{t,m}(\varphi_0) = B_t \tau^{1-\sigma} e^{\beta(\sigma-1)(t-m)} \varphi_0^{\sigma-1} - f_x, \tag{5}$$

where $B_t = \frac{1}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} EP_t^{\sigma-1}$ is the market-specific demand component at time t. As shown in Schröder and Sørensen (2012) a balanced growth path with a stable industry structure exists. A stable industry structure requires stable exit and export *lottery* thresholds (relative productivity), which in turn requires the above flow profits to be time-invariant. This can only be achieved if the market-specific demand component B decreases to balance the technological improvement, i.e. B must decrease at rate $\beta(\sigma-1)$.[4] Writing $B_t = B_0 e^{-\beta(\sigma-1)t}$, the flow profit expressions read

$$\pi_m(\varphi_0) = B_0 \varphi_0^{\sigma-1} e^{-\beta(\sigma-1)m} - f \tag{6}$$

$$\pi^x_m(\varphi_0) = B_0 \tau^{1-\sigma} \varphi_0^{\sigma-1} e^{-\beta(\sigma-1)m} - f_x\,. \tag{7}$$

A new firm serves the domestic market provided $\pi_0(\varphi_0) \geq 0 \Leftrightarrow \varphi_0 \geq \varphi_0^{exit} = \left(\frac{f}{B_0}\right)^{\frac{1}{\sigma-1}}$ and export markets provided $\pi^x_0(\varphi_0) \geq 0 \Leftrightarrow \varphi_0 \geq \varphi^x_0 = \left(\frac{f_x}{B_0}\right)^{\frac{1}{\sigma-1}} \tau$.[5] Due to the exogenously technological progress, a firm observes that its productivity falls

[4] The stable industry structure implies that the distribution of marginal productivities increases at rate β which in turn implies that prices decrease at rate β. The time-invariant nominal expenditures and fixed/sunk costs imply, given the constant mark-ups, a time-invariant number of varieties, M_t. Thus it follows that $B_t = \frac{1}{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} EP_t^{\sigma-1}$ decreases at rate $\beta(\sigma-1)$. See Schröder and Sørensen (2012) for further details.

[5] We impose the conventional parameter restriction $f_x \tau^{\sigma-1} > f$ that ensures that firms are partitioned into exporters and non-exporters. Notably, Jørgensen et al. (2012) study the situation beyond partitioning, where all firms export and low productivity firms compensate losses on the home market via exporting activity, such situations are not dealt with in the present set-up.

over time relative to younger competitors. Eventually, the market share will fall to a level where the firm is unable to cover fixed costs in a given market, and the firm therefore endogenously exits the market at that point in time. The ages at which a firm shuts down (m^{exit}) and leaves a given export markets (m^x) are determined by (6) and (7), respectively. They are

$$m^{exit}\left(\varphi_0\right) = \frac{1}{\beta\left(\sigma-1\right)} \ln\left(\frac{B_0\varphi_0^{\sigma-1}}{f}\right) \tag{8}$$

$$m^{x}\left(\varphi_0\right) = \frac{1}{\beta\left(\sigma-1\right)} \ln\left(\frac{B_0\varphi_0^{\sigma-1}\tau^{1-\sigma}}{f_x}\right). \tag{9}$$

3.4 Industry Equilibrium

The industry equilibrium is pinned down by the free entry condition. Taking the endogenous market durations (8) and (9) into account, net present value of flow profits read[6]

$$\begin{aligned}
\pi\left(\varphi_0\right) &= \int_0^{m^{exit}\left(\varphi_0\right)} \left(B_0\varphi_0^{\sigma-1}e^{-\beta(\sigma-1)m} - f\right) dm \\
&= \frac{f}{\beta\left(\sigma-1\right)}\left[\frac{B_0\varphi_0^{\sigma-1}}{f} - 1 - \ln\left(\frac{B_0\varphi_0^{\sigma-1}}{f}\right)\right] \\
\pi^x\left(\varphi_0\right) &= \frac{f_x}{\beta\left(\sigma-1\right)}\left[\frac{B_0\tau^{1-\sigma}\varphi_0^{\sigma-1}}{f_x} - 1 - \ln\left(\frac{B_0\tau^{1-\sigma}\varphi_0^{\sigma-1}}{f_x}\right)\right]
\end{aligned}$$

The free entry condition balances the expected net present value of flow profits with the sunk cost of entry costs, i.e.

$$\int_{\varphi_0^{exit}}^{\infty} \pi\left(\varphi_0\right) dG\left(\varphi_0\right) + n\int_{\varphi_0^{x}}^{\infty} \pi^x\left(\varphi_0\right) dG\left(\varphi_0\right) = f_e,$$

[6] On the balanced growth path, the interest rate equals the discount rate which by assumption equals zero.

which pins down the productivity thresholds as

$$\varphi_0^{exit} = \left(\frac{(\sigma-1)}{k-(\sigma-1)}\frac{1}{\beta k}\frac{f}{f_e}\right)^{\frac{1}{k}}\left(1+n\left(\frac{f_x}{f}\right)^{\frac{\sigma-1-k}{\sigma-1}}\tau^{-k}\right)^{\frac{1}{k}}\bar{\varphi}_0 \tag{10}$$

$$\varphi_0^{x} = \left(\frac{f_x}{f}\right)^{\frac{1}{\sigma-1}}\tau\varphi_0^{exit}.$$

A comparison with the Melitz (2003) model conditional on productivities being Pareto-distributed reveals that the thresholds are only changed by a scalar of $\left(\frac{\delta}{\beta k}\right)^k$, where $\delta > 0$ is the conventional exogenous death probability. Hence, all the well-known predictions of the Melitz model—such as selection into export activity according to productivity or intra-industry reallocations—also apply in the extended version. Thus we do not elaborate these standard features that also apply to the dynamic model and which have been repeatedly shown to match the empirical facts. Instead we turn to the predictions of export market exit and firm survival from above.

3.5 *Proof of the Theoretical Predictions*

From the above model elements and (8) it follows that in the extended dynamic version of the model firm death is deterministic conditional on the firm-specific productivity draw. Thus the exogenous firm death is replaced by a model endogenous firm closure decision.

Predictions 1 and 2, i.e. that more productive and larger firms survive longer, follow directly from (8) when noting that entry size increases in marginal productivity.

Furthermore, from (8) and the declining density of the Pareto distribution it follows that for each cohort the absolute number of exits declines with maturity. Accordingly, in the age distribution of a given exit wave the most recent entries have the largest share (Prediction 3).[7]

The market shares of a firm with productivity φ on the domestic and export markets are

$$s_t(\varphi) = \frac{\frac{E}{P_t}\left(\frac{p(\varphi)}{P_t}\right)^{-\sigma}p(\varphi)}{E} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\varphi^{\sigma-1}P_t^{\sigma-1} \tag{11}$$

$$s_t^x(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\tau^{1-\sigma}\varphi^{\sigma-1}P_t^{\sigma-1} \tag{12}$$

[7] It can be shown that the age distribution for the group of active firms exiting the market at any point in time is exponential with parameter βk (see the Schröder and Sørensen 2012 for details on distributions).

Combining (4), (5), (11) and (12), we find that $\pi_t(\varphi) = \frac{s_t(\varphi)E}{\sigma} - f$ and $\pi_t^x(\varphi) = \frac{s_t^x(\varphi)E}{\sigma} - f_x$. Accordingly, flow profits increase in the market shares and exiting firms will thus be small (have a small market share), i.e. Prediction 4. Market shares decline over time due to a decreasing price level, P_t, and exiters thus observe periods of declining market shares prior to exit (Prediction 5). Finally, exporters that cease their exporting activity continue for a while to exist as pure domestic firms, since $\phi_x^* > \phi^*$, c.f. footnote 5. This proves Prediction 6.

4 Data and Empirical Strategy

We assess the central predictions of the model provided in Sect. 2 using a sample of Danish manufacturing firms spanning the years 1996–2007, which is provided by Statistics Denmark. The observational unit is the firm. We consider a sample of manufacturing firms defined by the latest NACE revision on the two-digit level.

For the sake of comparability, our sample is defined as in the related literature. In particular, we follow Wagner (2009) and compare different exit cohorts to their surviving counterparts. A firm exit is defined as a pattern where a firm exists in time t but not in $t+1$ and does not reappear during the sample period. A survivor firm of time t is correspondingly defined as a firm which exists at least in time $t+1$, $t+2$, and $t+3$ (i.e. what happens after $t+3$ does not enter the classification). Firms with other existence patterns are not taken into account, such that the total number of firms N_t consists of survivors S_t and dying firms D_t. By construction, we observe for each cohort a firm's immediate pre-exit performance at time t. As we require surviving firms to survive at least 3 years longer than the exiting firms, the last observable and comparable exit cohort is from the year 2004. Moreover, Statistics Denmark has changed reporting rules in 1999. Before the change, all private firms subject to VAT had to report to the statistical office, irrespective whether they were truly active. After 1999 all truly active firms had to report to the general firm statistics and are thus included in our sample. Thus, the first exit cohort we consider is the one observed in 2000. We discard all firms which exit from the domestic market, but re-enter at a later point in time, as this pattern is likely to be due to misreporting rather than true firm closure. This initial cleaning decreases the sample of manufacturing firms by 1.17 %. Finally, we require that an existing firm employs at least one full-time employee. Overall, for the sampling period from 1996 to 2007, we observe 3080 distinct manufacturing firms and focus on the exit cohorts of the years 2000, 2001, 2002, 2003 and 2004.

In order to classify the export status of firms, any firm with positive export sales is defined to be an exporter. In our analysis, we exclude the exporters who trade with oil products or who export to unknown or undisclosed countries of destinations. Moreover, we restrict our sample to countries included in the CEPII data set (215 countries of destination). An export survivor is a firm that exports to at least one country at each point in time t, $t+1$, $t+2$ and $t+3$. An export market exit is defined

accordingly to firm death, but we do not generally rule out re-entry of the export market, such that an export exiting firm exports at time t, but does not export in time $t+1$, $t+2$, $t+3$ (but it may start again subsequently). In light of the related literature, it is likely that a firm's export knowledge fully decays in this time (Roberts and Tybout, 1997). With this definition, we compare continuous exporters with export stoppers rather than with temporary exporters, which matches most closely to the theoretical model.

As a matter of convention we subsequently use the term exit to refer to export market exit as well as to firm death. The respective tables presented below specify which type of exit is considered.

Productivity is measured as total factor productivity, which we estimate separately at the industry level using the method suggested by Levinsohn and Petrin (2003), thereby using the Stata function provided by Petrin et al. (2004).[8]

Some First Observations Based on the above specified sample, Table 1 displays the exit rates by exit cohorts for all manufacturing firms, pure domestic firms and exporters. Firms are regarded as small (large), when they exhibit less (more) than median employment. Table 1 reveals three main points. First, exit rates have increased between 2000 and 2004 for both domestic and exporting firms, leading to a total increase from 2.5 to 4.7 %. Secondly, firms with less than median employment, i.e. small firms, account for the larger share of exiting firms

Table 1 Exit rates by cohort, by firm type and firm size

	Overall	Small	Large
All firms			
2000	2.5	1.4	1.0
2001	2.5	1.7	0.8
2002	3.2	2.0	1.3
2003	3.1	2.1	1.0
2004	4.7	2.8	1.9
Domestic firms			
2000	2.8	2.1	0.7
2001	3.0	2.8	0.2
2002	3.0	3.0	0.0
2003	4.0	2.6	1.4
2004	6.0	5.3	0.6
Exporters			
2000	2.4	1.2	1.1
2001	2.3	1.4	0.9
2002	3.3	1.7	1.5
2003	2.9	2.0	0.9
2004	4.4	2.2	2.1

[8]Full results available from the authors upon request.

Table 2 Exit rates by industry and firm size: all firms, all years

	Overall	Small	Large
Basic metals	3.6	1.6	2.0
Beverages	0.0	0.0	0.0
Chemicals and chemical products	1.4	0.8	0.6
Coke and refined petroleum products	3.8	3.8	0.0
Computer, electronic and optical products	3.5	0.9	2.6
Electrical equipment	3.4	2.4	1.0
Fabricated metal products	2.4	1.5	0.9
Food products	1.9	1.1	0.8
Furniture	5.9	3.6	2.2
Leather and related products	1.6	1.6	0.0
Machinery	2.8	1.8	1.0
Motor vehicles, trailers and semi-trailers	0.9	0.9	0.0
Non-metallic mineral products	2.5	2.2	0.4
Other manufacturing	2.9	2.1	0.8
Other transport equipment	4.4	2.2	2.2
Paper and paper products	5.7	1.9	3.8
Printing and reproduction of recorded media	5.2	4.0	1.2
Rubber and plastic	2.6	1.0	1.5
Textiles	3.4	3.3	0.2
Tobacco products	0.0	0.0	0.0
Wearing apparel	2.7	2.4	0.3
Wooden products	3.7	2.1	1.6

throughout the entire period and across both domestic and exporting firms. Overall, the exit rates in Denmark are slightly lower than the ones found for the German manufacturing sector for the period from 1975 to 2006 (Fackler et al., 2013). They hide considerable heterogeneity as the industry decomposition in Table 2 shows. Table 2 presents an industry decomposition of firm exit rates by exit cohort for all firms in a two-digit NACE industry. Firms are regarded as small (large), when they exhibit less (more) than median employment.

Among all manufacturing subsectors, the exit rate is highest for the furniture sector, amounting to 5.9 % for the years 2000–2004. In sharp contrast, beverages and tobacco do not show any exit in the period of concern. The observation that the larger share of exit is driven by small firms carries over to the majority of sectors, apart from Paper and Paper Products, Basic Metals as well as Computer, Electronic and Optical Products.

These different exit rates across industries can be related to considerable performance differences. Table 3 shows industry averages of total sales, employment and total factor productivity, as well as the respective growth rates. Whereas total sales naturally vary considerably across sectors, the overall picture in terms of sales growth seems to be more homogenous: Even though the extent of sales

Table 3 Industry characteristics, all years

	Obs.	Total sales		Employees		TFP	
		Mean	Avg. growth	Mean	Avg. growth	Mean	Avg. growth
Basic metals	272	138.4	7.2	99.2	0.5	100.8	0.04
Beverages	89	445.0	9.6	247.2	2.8	134.9	0.04
Chemicals and chemical products	512	196.8	5.2	108.7	1.2	119.7	0.03
Coke and refined petroleum products	28	1383.6	9.4	110.0	0.7	182.1	0.01
Computer, electronic and optical products	123	64.5	4.7	60.3	−0.3	97.1	0.003
Electrical equipment	313	103.5	2.2	84.9	−3.1	98.2	0.01
Fabricated metal products	1825	40.9	4.6	42.6	0.2	91.6	0.03
Food products	1042	485.6	6.9	227.2	3.2	115.8	0.03
Furniture	1281	59.3	2.4	53.8	−2.3	85.9	0.02
Leather and related products	65	222.0	−5.0	82.7	−5.7	91.1	−0.02
Machinery	1867	91.2	4.4	84.4	−0.4	103.7	0.02
Motor vehicles, trailers and semi-trailers	113	104.0	3.7	108.5	−4.4	100.9	0.1
Non-metallic mineral products	576	117.9	2.8	97.8	−1.6	111.5	0.004
Other manufacturing	654	134.4	3.2	74.5	−1.2	110.9	0.02
Other transport equipment	49	80.3	3.1	60.8	2.2	120.3	0.03
Paper and paper products	469	108.6	3.9	79.1	−0.01	107.6	0.03
Printing and reproduction of recorded media	941	35.5	1.3	31.7	−1.3	93.1	0.02
Rubber and plastic	1314	70.6	4.0	64.5	0.5	96.2	0.03
Textiles	600	46.1	1.6	36.8	−4.4	87.0	0.03
Tobacco products	27	731.7	5.0	233.6	−1.1	290.6	0.1
Wearing apparel	357	41.0	2.3	24.5	−6.3	97.7	0.01
Wooden products	720	69.4	4.0	63.9	−0.1	88.4	0.03

growth varies considerably, with the exception of Leather and Related Products, all industries exhibit positive sales growth. In sharp contrast, employment has declined for 14 out of 22 industries. The growth rates range from −6.3 % in Wearing Apparel to 3.2 % in Food Products, the latter one also being the one with the highest average employment level. With respect to TFP, growth rates are moderate and positive but close to zero for all industries with the major exception of Leather and Related Products being characterized not only by negative sales, but also by a negative TFP growth.

The Danish manufacturer landscape displays similar patterns of exporting behavior as found for other countries (see for example Mayer and Ottaviano 2007): Danish firms are active on international export markets—on average, the number of markets served by an exporting manufacturer in our sample amounts to 11, whereby 50 % of firms in our sample do not serve more than 5 markets. Generally, most variation with respect to the number of markets served occurs across firms, but only to a lesser extent within firms. Around 26 % of manufacturers export to only one country of destination, but contribute by less than 1 % to the overall export revenue generated during the sample period.

Subsequently, we test the theoretical predictions of the model. The main tool of analysis is the Kolmogorov-Smirnov test, which allows us to compare size and productivity distributions of exiting and surviving firms. The Kolmogorov-Smirnov test is a non-parametric test, which tests the Null hypothesis of equality of distributions against first order dominance of one distribution.

5 Results

The presentation of results follows the predictions from Sect. 2.

Prediction 1 The theoretical model suggests that higher productivity firms survive longer on both domestic and export markets. This is reflected in higher productivity found among all survivors within an exit cohort. Figure 1 shows kernel density estimates of firms TFP at time t split among those that exit in $t+1$ (exits) and those that remain active in the domestic market until $t+3$ at least (survivors). Departing from pure visual inspection, Table 4 displays results for a t-test of equal means against three different alternative hypotheses, namely (a) survivors exhibit a higher average TFP, (b) means differ and (c) exiters exhibit a higher TFP. For all five cohorts, we reject the Null hypothesis in favor of mean inequality, and we reject mean equality between exits and survivors in favor of a higher average TFP of surviving firms. Our results, which are in line with Farinas and Ruano (2005), as well as Wagner (2009), thus confirm the theoretical prediction with respect to both firm death and export market exit as in both cases the productivity distribution of survivors first-order stochastically dominates the one of the exiting firms.

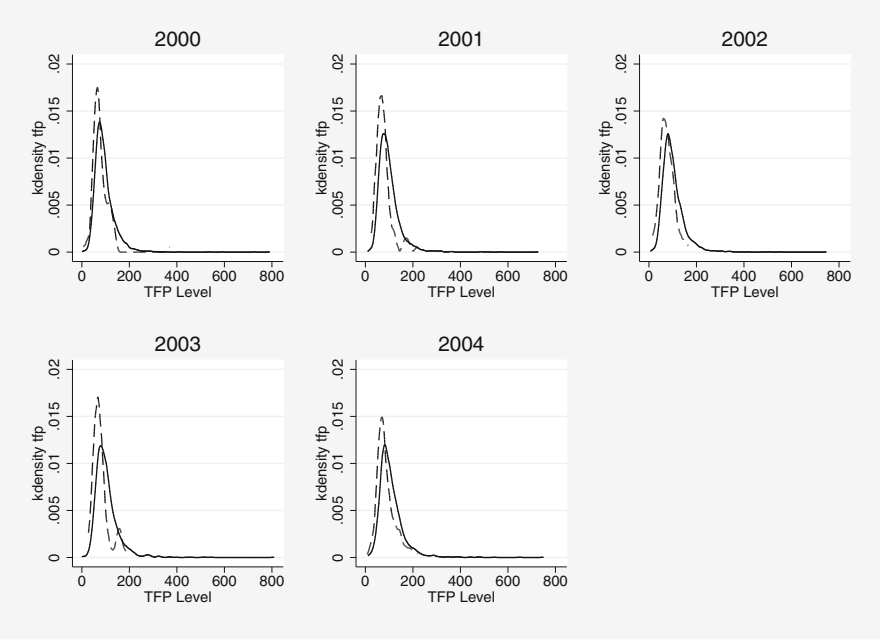

Fig. 1 Kernel Density of TFP for 5 years: Exits (*dashed line*) versus Survivors (*solid line*)

Prediction 2 Firms which enter a market with a higher level of productivity (and thus sales) survive longer. A high entry productivity should feed back into a persistently higher productivity level, leading to a persistent better chance of surviving. Along the lines of Farinas and Ruano (2005) and Wagner (2009), we consider the size distributions of exits and survivors at their time of entry. Since the model exhibits a clear link between productivity and size, we only examine the latter here. For firms which have been in the sample in the first year, the first observation on sales is used as a proxy for entry size. Table 5 shows that the role of entry size differs when considering export market exit and firm death. Concerning export market exit, only for two cohorts, we reject equality of the initial sales distributions of future exits and survivors. In case of firm death, equality of the initial sales distributions cannot be rejected. This result needs to be read with care, in particular since the proxy for entry size of those firms which are in the sample right from the beginning is likely to be imprecise.

Prediction 3 According to the theoretical results, younger firms constitute the majority share in a given exit cohort. This is strongly corroborated for both export market exit and firm death. As depicted in Table 6, young firms account for at least 75 % of export market exits, and for at least 60 % of dying firms. Thus, also for the Danish case, 'the liability of newness' is confirmed (compare Fackler et al. 2013, for a comprehensive literature review), even though results need to be taken with a pinch

Table 4 Productivity differences between exits and survivors

	Mean		t-test ($H_0 : \mu_S = \mu_E$)			Kolmogorov-Smirnov test ($H_0 : F_S = F_E$)		
	Survivors	Exits	$p(\mu_S > \mu_E)$	$p(\mu_S \neq \mu_E)$	$p(\mu_S < \mu_E)$	$p(F_S \succ F_E)$	$p(F_E \succ F_S)$	$p(F_S \neq F_E)$
Firm death								
2000	95.4	79.8	0.005	0.010	0.995	0.000	0.976	0.000
2001	99.2	77.0	0.000	0.000	1.000	0.000	0.999	0.000
2002	101.6	75.3	0.000	0.000	1.000	0.000	1.000	0.000
2003	103.8	79.0	0.000	0.000	1.000	0.000	0.996	0.000
2004	107.9	85.8	0.000	0.000	1.000	0.000	1.000	0.000
Exp. Mkt. exit								
2000	95.0	74.8	0.000	0.000	1.000	0.000	1.000	0.000
2001	97.8	82.4	0.000	0.000	1.000	0.004	1.000	0.003
2002	100.8	81.8	0.000	0.000	1.000	0.000	0.999	0.000
2003	103.3	79.4	0.000	0.000	1.000	0.000	0.994	0.000
2004	107.3	91.2	0.000	0.000	1.000	0.000	0.999	0.000

Table 5 Entry size differences among survivors and exits

	Mean		t-test ($H_0 : \mu_S = \mu_E$)			Kolmogorov-Smirnov test ($H_0 : F_S = F_E$)		
	Survivors	Exits	$p(\mu_S > \mu_E)$	$p(\mu_S \neq \mu_E)$	$p(\mu_S < \mu_E)$	$p(F_S \succ F_E)$	$p(F_E \succ F_S)$	$p(F_S \neq F_E)$
Firm death								
2000	102.1	89.2	0.238	0.477	0.762	0.687	0.394	0.720
2001	103.0	69.4	0.015	0.030	0.985	0.215	0.929	0.132
2002	103.4	62.3	0.000	0.000	1.000	0.867	0.566	0.516
2003	102.8	66.5	0.011	0.023	0.989	0.224	0.475	0.135
2004	101.9	74.2	0.027	0.053	0.973	0.306	0.331	0.175
Exp. Mkt. exit								
2000	101.0	75.7	0.036	0.073	0.964	0.097	0.926	0.060
2001	101.2	65.9	0.002	0.004	0.998	0.511	0.922	0.289
2002	102.4	54.4	0.000	0.000	1.000	0.421	0.964	0.236
2003	102.6	48.3	0.000	0.000	1.000	0.000	0.569	0.000
2004	101.8	69.5	0.005	0.010	0.995	0.120	0.755	0.071

Table 6 Firm age and exit

	Share of young firms	
	Export exit	Firm death
2000	75.3	60.0
2001	80.5	90.0
2002	83.3	81.0
2003	85.5	83.3
2004	87.1	86.4

of salt given our measurement of firm age: Due to data limitation, we compare 'old' firms founded in 1997 to 'young' firms founded afterwards, and thus the perspective is restricted to a comparison of relatively young firms. Thus, we cannot consider the 'liability of oldness' to the full extent (see Fackler et al. 2013, for a discussion).

Prediction 4 We test whether only small firms leave a market by using a Kolmogorov-Smirnov test. If only small firms leave a market, the distribution of total sales observed by exiting firms should be stochastically dominated by the one of stayers. Table 7 provides clear evidence in favor of the model's prediction. For all cohorts, the Kolmogorov-Smirnov test rejects the Null hypothesis of equal sales' distributions in favor of first order stochastic dominance of the stayers. This result is found for firms leaving the export market as well as for firms that quit entirely.

Prediction 5 Firms experience a decline in market share before exit (i.e. export market exit or firm death). Distinguishing between exits and survivors, Table 8 shows the share of firms which experience a decline in total sales, productivity (TFP) and employment prior to domestic exit (death) or export market exit. Sales, TFP and employment are calculated relative to the industry average at time t. In the case of export market exit, even in the cohort with the lowest share in sales declines (2000) more than 55.2 % of all exiting firms experience a decline in total sales prior to export market exit. This compares to 49.3 % of export stayers with declining sales in the same cohort. This pattern is in line with the theoretical prediction, and it is even more pronounced for firm death: The lowest share for exits with declining sales amounts to 62.9 % in 2000, which compares to a decline share of only 48.8 % among the surviving firms in the same cohort. This pattern is less pronounced for total factor productivity, but more pronounced for employment.

Prediction 6 The model predicts that firms which stop to export continue to exist as domestic firms before they eventually die. Table 9 shows that this is true for between 33 and 50 % of all firms that cease their exporting, i.e. they continue as pure domestic firms. This relatively low percentage of firms continuing domestically is partially due to our definition of export exits. Moreover, it is likely that one would find a different pattern for countries where the domestic market makes up for a larger share of firms' sales in general, i.e. for Denmark we expect the home market to matter relatively little for the closure decision of firms.

Table 7 Pre-exit firm size differences between survivors and exits

	Mean		t-test ($H_0 : \mu_S = \mu_E$)			Kolmogorov-Smirnov test ($H_0 : F_S = F_E$)		
	Survivors	Exits	$p(\mu_S > \mu_E)$	$p(\mu_S \neq \mu_E)$	$p(\mu_S < \mu_E)$	$p(F_S \succ F_E)$	$p(F_E \succ F_S)$	$p(F_S \neq F_E)$
Firm death								
2000	106.7	52.2	0.001	0.003	0.999	0.008	0.993	0.006
2001	121.2	37.0	0.000	0.000	1.000	0.000	1.000	0.000
2002	127.5	38.1	0.000	0.000	1.000	0.000	1.000	0.000
2003	130.7	27.3	0.000	0.000	1.000	0.000	0.999	0.000
2004	134.8	67.9	0.003	0.006	0.997	0.000	0.998	0.000
Export market exit								
2000	104.4	41.6	0.000	0.000	1.000	0.001	1.000	0.000
2001	116.4	42.8	0.000	0.000	1.000	0.002	1.000	0.002
2002	123.9	39.6	0.000	0.000	1.000	0.003	0.998	0.002
2003	128.8	25.3	0.000	0.000	1.000	0.000	0.994	0.000
2004	133.4	65.3	0.001	0.002	0.999	0.001	0.988	0.001

Table 8 Decline in sales, productivity and employment before firm death and export market exit

	Firm exit				Export market exit		
	Share of firms with decline in				Share of firms with decline in		
	Sales	TFP	Employment		Sales	TFP	Employment
Survivors				*Stayers*			
2000	48.8	46.7	46.9	2000	49.3	46.5	47.8
2001	53.3	46.5	53.4	2001	55.9	47.6	55.5
2002	54.4	47.5	48.1	2002	53.5	46.7	49.2
2003	53.6	48.0	48.2	2003	54.0	48.9	48.9
2004	50.9	47.6	50.5	2004	49.6	46.8	50.7
Exits				*Exits*			
2000	62.9	46.8	72.6	2000	55.2	53.3	61.0
2001	72.6	56.5	71.0	2001	65.2	52.8	62.9
2002	71.3	60.0	70.0	2002	60.2	54.9	58.6
2003	65.8	60.5	75.0	2003	65.6	60.9	60.2
2004	72.8	57.6	81.4	2004	58.7	54.7	68.7

Table 9 Export market exit and domestic survival

Exit year	Domestically active exits (%)
2000	49.5
2001	39.3
2002	39.2
2003	42.1
2004	33.3

6 Taking the Evidence Back to the Model

Our empirical results confirm the testable predictions from theory to a large extent, but they also disclose some new patterns that require further investigation. Most prominently, although the model showed the pattern of shrinking sales for exiting and dying firms, it did not confirm the strict sequence whereby firm death is preceded by systematic exit from export markets. This hints at two potential model extensions. Firstly, the presence of sizable sunk export cost—opposed to per period fixed export costs—would generate an exit-all-markets-at-once pattern, i.e. given that sunk costs have been paid, all orders from abroad will be fulfilled until the firm closes. Secondly, the presence of strong cross subsidization within firms' activity portfolio could alter the exit pattern. In the case of equal size countries—the current set-up in the theory presented here—break-even patterns on the different markets (and hence market exit) have a clear hierarchy. If in contrast the home market is very small, say so small that it does not suffice to cover the production fixed costs of firms on its own, we expect that the model would generate simultaneous firms closure and exit from the last export markets. Another complementary perspective, which is not captured within the present model, is the number of destinations that a firm

services. With more foreign destinations served the potential for cross-subsidization (and hence exit patterns that are more complex than what is examined in the present model) should be expected. Finally the present model as well as our empirical investigation ignore the role of multi-product firms and product exit as opposed to firm exit. Moreover, an additional perspective would be the link between firm closure and the preceding export market serving pattern, thereby exploring the role of temporary exports in the sense of Békés and Muraközy (2012), and by contrasting it to export market exit.

Future research will have to examine such extensions.

7 Conclusion

The present paper deploys the Melitz (2003) extension of Schröder and Sørensen (2012) to address firm survival and export market exit in an analytically solvable dynamic framework. The central driver of this approach is the inclusion of exogenous technological progress, such that newer firms draw from an improved productivity distribution and accordingly older firms eventually will be ousted from the market place. From this model several predictions for export market exit and firm survival are derived. In particular, high productivity firms (at any point in time) are likely also to produce in the future, i.e. they survive longer. Entry size and productivity of firms on a given market are positively linked to the duration of serving this market. Large exporters do not exit export markets, but smaller exporters do, i.e. firms exit markets after they have lost market shares, and exporters that cease to export will still serve their domestic market.

We contrast these theoretical results with the empirical stylized facts based on Danish firm-level data. In a nutshell, our results confirm the following predictions of the model. First, firms experience a decline in market share prior to death and export market exit. Secondly, firms that die or quit their exporting activity are small. Thirdly, our results show that small firms constitute the largest share of exits in a given cohort. Fourth, our empirical findings suggest that the productivity distribution of survivors first-order dominates the one of exits on both domestic and export markets. We find that between 33 to 50 % of firms that stop exporting continue to exist domestically subsequently.

Overall, our combination of theory and data shows that the intense focus on export market entry and market serving modes that the recent decades of international economics research based on firm-level data has generated, can usefully and practically be complemented with a systematic investigation into export market exit patterns and firms survival. The present chapter has mapped out some initial guidance and results.

Acknowledgements The authors acknowledge financial support from the Tuborg Foundation.

References

Arkolakis, C. (2016). A unified theory of firm selection and growth. *The Quarterly Journal of Economics*, 131(1), 89–155. doi:10.1093/qje/qjv039.

Atkeson, A., & Burstein, A. (2010). Innovation, firm dynamics, and international trade. *Journal of Political Economy, 118*(3), 433–484.

Békés, G., & Muraközy, B. (2012). Temporary trade and heterogeneous firms. *Journal of International Economics, 87*(2), 232–246.

Caves, R. (1998). Industrial organization and new findings on the turnover and mobility of firms. *Journal of Economic Literature, 36*(4), 1947–1982.

Costantini, J. A., & Melitz, M. J. (2008). The dynamics of firm-level adjustments to trade liberalization. In E. Helpman, D. Marin, & T. Verdier (Eds.), *The organization of firms in a global economy*. Cambridge: Harvard University Press.

Dasgupta, P., & Stiglitz, J. (1981). Entry, innovation, exit: Towards a dynamic theory of oligopolistic industrial structure. *European Economics Review, 15*, 137–158.

Dunne, T., Roberts, M., & Samuelson, L. (1988). Patterns of firm entry and exit in United-States manufacturing-industries. *Rand Journal of Economics, 19*(4), 495–515.

Eaton, J., Eslava, M., Kugler, M., & Tybout, J. (2008). Export dynamics in Colombia: Firm-level evidence. In E. Helpman, D. Marin, & T. Verdier (Eds.), *The organization of firms in a global economy*. Cambridge: Harvard University Press.

Fackler, D., Schnabel, C., & Wagner, J. (2013). Establishment exits in Germany: The role of size and age. http://economics.ucr.edu/undergrad_program/worksheets/BSEC_worksheet.pdf *Business Economics, 41*, 683. doi:10.1007/s11187-012-9450-z.

Farinas, J. C., & Ruano, S. (2005). Firm productivity, heterogeneity, sunk costs and market selection. *International Journal of Industrial Organization, 23*, 505–534.

Felbermayr, G., & Spiegel, G. (2014). A simple theory of trade, finance and firm dynamics. *Review of International Economics, 22*(2), 253–274.

Impullitti, G., Irarrazabal, A., & Opromolla, L. D. (2013). A theory of entry into and exit from exports markets. *Journal of International Economics, 90*, 75–90.

Jäkel, I. C. (2014). Import-push or export-pull? An industry-level analysis of the impact of trade on firm exit. *Empirica, 41*(4), 747–775. doi:10.1007/s10663-013-9235-x.

Jørgensen, J. G., Schröder, P. J. H., & Yu, Z. (2012). Globalization beyond partitioning: Back to Krugman's world. *Review of World Economics, 148*(1), 73–87.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica, 50*(3), 649–670.

Klepper, S. (2002). Firm survival and the evolution of oligopoly. *Rand Journal of Economics, 33*(1), 37–61.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies, 70*(2), 317–342.

Mayer, T., & Ottaviano, G. (2007). *The Happy Few: The Internationalisation of European Firms*. Blueprints, Bruegel, No. 12.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725.

Petrin, A., Poi, B. P., & Levinsohn, J. (2004). Production function estimation in Stata using inputs to control for unobservables. *The Stata Journal, 4*(2), 113–123.

Roberts, M. J., & Tybout, J. R. (1997). The decision to export in Colombia: An empirical model of entry with sunk costs. *American Economic Review, 87*(4), 545–564.

Schröder, P. J. H., & Sørensen, A. (2012). Firm exit, technological progress and trade. *European Economic Review, 56*(3), 579–591.

Schröder, P. J. H., & Sørensen, A. (2014). A welfare ranking of multilateral reductions in real and tariff trade barriers when firms are heterogenous. *Review of International Economics, 22*(3), 423–443.

Stigler, G. (1958). The economies of scale. *Journal of Law and Economics, 1*, 54–71.
Wagner, J. (2009). Entry, exit and productivity: Empirical results for German manufacturing industries. *German Economic Review, 11*, 78–85.
Wagner, J. (2013). Exports, imports and firm survival: First evidence for manufacturing enterprises in Germany. *Review of World Economics, 149*, 113–130.

Characteristics of International Trade Intermediaries and Their Location in the Supply Chain

Hale Utar

Abstract Wholesale trade firms and their role in international trade are examined using transaction and firm level data sets from Denmark for the period 1998–2006. Compared to internationally trading manufacturing firms, wholesale firms trading internationally are found to focus on fewer countries with more products and lower unit values, and their involvement in international trade transactions differ significantly across industries. Manufacturing industries with more competitive structure, lower firm size, lower capital intensity, higher production fragmentation and lower export/import intensities are found to have higher wholesale share of export. The analysis shows that export and import premia also exist among wholesale trade firms, which is in line with the idea that these premia result from fixed costs of exporting/importing. Systematic differences between wholesale trade firms in intermediate goods markets versus in consumption goods markets are also documented and found critical in understanding the role of intermediaries in international trade. While in intermediate goods export wholesale trade firms' unit prices are found to be significantly higher than manufacturers unit prices of the same good, the opposite holds true for consumption goods export. Wholesale trade firms that specialize in export of intermediate goods are found to be more skill intensive and pay more in comparison to other exporting wholesale trade firms. The wage premium for exporters of intermediate goods for professional level occupations is robust to controlling for detailed firm and worker characteristics. The results suggest that theories highlighting the potential roles of intermediaries should take the intermediaries' location in the supply chain into account.

The analysis is conducted while the author visited the Labor Market Dynamics and Growth Center at Aarhus University. The author is grateful to Henning Bunzel and the late Dale Mortensen for facilitating the access to the confidential data bases of the Statistics Denmark and for their support. Support of The Cycles, Adjustment, and Policy research unit, CAP, and the School of Economics and Business, Aarhus University are acknowledged with appreciation.

H. Utar (✉)
Department of Economics, Bielefeld University, Bielefeld, Germany
e-mail: haleutar@gmail.com

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_9

1 Introduction

Studies show that the export and import behavior of manufacturing firms are important factors in improving industrial productivity as well as in technology and knowledge transfer between firms and countries (Bernard et al. 2007). In this light, the significant presence of intermediaries in manufacturing trade revealed by recently available customs data has received well deserved academic attention. While previous economics and management literature on intermediaries emphasizes the potential roles of intermediaries in facilitating trade between manufacturers and final consumers, recent literature that makes use of transaction-level international trade data focus on the fixed costs saving nature of trade intermediaries in export.

So far this literature treats the intermediation between two suppliers within the supply chain and intermediation between producer and final consumer uniformly. This paper documents empirical regularities in the role and significance of intermediaries in international trade from a small, open, and advanced country and provides insight into the determinants of intermediation in relation to product and industry characteristics in international trade. It is also shown that characteristics of trade firms differ systematically depending on the location of intermediation in the supply chain. To do that, transaction-level data from Danish customs, which include detailed information on all export and import transactions, are matched with firm-level accounting data that includes detailed business activity, sales, employment, capital, investment and other firm-level expenditures for the years 1998–2006.

Wholesale trade firms sell any type of goods. They simplify flow of goods, payments and information by acting as intermediaries between the manufacturer and the customer. The share of trade transacted via the manufacturer or a wholesale trade company varies substantially between different product groups or industries. Between 1993 and 2006, wholesale trade firms' and manufacturers' average share of imports are 57 and 31 % respectively, while their respective shares of export are 32 and 54 % in Denmark.[1] Intermediaries are found to be more active in import than in export, but in both export and import there is substantial heterogeneity in their involvement across industries. While the wholesale trade share is found to be on average 90 % in leather export, its share is 20 % in plastics export in Denmark.[2] Crude analysis of the data shows that the room for intermediaries is bigger in non-manufactured products, such as farm products, as would be expected. But even in the

[1]These numbers are higher in comparison to the similar numbers from the US as reported by Bernard et al. (2010) and China as reported by Ahn et al. (2011). For the year 2002, Bernard et al. (2010) reports the value share of intermediaries in export as 10 % and in import as 42 %. The value share of intermediaries in export for the year 2002 for China is reported as 29 % in Ahn et al. (2011).

[2]These statistics are average of data between 1993 and 2006.

manufacturing sector intermediaries, and wholesale trade firms in particular, play a significant role in export and import.[3]

To understand the relative concentration of wholesalers across manufacturing sectors in international trade, manufacturing sector characteristics are linked with the wholesale trade shares in those industries. Manufacturing industries with higher median firm size and capital intensity exhibit higher export share of wholesale trade firms. The results are in line with theories that relate fixed costs of exporting to the presence of wholesale trade firms.

Manufacturing industries with more competitive structure are found to have higher wholesale share of export. More specifically, controlling for firm and industry size, the wholesale trade share in export is negatively associated with firms' age and industry concentrations indices, as measured by the Herfindahl-Hirschman index and the 4-firm concentration index, and it is positively associated with entry and exit rates. The wholesale trade share in export is also found to be higher in manufacturing industries where the degree of production fragmentation is higher. In general the wholesale trade shares in export and import at the industry level are found to be negatively correlated with the export and import intensities of manufacturing firms. This finding points to the role of intermediaries as trade facilitators.

Wholesale trade firms trading internationally are found to focus on fewer countries with more products and lower unit values, confirming that the Danish data exhibit stylized facts similar to what has recently been highlighted on the role of wholesale trade firms in export and import.

But this paper also highlights a number of new and interesting features about the wholesale trade firms. First it shows that similar to manufacturing firms, wholesale trade firms that export are bigger, more productive and more capital-intensive; they pay more and employ more educated employees in comparison to non-exporters. Importing wholesale trade firms also share most of these features, showing that both export and import premia exist also among wholesale trade firms. These findings are in line with the idea that these premia result from fixed costs associated with exporting/importing (Bernard et al. 2007).

Second, there are important differences among wholesale trade firms in their involvement in international trade depending on the distance in the supply chain to the final consumers. In consumption goods markets, wholesale trade firms are found to focus on fewer countries with relatively more products in comparison to manufacturing firms. In intermediate goods markets, on the other hand, they are found to focus on products as well as countries. Comparing unit prices after controlling for detailed products and country fixed effects reveal that, while in consumption goods export wholesale trade firms' unit prices are significantly lower in comparison to manufacturers', their prices are significantly higher than manufacturers' in intermediate goods export. Similarly, in consumption goods

[3]The analysis focuses on wholesale trade firms, including export and import agents but excluding the retail sector. Because of this focus, the terms "intermediary" and "wholesale" trade firms are used interchangeably throughout the paper.

import wholesale trade firms' unit prices found to be significantly lower but this is not the case in intermediate goods import.

Firm-level data also show that wholesale trade firms specializing in export of intermediate goods are bigger than other exporting wholesale trade firms. Controlling for size, wholesale trade firms that specialize in export of intermediate goods are found to be more skill intensive and pay more in comparison to other exporting wholesale trade firms. They are not found to be significantly different in terms of capital-labor ratio, investment and labor productivity. The wage premium for exporters of intermediate goods for professional level occupations is robust to controlling for detailed firm and worker-level characteristics as well as intensity of high-tech goods sales. Wholesale trade firms in intermediate goods markets may be developing product specific knowledge. More demanding firm to firm communication required in global production chains may be one reason behind this wage premium. These results indicate that in order to understand the role of wholesale trade firms in international trade it is important to consider their distance in the supply chain to the final consumers.

The presence of middlemen or intermediaries in markets, in general, is motivated by several possible factors including adverse selection and moral hazard (Biglaiser 1993; Biglaiser and Friedman 1994), and the existence of search and information frictions (Rubinstein and Wolinsky 1987). Biglaiser (1993) and Biglaiser and Friedman (1994) predict that intermediaries sell higher quality products by acting as quality guarantor. The results presented here suggest that such quality sorting may be more relevant considerations in explaining intermediaries role in intermediate goods market compared to consumption goods markets.

Among the recent studies that use transaction level trade data, Bernard et al. (2010) highlight a number of stylized facts about intermediaries engaging in international trade in the US using data from 2002. This paper complements theirs by providing additional detail to the understanding of the nature of intermediaries in international trade.

Ahn et al. (2011) and Akerman (2010) extend the heterogeneous trade model with intermediation technology. Their models predict that the share of trade handled through intermediaries increases with fixed costs of exporting. Similarly Bernard et al. (2011) provide empirical regularities on the relationship between intermediaries' involvement in export and country specific fixed and variable costs using Italian data. Using Colombian and Chilean matched transaction data Blum et al. (2009) document that in a majority of exporter and importer matches at least one of the parties is a large international trader and that more than half of the Chilean exporters sell to a single Colombian importer. The authors then develop a model with matching frictions that replicate these findings. Felbermayr and Jung (2011) and Tang and Zhang (2012) on the other hand focus on hold-up problems to relate the country and product characteristics to the presence of export intermediaries. Except for Felbermayr and Jung (2011) and Tang and Zhang (2012), these papers do not look at the extent of intermediaries' involvement across industries and products,

which is the focus of this paper.[4] But all of these papers, including Felbermayr and Jung (2011) and Tang and Zhang (2012) uniquely treat intermediation between two suppliers and intermediation between producer and final consumer/retailers. The results presented in this paper show that empirical regularities may show contrast depending on the location of intermediation in the supply chain. The theories that highlight one or the other potential role of intermediaries should take these differences into account.

The paper is organized as follows. Data sets used in this study are described in the next section. Empirical analyses are presented in Sects. 3–5 followed by concluding remarks.

2 Data

The main data sets used in this study are transaction-level custom records and firm-level accounting data sets from Denmark, but other supplemental micro and macro data sets are also used, such as labor market surveys. This section summarizes the main data sets while details e.g. variable constructions, information on the additional data sets, are presented in the Appendix.

The firm-level data set (business statistics data) is compiled from survey results of firms that take part in an annual financial survey as well as from the annual tax reports, vat reports, and annual reports from incorporated companies. Wholesale trade firms are included in this data set starting from 1998. So the sample period used in this study is 1998–2006. The general business statistics include only firms that employ at least 0.5 FTE (full-time equivalent employment) and/or have had an estimated earnings of a certain size. Earning sizes are estimated differently for different industries. In the wholesale trade sectors, the lower limit of earnings is typically 500,000 Danish Kroner, while in the manufacturing industry, it ranges

[4]Felbermayr and Jung (2011) approaches the presence of trade intermediaries in export as a firm boundary problem. As in the spirit of Helpman et al. (2014) manufacturing firms face a trade off in their decision to choose an export mode due to the lack of enforceable cross-country contracts. They can use their own wholesale affiliate in the foreign country to avoid distortion due to hold up problem and incur fixed costs of distribution or that they use a trade intermediary but then face lower export revenues. Their model predicts productivity/quality sorting within industries similar to Ahn et al. (2011) and Akerman (2010). While their focus is still on the country specific costs, their model predicts firms producing high quality products with strong brand reputation are more likely to invest in distribution channels in foreign markets. Similarly Tang and Zhang (2012) consider a hold up problem in a heterogenous firm framework where intermediaries provide fixed cost saving technology. Distortions caused by the hold up problem in quality verification efforts necessary for foreign buyers drives the relationship between quality differentiation and the propensity to use an export intermediary. Their model predicts that the propensity to export via an intermediary decreases with vertical differentiation while it increases with horizontal differentiation of the products.

between 150,000 and 200,000 Danish Kroner. Table 14 in the Appendix provide summary statistics for the wholesale trade firms.

International trade data are available at the transaction-level starting from 1993. They contain firm id, the type of transaction (whether it is export or import), the value of transaction in Danish Kroner, the name and the code of the partner country, the amount of the transacted good, the unit of the amount, the name and the 8-digit combined nomenclature (CN) code of the good as well as the year of the transaction. The details of this data set is given in Pedersen (2009). For the years 1993–2006, firm id's in the transactions data sets are matched with the main industry affiliation of firms using supplemental data sets within Statistics Denmark. As a result of this match 89 % of the firm ids' in the export data and 94 % of the firm ids' in the import data are matched with industry affiliations.

As wholesale firms specialize in logistics, marketing and distribution, they can be expected to employ fewer employees than manufacturing firms. Figure 1 shows the distribution of size (the logarithm of employment and the logarithm of capital assets) among manufacturing firms and among wholesale firms for the year 2000. From the figure it is apparent that wholesale firms employ fewer employees in general and have lower level of capital assets.

Figure 2 shows that wholesale trade firms on average sell more in comparison to the manufacturing firms, but, as one expects, their rate of value-added over sales is on average much lower compared to manufacturing firms.

While wholesale trade firms employ less people, they pay more on average than manufacturing firms as indicated by Table 1 and they also employ more educated employees.

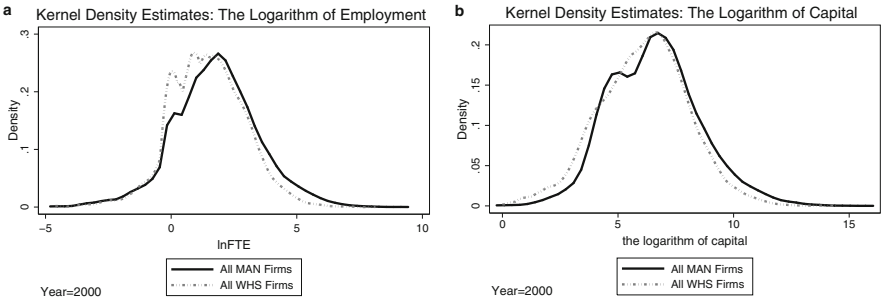

Fig. 1 Distribution of employment and capital among manufacturing and wholesale firms (values are expressed in constant 2000 prices in thousand Danish Kroner. *Source*: Statistics Denmark)

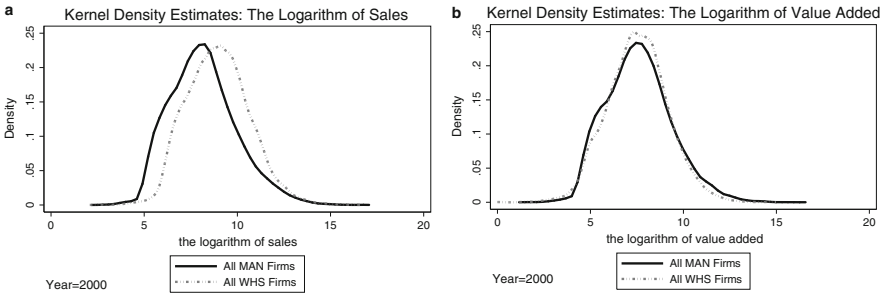

Fig. 2 Distribution of sales and value added among manufacturing and wholesale firms (values are expressed in constant 2000 prices in thousand Danish Kroner. *Source*: Statistics Denmark)

Table 1 Wholesale and manufacturing firms' characteristics

	WHS firms		MAN firms	
	Mean	# of obs	Mean	# of obs
Value added	8190.492	133,527	14,669.377	151,770
FTE	10.980	133,527	22.623	151,770
Average hourly wage	0.182	106,462	0.160	108,185
Professional occupation rate	0.184	128,862	0.094	145,879
College rate	0.171	106,462	0.125	108,185

The sample period is 1998–2006. Values are expressed in thousand 2000 Danish Kroner. FTE is the full-time equivalent number of employees. Professional Occupation Rate is the ratio of employees with at least mid-level occupations over the total number of employees. College Rate is the ratio of employees with at least some college education over the total number of employees

3 Understanding Across Industry Distribution of Wholesale Share

The share of trade transacted via manufacturers or wholesale trade companies varies substantially between different product groups or industries. Figure 3 presents the average shares of export/import transacted directly by manufacturers across broad product categories (CN chapters) between 1993 and 2006. Figure 4 presents the average shares of trade transacted via wholesale trade firms.[5] In the Appendix, Table 15 presents the respective shares across all broad product categories. Intermediaries are found to be more active in import than in export (except vegetable products, fats and oils, leather, footwear, and arms and arts) but on both sides there is also substantial heterogeneity across broad product categories. A higher share of wholesale trade in import compared to export is expected as wholesaler involvement

[5]Most of the trade is conducted via manufacturer and wholesale trade firms, but retail firms as well as other service firms are also present in international trade.

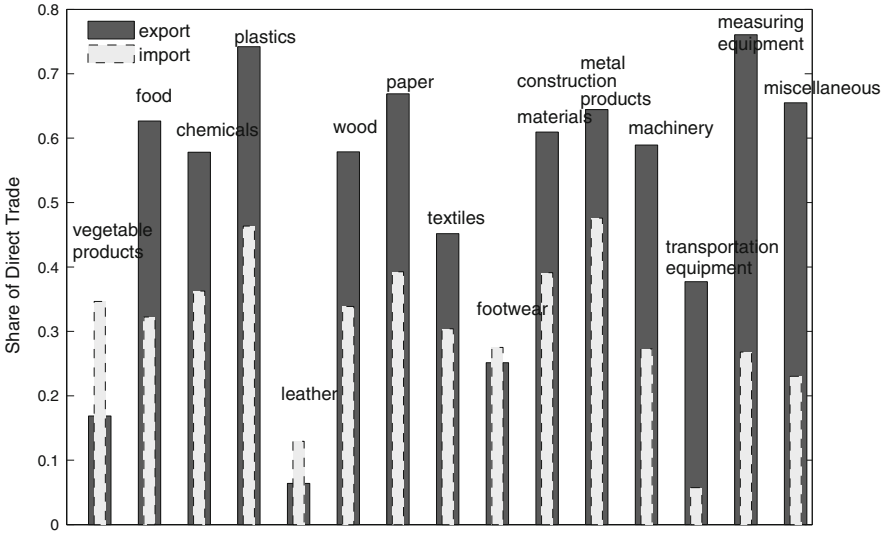

Fig. 3 Shares of manufacturing trade across broad product groups (*Source*: Statistics Denmark)

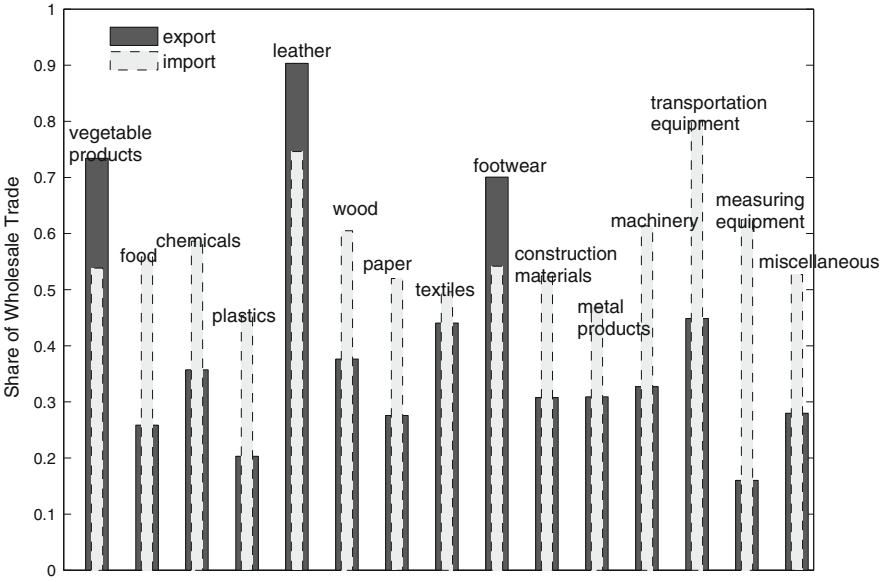

Fig. 4 Shares of wholesale trade across broad product groups (*Source*: Statistics Denmark)

at the later stages of the value chain is more likely due to the distribution and logistics services they provide.

Very low shares of direct manufacturing export are observed in vegetable products and leather coupled with proportionately high shares of wholesale trade

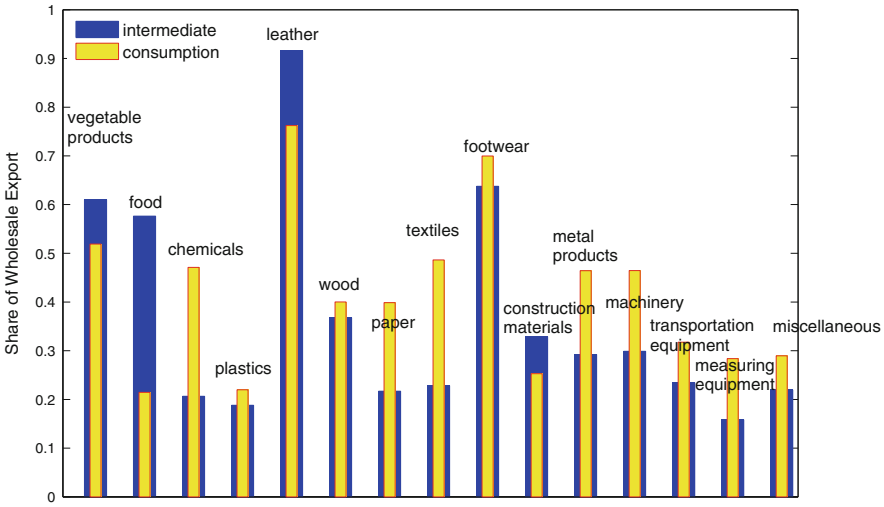

Fig. 5 Shares of export by wholesale traders across intermediate and consumption goods within industries (*Source*: Statistics Denmark)

in those product categories. This is most probably because these categories contain non-manufactured products mostly produced by farmers. What is most interesting is intermediaries' significant involvement in export of manufactured products (as also documented in Ahn et al. (2011) for China or Bernard et al. (2010) for the US). Part of their involvement in export could be due to the firm boundary issue that manufacturers of final products may prefer to outsource distribution services and hence intermediaries export as part of their service in the distribution of the products domestically and internationally. In Fig. 5, the wholesale shares in export are shown separately for intermediate and consumption or final goods across broad product categories. The figure shows that intermediaries are more involved in final goods' export except in vegetable products, food, leather and construction materials all of which largely contain non-manufactured goods (farming and mining).

In order to relate (manufacturing) industry characteristics to the wholesale export and import shares in that industry, the following industry-level equations are estimated[6]:

$$WHS_{jt} = \beta_0 + \beta_1^a IndustryCharacteristic_{jt} + \sum_t \delta_t^Y Year_t + \epsilon_{jt} \quad (1)$$

[6]Since the dependent variable is a share, the results are obtained using fractional logit model with robust standard errors as suggested by Papke and Wooldridge (1996). The results are robust to transforming the share variable as a log-odds ratio and are available upon request.

$$WHS_{jt} = \beta_0 + \beta_1^b IndustryCharacteristic_{jt} + \beta_2 FirmSize_{jt} + \sum_t \delta_t^Y Year_t + \epsilon_{jt} \quad (2)$$

$$WHS_{jt} = \beta_0 + \beta_1^c IndustryCharacteristic_{jt} + \beta_2 FirmSize_{jt} + \beta_3 NumberOfFirms_{jt} + \sum_t \delta_t^Y Year_t + \epsilon_{jt} \quad (3)$$

On the left hand side, there is the wholesale export/import share within manufacturing industry j (2-digit NACE) at year t.[7] On the right hand side, there are an industry characteristic and year fixed effects. Equation (2) controls for firms' size, $FirmSize_{jt}$, as measured by the logarithm of the median firm's employment within manufacturing sector j. Equation (3) additionally controls for industry size, $NumberOfFirms_{jt}$, which is the logarithm of the number of manufacturing firms in industry j at year t.

The industry characteristics are technology, openness, market structure and product type indicators for the corresponding manufacturing industries. Technology indicators are the median capital intensity as measured by the ratio of total fixed assets over total revenue (Capital Intensity), and production defragmentation as measured by the median of the ratio of the value-added over the value of production (Production Defragmentation). Openness indicators are export/import intensity defined as the median of the ratio of the export/import value over the total revenue (Export/Import Intensity). Market structure indicators are the logarithm of the median firm's age (Log Firm Age), 4 Firm Concentration Rate, Herfindahl-Hirschman Index (HHI), and finally Entry and Exit Rates.[8]

Table 2 reports estimates of β_1^a, β_1^b, and β_1^c in Eqs. (1) to (3) for export and Table 3 reports the same estimates for import.

The results on export as reported in Table 2 show that the wholesale share in export is negatively correlated with the number of manufacturing firms and the median firm-level employment in that industry. These findings are intuitive: Bigger firms or firms in industries with higher comparative advantage (as indicated by the number of producing firms) are more likely to pay the sunk entry costs of exporting a la Melitz (2003) and export directly rather than through wholesale trade firms.[9] In relatively big industries manufacturing firms may also have more opportunity to learn from each other about potential export opportunities.

[7] CN product codes are matched with 2-digit industry (NACE) codes using correspondence tables between prodcom and CN provided by EuroStat RAMON.

[8] The median industry characteristics are calculated using firm-level data on the manufacturing industry between 1998 and 2006. Herfindahl-Hirschman Indices and 4-firm concentration indices are calculated by taking both domestic and foreign sales into account.

[9] Recent studies emphasize a role of intermediaries as reducing fixed costs of exporting, e.g. Akerman (2010), Ahn et al. (2011), and Tang and Zhang (2012).

Table 2 Wholesale share in export and industry characteristics

	$\hat{\beta}_1^a$	$\hat{\beta}_1^b$	$\hat{\beta}_1^c$
NumberOfFirms	−0.108*	−0.319***	
FirmSize	−0.239***		−0.471***
Capital intensity	−3.611***	−3.656***	−1.765**
Production defragmentation	−6.223***	−5.549***	−3.705***
Export intensity	−3.881***	−1.309	−4.837***
Import intensity	−3.830***	−0.696	−7.668***
Market structure			
Log Firm Age	−0.871**	−0.652*	−1.013***
4 firm concentration rate	0.007*	0.012***	−0.011**
Herfindahl-Hirschman index	−0.000	0.000	−0.000***
Entry rate	8.260***	6.655***	5.478***
Exit rate	7.768***	6.774***	4.035*

The results are obtained using the generalized linear model with binomial family, logit link and robust standard errors. The number of observations in all regressions is 195. The sample period is 1998–2006

*, ** and *** indicate significance at the 5 %, 1 % and 0.1 % levels respectively

Capital intensity, and production (de)fragmentation are found to be negatively associated with the wholesale share in export. This holds true even after controlling for the median (manufacturing) firm size and the number of (manufacturing) firms as well. Manufacturers may prefer to control distributional channels as a part of brand and product differentiation (Dent 2008) especially for industries that exhibit increasing returns to scale or industries with lower degree of production fragmentation.[10] Negative relationships between export and import intensities and wholesale trade share in export are in line with the trade facilitator role of traders where export and import are relatively rare activities.

The wholesale trade share in export is also found to be higher in more competitive industries as indicated by negative and significant coefficient estimates of the concentration indices (HHI and 4-firm) and firms' age and positive and significant coefficients of entry and exit rates of manufacturing firms. Lower entry barriers in more competitive industries must allow small and young firms to operate easily and these firms are more likely to export through intermediaries. Concentration of industries may also be driven by increasing returns at the firm level so firms may be less likely to outsource distribution services in less competitive industries.

Table 3 presents the results for import. As in export, the wholesale import share is negatively related with the number of manufacturing firms operating in

[10] A company with a high level of brand recognition may be hurt by using the same distribution channels as used for cheaper generic products. Consider a product with a highly advertised specific function sold together with a cheaper alternative. The distributors may extract higher profit margin from the cheaper alternative by selling it together with the expensive one so that they can get a price which is close to the price of the expensive one.

Table 3 Wholesale share in import and industry characteristics

	$\hat{\beta}_1^a$	$\hat{\beta}_1^b$	$\hat{\beta}_1^c$
NumberOfFirms	−0.089*	−0.139***	
FirmSize	−0.005		−0.114*
Capital intensity	−1.409*	−1.408*	−0.663
Production defragmentation	−2.998***	−3.237***	−2.598**
Export intensity	−2.086***	−3.165***	−4.762***
Import intensity	−1.078	−1.875	−5.264**
Market structure			
Log firm age	0.126	0.144	−0.017
4 firm concentration rate	0.002	0.003	−0.012**
Herfindahl-Hirschman index	0.000	0.000	−0.000*
Entry rate	1.511	1.829	1.603
Exit rate	2.051	2.637*	1.532

The results are obtained using the generalized linear model with binomial family, logit link and robust standard errors. The number of observations in all regressions is 195. The sample period is 1998–2006

*, ** and *** indicate significance at the 5 %, 1 % and 0.1 % levels respectively

the industry and the median firm size, although the size effect is not found to be as important as in export. While all other characteristics have the same signs as in export, some such as median capital intensity, median firm's age, entry and exit rates are not found to be significantly correlated with the wholesale import share. A weaker relationship between the share of wholesale import and manufacturing characteristics is expected since the manufacturing industry constitutes only part of the customers of the import traders as the import traders also import for retail and service sectors.

4 Comparing Manufacturing and Wholesale Trade Firms in International Trade

4.1 Firm-Level Differences

Table 4 reports that on average wholesale trade firms export ten 8-digit products, while manufacturing firms export eight products during the 1998–2006 period. Wholesale trade firms export to on average 4.5 countries while manufacturing firms export to 7.4 countries. That is, wholesale exporters export more products to less countries in comparison to manufacturing firms. This also holds true for imports, but to a lesser extent.

To examine differences between wholesale intermediaries and manufacturing firms engaging in international trade within broad product categories, Eq. (4) is

Table 4 Manufacturing and wholesale trade firms in international trade

	WHS firms		MAN firms	
	Mean	# of obs	Mean	# of obs
Number of exported products	10.005	63,794	8.013	45,147
Number of export countries	4.547	63,794	7.413	45147
Number of imported products	18.130	83,783	15.032	44,596
Number of import countries	5.013	83,783	6.113	44,596

International trade data set that is matched with industry classifications is used. The sample period is 1998–2006. Products are defined at the 8-digit CN level

estimated separately with export and import transaction data aggregated at the firm-level.

$$x_{it} = \beta_0^{WH} + \beta_1^{WH} I(WH_{it}) + \sum_t \delta_t Year_t + \sum_{it} \gamma_{it} Industry_{it} + \epsilon_{it} \quad (4)$$

where $I(WH_{it})$ is an indicator whether firm i is an wholesale trade firm, and x_{it} denotes characteristics (in logarithm) of firm i at period t: value of export/import, average price of exports/imports, the number of products exported/imported, the number of countries, the number of export/import transactions, and the number of years in the export/import market. The sample only includes firms that are identified by manufacturer or wholesale trade firms so the estimates of β_1^{WH} indicate the percentage difference in the characteristics for wholesale trade firms in comparison to manufacturing firms after controlling for industry and year fixed effects.[11]

The results for export and import are presented in column (a) and in column (b) of Table 5 respectively. Starting from export, the results show that wholesale trade firms' export is about 80 % less in comparison to manufacturing firms' exports, they sell to a smaller number of countries (relative country focus); on the other hand they sell about 6.3 % more products and they are shorter lived in the export market. Wholesale trade firms do not seem to have significant price differences at the firm level in comparison to manufacturing firms. In column (b) the results show that the value of import is about 28 % more than the manufacturers' imports, they import more products (about 9 %), their prices are lower, and they buy from a smaller number of countries (13.8 %). Finally wholesale importers are also found to be 8.3 % less tenured in the import market in comparison to manufacturer importers. The results indicate focus on country more than product as well as shorter tenure in the international markets are common properties of both export and import traders.

In order to gain insights into potentially distinct roles of intermediaries in consumption and intermediate goods markets, Eq. (4) is also estimated separately

[11] In Eq. (4) industry fixed effects are broad product category (CN Chapter) affiliations of firms. They do not indicate whether a firm is a manufacturer or trader of these products. CN Chapters are listed in Table 15 in the Appendix.

Table 5 Firm characteristics in international trade

	(a) Export	(b) Import
Firm characteristics	β_1^{WH}	β_1^{WH}
Log export-import value	−0.806***	0.281***
Log weighted average price	0.010	−0.101***
Log number of products	0.063***	0.091***
Log number of countries	−0.304***	−0.138***
Log number of transactions	−0.158***	0.083***
Log number of years export/import	−0.104***	−0.083***
Number of observations (min/max)	108,272/108,855	127,541/128,375
Intermediate goods trade		
Log export-import value	−0.961***	−0.224***
Log weighted average price	0.076**	0.005
Log number of products	−0.037*	−0.046*
Log number of countries	−0.330***	−0.152***
Log number of transactions	−0.252***	−0.043*
Log number of years export/import	−0.112***	−0.077***
Number of observations (min/max)	76,030/ 76,040	97,933/97,939
Consumption goods trade		
Log export-import value	−0.504***	−0.044
Log weighted average price	−0.014	−0.021
Log number of products	0.150***	0.101***
Log number of countries	−0.208***	−0.151***
Log number of transactions	−0.039	0.011
Log number of years export/import	−0.084***	−0.102***
Number of observations (min/max)	61,810/63,513	90,030/93,003

Robust standard errors, that are clustered for firms, are reported in parentheses. A constant term is included but not reported. All regressions include industry and year fixed effects. Industry is defined as a broad product category (CN Chapter) of a firm's activity. The sample only includes wholesale trade firms and manufacturers between 1998 and 2006. Value shares are used as weights in calculating firm-level average prices

*, **, and *** indicate significance at the 5 %, 1%, and 0.1% levels respectively

among intermediate and consumption goods.[12] To do that, before aggregating the transaction data at the firm-level, only transactions with either intermediate or consumption goods are kept in the sample. The results are presented at the lower panels of Table 5.

The result show that relative country focus among wholesale traders holds true whether we look at intermediate goods or consumption goods trade. But the number of exported/imported intermediate products is found to be smaller among traders

[12]Intermediate and consumption goods classification is based on BEC Rev. 3. See the Appendix for details.

in both export and import compared to manufacturers, indicating that intermediate goods traders may be building product specific knowledge. This seems not to be the case for consumption goods traders. The average price of intermediate goods exporters is found to be higher (9.2 %) in comparison to manufacturer exporters' prices. There is no statistically significant difference found between wholesalers and manufacturers' prices in consumption goods export. Only broad product categories are controlled for in these regressions so the price differences are expected to contain differences in the type and quality of the products as well. This will be taken into account in the next analysis below.

4.2 Unit Price Differences

This section presents an analysis to understand if there is any systematic differences in the unit prices of goods transacted by intermediaries versus manufacturers. On the one hand, manufacturing firms may be able to extract more surplus and so charge a higher price by controlling distributional channels.[13] Models with productivity sorting with fixed costs of exporting on the other hand, predict higher unit prices for wholesale trade exporters. Intermediation, in general, may result in double marginalization if the markets are not competitive or intermediaries may have a value creating role by providing additional services.[14]

$$ln(p_{ifc}) = \gamma_0 + \gamma_1 I(W) + \gamma_2 I(R) + \sum_{ic} \lambda_{ic}(Product_i \times Country_c) + \epsilon_{ifc} \tag{5}$$

Here p_{ifc} is the unit price of good i (at CN-8 digit) imported/exported from/to country c by firm f. $I(W)$ and $I(R)$ are wholesaler and retailer dummies respectively.

To see if there is any systematic difference for unit price differential between wholesaler and manufacturer depending on the type of goods they transact, the wholesaler dummy is also interacted with an intermediate good indicator, $I(IG)$, as below.

$$ln(p_{ifc}) = \gamma_0 + \gamma_1 I(W) + \gamma_2 I(R) + \gamma_3 * I(W) * I(IG) + \sum_{ic} \lambda_{ic}(Product_i \times Country_c) + \epsilon_{ifc} \tag{6}$$

Tables 6 and 7 show unit price differentials between intermediaries and manufacturers for the period 1998 through 2006 for export and import respectively.

[13] Dent (2008) emphasizes that routes to market may involve product/brand differentiation. Models of hold-up predict that manufacturing firms producing higher quality choose to export directly. See for example, Tang and Zhang (2012).

[14] The models with adverse selection (e.g. Biglaiser 1993) predict that intermediaries on average sell higher quality products, and their average prices are higher.

Table 6 Unit price differences-export

Year	1998	1998	1999	1999	2000	2000	2001	2001	2002
Variables	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price
Wholesaler dummy (I(W))	−0.029**	−0.093***	−0.012	−0.074***	−0.013	−0.077***	0.007	−0.065***	−0.015
	(0.011)	(0.016)	(0.011)	(0.016)	(0.010)	(0.014)	(0.009)	(0.012)	(0.008)
Retailer dummy (I(R))	−0.036	−0.065	−0.054	−0.083*	−0.032	−0.062	0.046	0.014	−0.128***
	(0.033)	(0.035)	(0.039)	(0.041)	(0.033)	(0.034)	(0.030)	(0.031)	(0.023)
Wholesaler*intermediate good		0.161***		0.158***		0.169***		0.186***	
		(0.021)		(0.020)		(0.019)		(0.018)	
Product by market fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observation	188,618	188,618	199,497	199,497	214,210	214,210	240,978	240,978	311,288
Number of clusters	77,636	77,636	80,017	80,017	82,738	82,738	88,459	88,459	105,627
F	3.8	23.6	1.0	21.7	0.9	28.4	1.2	44.4	17.6

Year	2002	2003	2003	2004	2004	2005	2005	2006	2006
Variables	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price
Wholesaler dummy (I(W))	−0.062***	−0.069***	−0.142***	−0.081***	−0.153***	−0.132***	−0.214***	−0.092***	−0.201***
	(0.011)	(0.008)	(0.010)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.009)
Retailer dummy (I(R))	−0.145***	−0.129***	−0.153***	−0.048**	−0.075***	−0.417***	−0.451***	−0.220***	−0.273***
	(0.024)	(0.018)	(0.019)	(0.016)	(0.016)	(0.013)	(0.013)	(0.016)	(0.016)
Wholesaler*intermediate good	0.111***		0.174***		0.173***		0.196***		0.236***
	(0.014)		(0.014)		(0.013)		(0.012)		(0.014)
Product by market fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observation	311,288	337,461	337,461	339,485	339,485	452,254	452,254	355,370	355,370
Number of clusters	105,627	109,908	109,908	110,801	110,801	138,614	138,614	111,285	111,285
F	23.2	40.6	70.7	70.6	118.9	532.1	466.3	120.6	198.6

Robust standard errors are reported in parentheses. They are clustered for each CN-8 digit product and country pair. A constant term is included but not reported
*, **, and *** indicate significance at the 5 %, 1%, and 0.1% levels respectively

Table 7 Unit price differences-import

Year	1998	1998	1999	1999	2000	2000	2001	2001	2002
Variables	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price
Wholesaler dummy (I(W))	−0.029***	−0.056***	−0.053***	−0.068***	−0.038***	−0.059***	−0.025**	−0.055***	−0.039***
	(0.009)	(0.015)	(0.009)	(0.017)	(0.009)	(0.016)	(0.009)	(0.015)	(0.008)
Retailer dummy (I(R))	0.061***	0.046**	0.075***	0.067***	0.049***	0.036*	0.044**	0.027	0.003
	(0.013)	(0.015)	(0.014)	(0.018)	(0.013)	(0.017)	(0.013)	(0.016)	(0.011)
Wholesaler*intermediate good		0.046**		0.025		0.037*		0.052**	
		(0.017)		(0.018)		(0.017)		(0.017)	
Product by market fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observation	329,085	329,085	336,968	336,968	353,146	353,146	374,009	374,009	461,178
Number of clusters	76,640	76,640	78,425	78,425	80,313	80,313	84,352	84,352	97,949
F	47.0	33.1	109.5	72.9	53.3	36.8	32.3	25.8	19.1sss

Year	2002	2003	2003	2004	2004	2005	2005	2006	2006
Variables	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price	Log price
Wholesaler dummy (I(W))	−0.075***	−0.055***	−0.088***	−0.076***	−0.155***	−0.090***	−0.178***	−0.104***	−0.184***
	(0.014)	(0.008)	(0.013)	(0.009)	(0.014)	(0.009)	(0.014)	(0.010)	(0.016)
Retailer dummy (I(R))	−0.017	−0.016	−0.033**	−0.051***	−0.092***	−0.067***	−0.115***	−0.070***	−0.115***
	(0.013)	(0.011)	(0.012)	(0.012)	(0.014)	(0.012)	(0.014)	(0.013)	(0.016)
Wholesaler*intermediate good	0.067***		0.063***		0.142***		0.153***		0.139***
	(0.016)		(0.015)		(0.015)		(0.015)		(0.016)
Product by market fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observation	461,178	473,536	473,536	483,382	483,382	457,186	457,186	526,599	526,599
Number of clusters	97,949	98,941	98,941	99,433	99,433	94,218	94,218	101,158	101,158
F	14.5	28.7	19.1	42.4	45.0	58.2	59.5	70.6	57.6

Robust standard errors are reported in parentheses. They are clustered for each CN-8 digit product and country pair. A constant term is included but not reported

*, **, and *** indicate significance at the 5 %, 1%, and 0.1% levels respectively

In these regressions the sample only includes transactions that are conducted via either manufacturing, wholesale or retailer companies excluding other types of firms such as business service firms. So the estimate of γ_1 in Eq. (5) indicates a unit price differential for wholesalers in comparison to manufacturing companies after controlling for 8-digit product by country fixed effects in a given year.

The results show that wholesalers' price is on average lower compared to manufacturers, both in export and in import after controlling for detailed product (CN-8 digit) by country fixed effects. In export, the difference is not always significant and the F statistics are low before 2002. However, when the intermediate good dummy is interacted with the wholesale dummy, F statistics grow sizably and estimates of γ_1 become significant at the 1 % level. Estimates of γ_3, on the other hand are always positive and significant at the 1 % level. In 1998 wholesale prices are found to be on average about 2.9 % lower than manufacturers' prices regardless of the type of goods. But when one controls for intermediate goods, wholesalers' prices are found to be 9.3 % lower, while prices of intermediate goods of wholesalers are found to be about 7 % higher than manufacturers' intermediate goods' prices.[15] The findings are very similar across all the years, suggesting that wholesalers' involvement in export in intermediate and consumption goods markets respectively have important distinctions. These results are quite different from Ahn et al. (2011) where intermediaries' unit prices are shown to be higher than manufacturers' in China.[16] Bernard et al. (2010), on the other hand, show wholesale traders' unit prices on average lower than manufacturers. The results presented here for Denmark are in line with Bernard et al. (2010). But none of these studies look at the intermediate and consumption goods markets separately. Doing that, results show that the unit price differential between wholesalers and manufacturers in general depend on the type of goods, intermediate versus consumption. While unit prices of exporting wholesale companies are found to be lower in general compared to manufacturers' prices, their prices are found to be significantly higher for intermediate goods.

The results for import presented in Table 7 show that retailers' unit prices are significantly higher in comparison to manufacturers, probably indicating their closer distance to the final customers. The prices of wholesalers are also found to be lower in general. Wholesalers that import intermediate goods on the other hand are not found to have significantly lower prices in comparison to manufacturers' unit prices.

[15] Separate estimation of Eq. (5) among intermediate and consumption goods also confirm these findings. They are available upon request.

[16] Ahn et al. (2011) control for the size as measured by employment when analyzing unit price differences between manufacturers and intermediaries. Since wholesale trade firms are significantly smaller in terms of employment than manufacturers, one expects upward bias on the coefficients for unit prices of intermediaries they find. The different results obtained with Denmark as opposed to China may also be due to (potentially) higher share of intermediate goods in Chinese export data.

The models with adverse selection (Biglaiser 1993) predict that intermediaries on average sell higher quality products, and their average prices are higher. The results here suggest a possibility of adverse selection problems in the intermediate goods markets. The differences in market structure of intermediaries in intermediate and consumption goods markets may also result in differences in prices as the possibility of double marginalization depends on competitiveness of the markets. Tables 17 to 18 show that concentration patterns are also different for wholesalers and manufacturers across these two different types of products. While concentration patterns of wholesale trade firms and manufacturers are similar in general (Table 16); wholesale trade firms' concentration is higher in intermediate goods while manufacturers concentration is especially higher in consumption goods.

These results suggest hold-up problems that give rise to quality sorting of goods traded between manufacturers and wholesale traders may be more relevant in the consumption goods markets, while information frictions that give rise to adverse selection may be more important in the intermediate goods markets. These results overall suggest that in understanding the role of intermediaries in international trade it is critical to explicitly consider their location in the supply chain.

5 Export and Import Premia Among Wholesale Traders

International trade literature emphasizes the importance of export and import behavior in manufacturing firms' performances (Bernard et al. 2007). There is a large literature on the sources of export premia among manufacturing firms. The two competing hypotheses are whether export premia are due to self-selection of more productive firms into export markets or whether they are due to learning by exporting, which is often thought to happen via the buyer supplier link that foreign buyers actively or passively channel knowledge to the local suppliers. One implicit assumption mostly made to motivate the learning by exporting hypothesis is that knowledge is channeled into manufacturing processes. To further this understanding it is important to see whether similar export and import premia also exist among wholesale trade firms as these firms do not manufacture, so no such learning can drive possible premia.

Table 8 show that 38 (47) % of wholesale firms export (import) during the sample period. Table 8 also indicates that more than 80 % of wholesale firms that export also import. Using business statistics data from 2000, Fig. 6 shows that both exporting wholesale firms and importing wholesale firms are on average larger (employment and capital) than non-exporting and non-importing wholesale firms respectively. It also indicates that exporting wholesale firms are on average larger than importing wholesale firms.

Table 8 Percentage of exporters and importers

Year	WHS firms		
	Exporters	Importers	Exporter-importers
1998	0.342	0.445	0.279
2000	0.362	0.451	0.289
2002	0.404	0.489	0.332
2004	0.393	0.498	0.326
2006	0.388	0.498	0.330
1998–2006	0.379	0.474	0.311

Sample: Wholesale Trade Firms, 1998–2006. The sample excludes agents (NACE 5111–5119). *Source*: Statistics Denmark

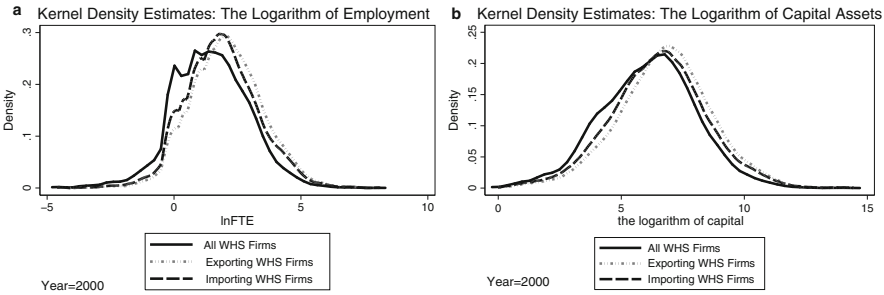

Fig. 6 Size distribution among wholesale firms (*Source*: Statistics Denmark)

To quantify possible export and import premia, Eqs. (7) and (8) are estimated using business statistics data for the wholesale trade firms for 1998–2006.

$$x_{ijt} = \beta_0^X + \beta_1^X I(XP_{ijt}) + \beta_2^X lnFTE_{ijt} + \sum_j \eta_j Industry_j + \sum_t \delta_t Year_t + \epsilon_{ijt} \quad (7)$$

$$x_{ijt} = \beta_0^M + \beta_1^M I(MP_{ijt}) + \beta_2^M lnFTE_{ijt} + \sum_j \eta_j Industry_j + \sum_t \delta_t Year_t + \epsilon_{ijt} \quad (8)$$

In these equations x_{ijt} denotes characteristics of wholesale trade firm i at period t in industry j in logarithm, *lnFTE* is the logarithm of the full-time equivalent number of employees, $I(XP_{ijt})$ is an export dummy and $I(MP_{ijt})$ is an import dummy. 4-digit industry and year dummies are also included. The coefficient β_1 indicates the percentage differences in the relevant firm characteristics controlling for size (measured by the number of employees), industry and time effects. The first column of Table 9 reports the differences between exporters and non-exporters among wholesale trade firms and the second column reports the differences between importers and non-importers among wholesale trade firms.

Table 9 Export and import premia

Firm characteristics	(a) β_1^X	(b) β_1^M
FTE (employment)	1.157***	1.077***
Value added per worker	0.277***	0.288***
Capital per worker	0.076***	0.017
Investment per worker	0.096***	0.040**
Average hourly wage	0.080***	0.109***
Average hourly wage of base level	0.046***	0.074***
Average hourly wage of professional level	0.070***	0.100***
$\frac{CollegeEducated}{TotalNumberofEmployees}$	0.087***	0.101***
$\frac{ProfessionalLevel}{TotalNumberofEmployees}$	0.106***	0.126***
Number of observations (max/min)	116,231/57,077	116,231/57,077

All regressions include the logarithm of employment except for the employment regression. Employment is measured as the full-time equivalent number of employees (FTE). All dependent variables are in logarithm. All monetary variables are deflated by the appropriate deflator. Standard errors are clustered for firms. *Source*: Business, Labor and Trade Statistics, 1998–2006, Statistics Denmark

*, **, and *** indicate significance at the 5 %, 1 %, and 0.1 % levels respectively.

The results show that both export and import premia exist among wholesale trade firms. More specifically exporting wholesale trade firms are larger, they pay higher wages (8 %), and invest more (9.6 %). Their technology is relatively more capital intensive (7.6 %) and they have higher labor productivity (27.7 %) and employ more skill intensive employees. Similar differences also hold between importers and non-importers except for the capital intensity. These results are in line with the sunk costs driven self-selection hypothesis.

The previous analysis revealed that export intermediaries that focus on intermediate goods charge higher prices on average in comparison to manufacturers. To see if export premia also change depending on the location of the intermediaries in the supply chain, Eqs. (9) and (10) are estimated among wholesale trade exporters.

$$x_{ijt} = \beta_0^I + \beta_1^I IGIntensity_{ijt} + \beta_2^I lnFTE_{ijt} + \sum_j \eta_j Industry_j + \sum_t \delta_t Year_t + \epsilon_{ijt} \quad (9)$$

$$x_{ijt} = \beta_0^C + \beta_1^C CGIntensity_{ijt} + \beta_2^C lnFTE_{ijt} + \sum_j \eta_j Industry_j + \sum_t \delta_t Year_t + \epsilon_{ijt} \quad (10)$$

Table 10 Intermediate goods premium among wholesale trade exporters. Robust standard errors are reported in parentheses. The dependent variables in columns a–d are the logarithm of FTE, the logarithm of the value added over FTE, the logarithm of the value of capital assets over FTE and the logarithm of the value of total investment over FTE respectively

Dependent variable	(a) lnFTE	(b) lnLabor Productivity	(c) lnCapital PerWorker	(d) lnInvestment PerWorker
IGIntensity	0.167***	0.014	−0.020	−0.034
	(0.034)	(0.013)	(0.030)	(0.028)
lnFTE		−0.105***	−0.133***	−0.163***
		(0.005)	(0.009)	(0.008)
Industry (4-digit) fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Adjusted R^2	0.065	0.087	0.064	0.043
F	11.843	21.758	18.216	15.166
Number of observations	49,950	49,656	48,934	46,174
Number of clusters	11,684	11,659	11,518	11,310
Consumption goods premium among wholesale trade exporters				
CGIntensity	−0.116***	0.002	−0.052	−0.056
	(0.037)	(0.014)	(0.033)	(0.030)
lnFTE		−0.105***	−0.134***	−0.164***
		(0.005)	(0.009)	(0.008)
Industry (4-digit) fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Adjusted R^2	0.064	0.087	0.064	0.044
F	11.592	21.645	18.218	15.204
Number of observations	49,950	49,656	48,934	46,174
Number of clusters	11,684	11,659	11,518	11,310

Source: Business, Labor and Trade Statistics, 1998–2006, Statistics Denmark
*, **, and *** indicate significance at the 5 %, 1 %, and 0.1 % levels respectively

$IGIntensity_{ijt}$ is the ratio of exported intermediate goods over the total number of exported goods by wholesale trade firm i in industry j at year t. Similarly, $CGIntensity_{ijt}$ is the ratio of exported consumption goods over the total number of exported goods by wholesale trade firm i in industry j at year t.[17]

[17]The sets of intermediate and consumption goods are not exhaustive. First, the definition of intermediate goods do not include fuels and lubricants, second there are also capital and non-classified goods. Hence two equations, one with intermediate goods intensity and the other with consumption goods intensity are estimated instead of one.

The results in Table 10 show that wholesale trade exporters that specialize in exporting intermediate goods instead of consumption goods are larger on average. But after controlling for size, there are no significant differences found between wholesale exporters in intermediate goods versus in consumption goods in terms of labor productivity, capital labor ratios and investment intensity.

Table 11 reports intermediate goods premia in employee wages and characteristics. After controlling for firm employment, there are significant differences in employee characteristics and wages between wholesale trade exporters depending on their location in the supply chain. Wholesale firms that export intermediate goods employ significantly more skill-intensive employees. This is manifested in a significantly higher ratio of employees with college education as well as a higher ratio of employees with professional level of occupations. They also pay higher wages. The coefficients in columns 2 and 3 in Table 11 indicate that one standard deviation increase in the intermediate goods intensity is associated with 0.030 and 0.053 standard deviation increases in average hourly wages of basic level employees and of professional level employees respectively.

To see if the wage premium can be explained by firm and labor characteristics, firm controls and worker characteristics at the firm level are added. The firm controls are whether a firm is single-plant, whether it is a proprietorship and firm's age. The worker characteristics are female employee ratio, college rate, average tenure of employees and a quadratic term of average tenure of employees.[18] The results in Table 12 show that the premium of the overall average hourly wage and the average hourly wage of basic level employees can be explained by firm and worker characteristics. The average hourly wage of professional employees is still found to be significantly higher with the intensity of intermediate goods. High-tech good intensity is added as an additional control in Table 13 but the results are not affected. The coefficients of high-tech good intensity are found to be positive and significant indicating an additional wage premium associated with high tech goods.

So wholesale trade firms that export intermediate goods employ more educated employees and pay them proportionately more. The higher average hourly wages of basic level employees can be explained by firm and employee characteristics. But even after controlling for these factors as well as intensity of high-tech goods sales, average hourly wages of professional level employees are still found to be positively associated with the intermediate goods intensity. This could be due to communication requirements for intermediate goods which increases the need for professional level functions.

[18]Empirical studies analyzing the impact of tenure on earnings usually find positive effect at a diminishing rate.

Table 11 Intermediate goods premium among wholesale trade exporters. Robust standard errors are reported in parentheses. The dependent variables in columns (a)–(e) are the logarithm of the average hourly salary, the logarithm of the average hourly salary of employees who do jobs that require basic-level skills, the logarithm of the average hourly salary of employees with professional and technical skill required occupations, the logarithm of the share of the college educated employees over the total number of employees, and the logarithm of the share of the employees with professional and technical occupations over the total number of employees respectively. Hourly wages do not include the benefits

	(a)	(b) lnAvgHourlyWage	(c) lnAvgHourlyWage	(d)	(e)
Dependent variable	lnAvgHourlyWage	Base level	Professional level	lnCollegeRate	lnProfessionalRate
IGIntensity	0.040***	0.028***	0.052***	0.135***	0.077***
	(0.008)	(0.008)	(0.009)	(0.020)	(0.018)
lnFTE	0.033***	0.019***	0.064***	−0.291***	−0.074***
	(0.002)	(0.003)	(0.003)	(0.007)	(0.006)
Industry (4-digit) fixed effects	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓
Adjusted R^2	0.103	0.048	0.094	0.375	0.140
F	29.005	15.660	18.706	60.820	25.624
Number of observations	49,015	37,359	34,994	32,870	35,016
Number of clusters	11,418	9250	8820	8037	8827

(continued)

Table 11 (continued)

Dependent variable	(a) lnAvgHourlyWage	(b) lnAvgHourlyWage Base level	(c) lnAvgHourlyWage Professional level	(d) lnCollegeRate	(e) lnProfessionalRate
Consumption goods premium among wholesale trade exporters					
CGIntensity	−0.047***	−0.033***	−0.054***	−0.102***	−0.068***
	0.008	(0.009)	(0.010)	(0.022)	(0.020)
lnFTE	0.033***	0.019***	0.064***	−0.290***	−0.073***
	(0.002)	(0.003)	(0.003)	(0.007)	(0.006)
Industry (4-digit) fixed effects	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓
Adjusted R^2	0.104	0.048	0.094	0.374	0.139
F	28.721	15.639	18.685	59.200	24.759
Number of observations	49,015	37,359	34,994	32,870	35,016
Number of clusters	11,418	9250	8820	8037	8827

Source: Business, Labor and Trade Statistics, 1998–2006, Statistics Denmark
*, **, and *** indicate significance at the 5 %, 1 %, and 0.1 % levels respectively

Table 12 Intermediate goods premium, controlling for firm and worker characteristics

Dependent variable	lnAvgHourlyWage	lnAvgHourlyWage among base level occupations	lnAvgHourlyWage among professional level occupations
IGIntensity	0.009	−0.001	0.032***
	(0.007)	(0.008)	(0.009)
lnFTE	0.022***	0.008*	0.065***
	(0.003)	(0.003)	(0.004)
Singleplant	−0.006	−0.011	0.009
	(0.006)	(0.006)	(0.007)
lnFirmAge	−0.010***	−0.007*	0.002
	(0.003)	(0.003)	(0.003)
Proprietorship	−0.127***	−0.035**	−0.150***
	(0.011)	(0.013)	(0.019)
FemaleRatio	−0.091***	−0.143***	−0.039*
	(0.013)	(0.015)	(0.017)
AvgTenure	0.048***	0.034***	0.030***
	(0.004)	(0.005)	(0.005)
$AvgTenure^2$	−0.001***	−0.000**	−0.000**
	(0.000)	(0.000)	(0.000)
CollegeRate	0.255***	0.211***	0.169***
	(0.015)	(0.016)	(0.017)
Industry (4-digit) fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓
Adjusted R^2	0.227	0.117	0.139
F	114.518	69.247	68.716
Number of observations	47,548	36,338	33,964
Number of clusters	10,925	8892	8437

Robust standard errors are reported in parentheses. IGIntensity is the number of intermediate goods that are exported over the total number of exported goods. Singleplant is a dummy variable indicating single plant firms. logFirmAge is the logarithm of a firm's age. Proprietorship is a dummy variable indicating the ownership type. FemaleRatio is the number of female employees over the total number of employees. AvgTenure is the average tenure of employees in a firm. $AvgTenure^2$ is the square of the average tenure of employees in a firm. CollegeRate is the number of employees with at least some college education over the total number. *Source*: Business, Labor and Trade Statistics, 1998–2006, Statistics Denmark

*, **, and *** indicate significance at the 5 %, 1 %, and 0.1 % levels respectively

Table 13 Intermediate goods premium, controlling for firm and worker characteristics and high-tech intensity

Dependent variable	lnAvgHourlyWage	lnAvgHourlyWage among base level occupations	lnAvgHourlyWage among professional level occupations
IGIntensity	0.009	−0.001	0.032***
	(0.007)	(0.008)	(0.009)
HTechIntensity	0.037***	0.042***	0.032***
	(0.008)	(0.008)	(0.010)
lnFTE	0.022***	0.008*	0.065***
	(0.003)	(0.003)	(0.004)
Singleplant	−0.006	−0.011	0.009
	(0.006)	(0.006)	(0.007)
lnFirmAge	−0.010***	−0.007*	0.002
	(0.003)	(0.003)	(0.003)
Proprietorship	−0.126***	−0.034**	−0.150***
	(0.011)	(0.013)	(0.019)
FemaleRatio	−0.088***	−0.139***	−0.036*
	(0.013)	(0.015)	(0.017)
AvgTenure	0.048***	0.034***	0.030***
	(0.004)	(0.005)	(0.005)
$AvgTenure^2$	−0.001***	−0.000**	−0.000**
	(0.000)	(0.000)	(0.000)
CollegeRate	0.252***	0.206***	0.165***
	(0.015)	(0.016)	(0.017)
Industry (4-digit) fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓
Adjusted R^2	0.228	0.118	0.139
F	109.889	68.732	66.815
Number of observations	47,548	36,338	33,964
Number of clusters	10,925	8892	8437

Robust standard errors are reported in parentheses. HTechIntensity is the number of high tech goods that are exported over the total number of exported goods. For definitions of other variables see Table 12. *Source*: Business, Labor and Trade Statistics, 1998–2006, Statistics Denmark
*, **, and *** indicate significance at the 5 %, 1 %, and 0.1 % levels respectively

6 Concluding Remarks

I examine the presence of intermediaries in international trade using data from a small, open and advanced country, Denmark. A number of stylized facts about the wholesale trade firms are highlighted to understand the role of these firms and distribution channels in international trade in general.

Wholesale trade firms are found to employ fewer but more educated employees than manufacturing firms. They sell more but generate lower value-added. They are

more likely to engage in import than in export but those wholesale trade firms that import are more likely to export as well. In general, they are found to focus on fewer countries with more products in comparison to manufacturing firms. Their share of export and import are found to be higher in manufacturing industries with lower export and import intensities and with higher degree of production fragmentation. Manufacturing industries with less market concentration, higher entry and exit rates and lower median firm age are found to exhibit higher export share of wholesale trade firms. Manufacturing industries with lower firm size and capital intensity are also found to have higher export share of wholesale trade firms, confirming theories that relate fixed costs of exporting to the presence of export intermediaries.

Characteristics of wholesale trade firms in international trade are shown to differ depending on whether they trade consumption goods or intermediate goods, trade in the latter of which has been growing steadily over the last decades due to internationalization of production and increasing outsourcing. Particularly, while wholesale traders in general command lower unit prices in international trade than manufacturers, exporting wholesale traders of intermediate goods command higher unit prices.

Both export and import premia exist also among wholesale trade firms. The results show that exporting wholesale firms are significantly larger, more capital and skill intensive in comparison to non-exporting wholesale trade firms, they pay higher wages and they are more productive. Similar differences also exist between importers and non-importers. Among trade exporters, firms that specialize in intermediate goods are found to have additional premia. They are found to be bigger, more skill-intensive both in terms of employees' education levels and in occupation characteristics and pay higher wages in comparison to other exporting wholesale traders. The wage premium for professional occupations remains even after controlling for detailed firm and worker level characteristics.

In existing literature intermediaries are treated uniformly regardless of their location in the supply chain. In this paper I show that wholesale trade exporters differ systematically depending on whether they function in intermediate or consumption goods markets. The results suggest that when analyzing the role of intermediaries, attention should be given to whether the goods in question are consumption goods or intermediate goods.

Appendix 1: Additional Tables

See Tables 14, 15, 16, 17, and 18.

Table 14 Summary statistics for WHS firms

	Mean	Median	Standard deviation	N
1998				
Employment	9	1	50	15,125
Total revenue	34,046	4753	203,006	15,125
Value added	7009	1293	37,996	15,125
Capital	2916	346	41,714	15,125
Investment	746	73	6197	15,125
Total Assets	15,745	2143	151,078	15,125
Average Wage	340	335	124	11,164
2000				
Employment	11	3	53	15,911
Total revenue	38,714	7105	232,682	15,911
Value added	7686	1738	44,705	15,911
Capital	3018	402	29,100	15,911
Investment	958	92.000	8001	15,911
Total assets	18,616	3201	195,725	15,911
Average wage	308	300	161	13,135
2002				
Employment	11	3	51	15,491
Total revenue	39,163	6917	220,882	15,491
Value added	8278	1991	41,469	15,491
Capital	2976	446	27,041	15,491
Investment	772	109	6310	15,491
Total assets	18,415	3093	225,051	15,491
Average wage	329	307	747	12,977
2004				
Employment	10	2	52	16,181
Total revenue	37,513	6591	224,294	16,181
Value added	7914	1725	41,687	16,181
Capital	3120	424	30,562	16,181
Investment	809	96	7237	16,181
Total assets	17,438	3045	159,774	16,181
Average wage	323	325	158	13,239
2006				
Employment	11	2	53	16,567
Total revenue	43,569	6694	303,968	16,567
Value added	8607	1723	52,732	16,567
Capital	3339	410	37,778	16,567
Investment	964	68	10,153	16,567
Total assets	21,632	3197	321,283	16,567
Average wage	341	321	415	13,539

Sample: Wholesale Trade Firms, 1998–2006. Values are expressed in constant 2000 prices in thousand Danish kroner. *Source*: Statistics Denmark

Table 15 Share of wholesalers across CN product categories (average between 1993–2006)

	Import		Export	
CN product categories	Wholesale	Manufacture	Wholesale	Manufacture
CN 1: Live animals; animal products	0.612	0.302	0.423	0.553
CN 2: Vegetable products	0.539	0.346	0.734	0.169
CN 3: Animal and vegetable fats and oils	0.266	0.688	0.505	0.483
CN 4: Prepared foodstuffs; beverages,..	0.559	0.322	0.259	0.627
CN 5: Mineral products	0.487	0.309	0.087	0.140
CN 6: Products of the chemical or allied industries	0.586	0.363	0.357	0.578
CN 7: Plastics, rubber and articles thereof	0.454	0.464	0.203	0.742
CN 8: Raw hides and skins, leather, furskins and articles thereof	0.746	0.129	0.903	0.064
CN 9: Wood and articles of wood	0.605	0.339	0.376	0.579
CN 10: Pulp of wood or of other fibrous cellulosic material	0.520	0.392	0.276	0.669
CN 11: Textiles and textile articles	0.496	0.304	0.441	0.452
CN 12: Footwear, headgear, umbrellas, sun umbrellas, walking sticks, seat-sticks,...	0.542	0.275	0.701	0.251
CN 13: Articles of stone, plaster, cement, asbestos, mica and similar materials; ...	0.525	0.391	0.308	0.609
CN 14: Natural or cultured pearls, precious or semi-precious stones, precious metals, ...	0.673	0.206	0.413	0.503
CN 15: Base metals and articles of base metal	0.469	0.476	0.309	0.644
CN 16: Machinery and mechanical appliances; electrical equipment; parts thereof; ...	0.614	0.273	0.327	0.589
CN 17: Vehicles, aircraft, vessels and associated transport equipment	0.802	0.057	0.449	0.377
CN 18: Optical, photographic, measuring, checking, precision, medical instruments	0.625	0.268	0.160	0.761
CN 19: Arms and ammunition; parts and accessories thereof	0.619	0.038	0.698	0.249
CN 20: Miscellaneous manufactured articles	0.527	0.230	0.280	0.655
CN 21: Works of art, collectors' pieces, and antiques	0.298	0.031	0.598	0.025
CN 22: Other products	0.000	0.055	0.481	0.190

Table 16 Concentration in export market by type of firms

	All	Matched	Manufacturing	Wholesale
Top 1	0.562	0.515	0.426	0.470
Top 5	0.808	0.786	0.710	0.759
Top 10	0.900	0.886	0.834	0.866
Top 25	0.978	0.974	0.954	0.966
Top 50	0.998	0.997	0.995	0.995
Top 100	1.000	1.000	1.000	1.000

Table reports the share of export values accounted by participating firms. "All", "Matched" and "Manufacturing" and "Wholesale" refer to all export, export data that are matched with industry affiliations, export accounted by manufacturing firms and export accounted by wholesale trade firms respectively. *Source*: Statistics Denmark, International Trade, Export Data Set, Year 2002

Table 17 Concentration in export market by type of goods

	All	Consumption	Intermediate
Top 1	0.562	0.625	0.528
Top 5	0.808	0.864	0.805
Top 10	0.900	0.938	0.899
Top 25	0.978	0.990	0.979
Top 50	0.998	0.999	0.997
Top 100	1.000	1.000	1.000

Table reports the share of export values accounted by participating firms. "All", "Consumption" and "Intermediate" refer to all export, consumption goods export and intermediated goods export respectively. Consumption and intermediate goods definitions follow BEC. Intermediate goods do not include fuels and lubricants. *Source*: Statistics Denmark, International Trade, Export Data Set, Year 2002

Table 18 Concentration in export market by type of firms and type of goods

	WHS		MAN	
	Consumption	Intermediate	Consumption	Intermediate
Top 1	0.453	0.556	0.456	0.352
Top 5	0.779	0.805	0.781	0.700
Top 10	0.889	0.894	0.894	0.826
Top 25	0.979	0.975	0.982	0.954
Top 50	0.997	0.996	0.998	0.995
Top 100	1.000	1.000	1.000	1.000

Table reports the share of export values accounted by participating firms by type of goods and type of firms. *Source*: Denmark Statistics, International Trade, Export Data Set, Year 2002

Appendix 2: Data

Foreign Trade Data

The foreign trade data sets are compiled from the Danish Customs records. Each shipment record includes the date of the shipment, the value of shipment, the product code (CN-8 digit), and the name of the product, weight of the shipment, type of the weight and sometimes quantity information as well as the unique firm identifier. Statistics Denmark aggregated this data into annual shipments for each product (CN-8 digit), country and firm triplet. As provided by Statistics Denmark, the international transaction data set covers the universe of Danish firms' transactions for the period 1993–2007. However, only product shipments of 10,000 kr (approx. 1800 us $) or above are included in the data set for the transactions with the EU countries.

Business Statistics Data

Business statistics data are compiled from survey results of firms that take part in a yearly financial survey as well as from tax reports, vat reports, and annual reports from incorporated companies. The general business statistics include only firms that employ at least a 0.5 FTE (full-time equivalent number of employees) and/or have had an estimated earnings of a certain size. Earning sizes are estimated differently for different industries. In the wholesale trade sectors, the limit of earnings is typically over 500,000 Danish Kroner, while in the manufacturing industry, it ranges between 150,000 and 200,000 Danish Kroner. Some of the data for very small firms may be subject to imputation. This data set is available starting from 1995, but only manufacturing, construction and retail sectors are included until 1998. In 1998, the wholesale trade sector is included and starting from 1999 it covers almost all sectors including mining, and all business service sectors.[19] This data set is supplemented with the labor surveys (IDA) that provide information on wages, education and occupation characteristics for each individual in the labor force. In the labor (IDA) data set, for each employed person there is a unique firm identifier provided for the employer. Using this firm identifier, extracted information from IDA is merged with the Firm Accounting Data Set for each year. Only a couple of observations in firm accounting data were left unmatched from this matching.[20]

[19]Starting from 1999, the data set includes hospitality, transportation, telecommunication, real estate, rental services, information technology services, research and development services, and other consultancy and business services. It does not include agriculture, financial sector, public, education and medical service sectors.

[20]For the details of labor data set as well as other data sets used in this study see Utar (2014).

Intermediaries are defined as firms with their main economic activity in 2-digit Danish Industrial Classification 51 (wholesale except of motor vehicles) as well as 6-digit industry classifications equal to 501010, 501020, 501030, 503010, 503020, and 504000 which are sale of motor vehicles, parts and accessories.

Matching Foreign Trade Data with Firm-Level Data Sets

Foreign trade as compiled from the custom records contain firm id's but not a main business/industry affiliation, so it is not possible to identify the type of firms whether wholesale trade, retailer, manufacturer or service etc. from the foreign trade data alone. The analysis in this paper is carried out by matching the foreign trade data with the business statistics as well as other available data sets from Statistics Denmark such as tax data, and industry sales data. Between 1993 and 2007, most of the foreign data in the import side (94 % of firms) can be matched, less so in export (89 % of firms). A significant part of the transactions cannot be matched in the export side, probably due to reporting errors. Nevertheless during 1998 and 2006 which is the sample period used in the empirical analysis, 91 % of the exporting firms in custom data were matched with their corresponding industry affiliations.

Product Detail

Products description is based on the Combined Nomenclature (CN) 8 digit categories. The first 6 digits of the CN corresponds to the HS-6 digit classification. For example, 852812 product code refers to color television receivers with built-in picture tubes. In the CN-8 classification there are 19 different kinds of color television receivers depending on different characteristics such as display width, diagonal screen size, and lines of resolution.

Broad Product Classification

Product classification of the products as consumption, intermediate, or industrial good is based on BEC Rev. 3. Consumption goods are defined as (BEC=112, 122, 522, >=600). The rest are defined as industrial goods. Intermediate goods definition does not include fuels and lubricants and is defined as (BEC=111, 121, 210, 220, 420, 530).

CN Chapters are used as broad product classifications and they are listed in Table 15. CN codes are matched with the corresponding manufacturing industries using PRODCOM. Prodcom provides statistics on the production of manufactured goods. Prodcom uses the product codes specified on the Prodcom List, which contains about 4500 different types of manufactured products. Products are identified

by an 8-digit code: the first four digits are the classification of the producing enterprise given by the Statistical Classification of Economic Activities in the European Community (NACE). Most product codes correspond to one or more Combined Nomenclature (CN) codes, but some (mostly industrial services) do not. The matching between CN and PRODCOM are provided by EUROSTAT RAMON. The matches are executed for every year separately.

Rauch (1999) classification is used to classify products as homogenous, reference and differentiated goods. Classifications in Rauch (1999) are based on SITC codes. Correspondence tables between CN 8-digit and SITC 4-digit (provided by EUROSTAT RAMON) are used to link the classification with the Danish data.

High tech goods definitions follow OECD nomenclature (Loschky 2008). High-technology classification is based on both direct and indirect R&D intensities in relation to the production output or to the valued added. The indirect R&D intensity is defined as the R&D expenditures embodied in the intermediate products used in the production in another economic sector. See Loschky (2008) for more details.

References

Ahn, J. B., Khandelwal, A. K., & Wei, S.-J. (2011). The role of intermediaries in facilitating trade. *Journal of International Economics, 84*(1), 73–85.

Akerman, A. (2010). *A Theory on the Role of Wholesalers in International Trade Based on Economies of Scope*. Research Papers in Economics 2010:1, Stockholm University, Department of Economics.

Bernard, A. B., Grazzi, M., & Tomasi, C. (2011). *Intermediaries in International Trade: Direct Versus Indirect Modes of Export*. NBER Working Paper No. 17711.

Bernard, A. B., Jensen, B., Redding, S. J., & Schott, P. K. (2007). Firms in international trade. *Journal of Economic Perspectives, 21*(3), 105–130.

Bernard, A. B., Jensen, B., Redding, S. J., & Schott, P. K. (2010). Wholesalers and retailers in U.S. trade. *American Economic Review Papers and Proceedings, 100*, 408–413.

Biglaiser, G. (1993). Middlemen as experts. *Rand Journal of Economics, 24*(2, Summer 1993), 212–223.

Biglaiser, G., & Friedman, J. (1994). Middlemen as guarantors of quality. *International Journal of Industrial Organization, 12*, 509–531.

Blum, B., Claro, S., & Horstmann, I. (2009). *Intermediation and the Nature of Trade Costs: Theory and Evidence*. Working Paper, University of Toronto.

Dent, J. (2008). *Distribution channels: Understanding and managing the channels to market*. London: Kogan Page Limited.

Felbermayr, G., & Jung, B. (2011). Trade intermediation and the organization of exporters. *Review of International Economics, 19*(4), 634–648.

Helpman, E., Melitz, M., & Yeaple, S. (2014). Export versus FDI with heterogeneous firms. *American Economic Review, 94*(1), 300–316.

Loschky, A. (2008). *Reviewing the Nomenclature for High-Technology Trade – The Sectoral Approach*. Paper presented at the 1st Meeting of the OECD Working Party on International Trade in Goods and Trade in Services Statistics (WPTGS), Paris, September 22–24, 2008.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*, 1695–1725.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics, 11*, 619–632.

Pedersen, N. (2009). *Essays in International Trade*. Ph.D. dissertation, Northwestern University, Evanston, IL.

Rauch, J. E. (1999). Network versus markets in international trade. *Journal of International Economics, 48*, 7–35.

Rubinstein, A., & Wolinsky, A. (1987). Middlemen. *The Quarterly Journal of Economics, 102*(3), 581–593.

Tang, H., & Zhang, Y. (2012). *Quality Differentiation and Trade Intermediation*. Working Paper, Tufts University.

Utar, H. (2014). When the floodgates open: Northern firms' response to removal of trade quotas on Chinese goods. *American Economic Journal: Applied Economics, 6*(4), 226–250.

Part III
Financial Markets and Aggregate Fluctuations

Cyclical or Structural? Evidence on the Sources of U.S. Unemployment

Jinzhu Chen, Prakash Kannan, Prakash Loungani, and Bharat Trehan

Abstract We provide evidence on the relative importance of cyclical and structural factors in explaining unemployment, including the sharp rise in U.S. long-term unemployment during the Great Recession of 2007–09. About 75 % of the forecast error variance of unemployment is accounted for by cyclical factors—real GDP changes ("Okun's Law") and monetary and fiscal policies. Structural factors, which we measure using the dispersion of industry-level stock returns, account for the remaining 25 %. For long-term unemployment the split between cyclical and structural factors is closer to 60–40, including during the Great Recession. Examination of the industry-level stock returns suggests that adverse shocks to the construction sector and, to a lesser extent, the finance sector were responsible for the increase in structural unemployment. The Great Recession appears similar to the recession of 1973–75, as sectoral shocks played a large role at that time as well.

1 Introduction

Are persistent increases in unemployment cyclical or structural? The question is timely and timeless. It is timely because the sharp run-up in U.S. unemployment rates since 2007 has triggered a debate on the contribution of structural factors. Krugman (2010) states that the present "high unemployment in America is the result of inadequate demand—full stop", whereas Kocherlakota (2010) asserts that

J. Chen
Industrial and Commercial Bank of China, Beijing, China
e-mail: chen25@fas.harvard.edu

P. Kannan
Government of Singapore Investment Corporation (GIC), Singapore, Singapore
e-mail: prakashkannan@gic.com.sg

P. Loungani (✉)
International Monetary Fund, Washington, DC, USA
e-mail: ploungani@imf.org

B. Trehan
Federal Reserve Bank of San Francisco, San Francisco, CA, USA
e-mail: Bharat.Trehan@sf.frb.org

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_10

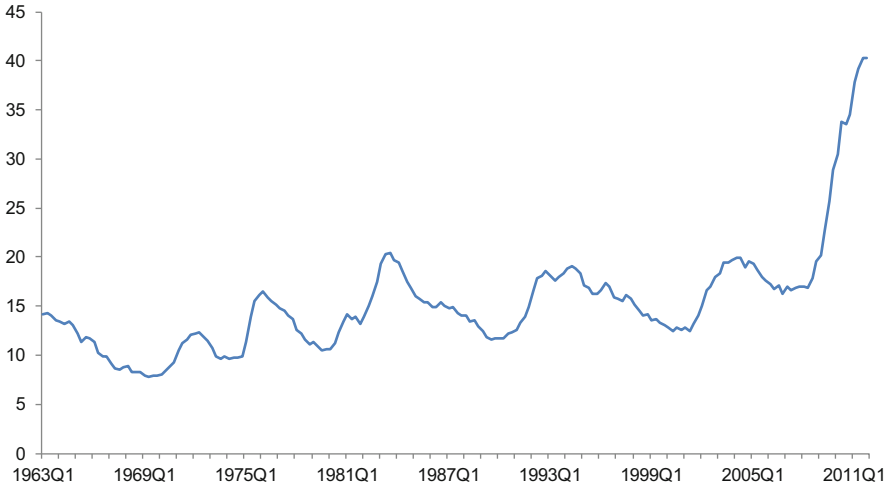

Fig. 1 Average duration of unemployment (weeks)

"firms have jobs, but can't find appropriate workers. The workers want to work, but can't find appropriate jobs." It is timeless because nearly every persistent increase in unemployment over the past 100 years has been marked by a debate of this kind. For instance, the persistent unemployment of the Great Depression was attributed by some to "a great shortage of labor of certain kinds" (Clague 1935). Likewise, the causes of the high unemployment in Great Britain during the interwar years were a matter of intense debate then and to this day (see Benjamin and Kochin 1979; Brainard 1992; Nason and Vahey 2006).[1]

During the Great Recession, the sharp increase in the U.S. unemployment rate from 4.4 % in May 2007 to 10.1 % in October 2009 was accompanied by a striking increase in the duration of unemployment. As shown in Fig. 1, while the average duration of unemployment has been inching upwards for a while, it rose sharply in the recession and continued to increase well after the peak in the unemployment rate. Recent readings, which show average unemployment spells in excess of 40 weeks, are about 20 weeks above the previous highest duration seen in data going back to the early 1960s.[2] Figure 2 shows the breakdown of unemployment by duration of unemployment spells, which underlie the changes in average duration. While

[1] Similarly, the jobless recovery following the 2001 U.S. recession led some observers, notably Groshen and Potter (2003), to assign a significant role to structural factors, while others took a more skeptical view (e.g. Aaronson et al. 2004).

[2] Effective January 2011, the Current Population Survey (CPS) was modified to allow respondents to report durations of unemployment of up to 5 years. Prior to that time, the CPS accepted unemployment durations of up to 2 years; any response of unemployment duration greater than this was entered as 2 years. For the first 6 months of 2011, the new measure of mean duration exceeded the old by 2.3 weeks on average.

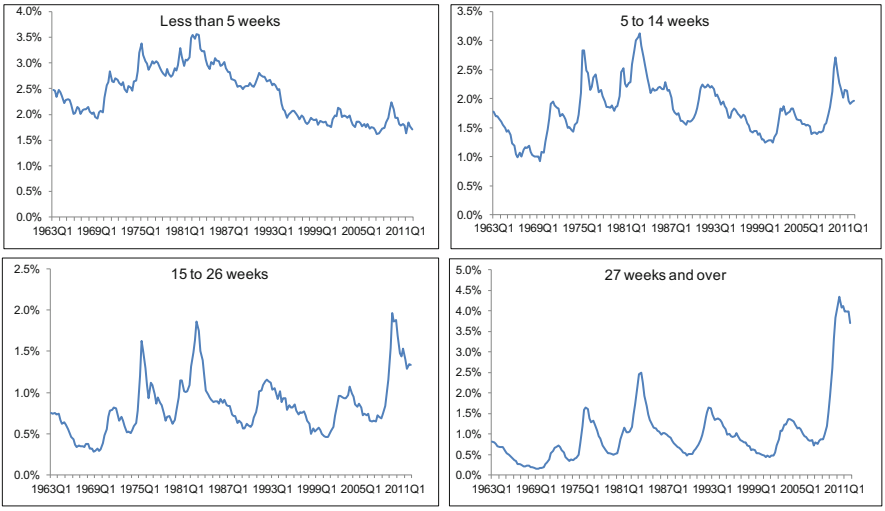

Fig. 2 Duration of unemployment (percent of labor force)

short-term spells (less than 5 weeks) also showed an uptick, it is the increase in medium-term and long-term spells (greater than 27 weeks) that is particularly alarming.

The severity and persistence of output declines during the Great Recession have likely had a significant effect on unemployment (Elsby et al. 2010). However, a number of studies suggest that structural factors have played a part as well. Kirkegaard (2009) extends the analysis of Groshen and Potter (2003) to study structural and cyclical employment trends during the Great Recession. He finds that there has been a sharp increase in "the relative employment weight of industries undergoing structural change in the current cycle", which he concludes "can be expected to increase the necessity for unemployed Americans to take new jobs in industries different from the ones in which they were previously employed."

Other studies have looked for evidence of structural change in terms of mismatch in the labor market, and tend to find that while mismatch is not the dominant factor, it has played a non-negligible role in the increase in unemployment over the recession. Sahin et al. (2011) define mismatch as the distance between the observed allocation of unemployment across sectors and the allocation that would prevail under costless mobility. They find that between 0.8 and 1.4 percentage points of the increase in U.S. unemployment during the Great Recession can be attributed to mismatch. This corresponds to between 20 % and 25 % of the observed increase in unemployment. They also find industrial and occupational mismatches, rather than geographic mismatch, are the sources of the increased unemployment.

Barnichon and Figura (2010) find that until 2006, most of the changes in matching efficiency could be explained by changes in the composition of the unemployed (for instance, the relative prevalence of workers on temporary vs. permanent layoffs). Since 2006, however, composition has played a much diminished role

relative to the role of "dispersion in labor market conditions, the fact that tight labor markets coexist with slack ones." They estimate that in the 2008–2009 recession, the decline in aggregate matching efficiency added 1½ percentage points to the unemployment rate.

Estevao and Tsounta (2011) find that "increases in skill mismatches in states with worse housing market conditions ... are associated with even higher unemployment rates, after controlling for all cyclical factors." They suggest that this could be because "bad local housing conditions may slow the exodus of jobless individuals from a depressed area, thus raising equilibrium unemployment rates." They estimate that the combined impact of skill mismatches and higher foreclosure rates might have raised the natural rate of unemployment by about 1½ percentage points since 2007. Lazear (2012) constructs a mismatch index based on the ratio of vacancies to unemployment. He concludes that while mismatch did rise substantially during the recent recession, it is no longer an issue for the labor market.

This paper provides further evidence on the role of structural factors in accounting for the recent rise in unemployment, especially the long-term component, using a different methodology. We use the stock market valuation of the firms in an industry to construct a measure of the shocks hitting that industry. More specifically, our measure of the intensity of structural shocks implements a conjecture by Black (1987) that periods of greater cross-industry dispersion in stock returns should be followed by increases in unemployment. The cross-industry dispersion of stock returns provides an "early signal of shocks that affect sectors differently, and puts more weight on shocks that investors expect to be permanent" (Black 1995). This latter point is important because it is presumably permanent shocks that motivate reallocation of labor across industries.

The paper makes three contributions relative to previous research on the topic (Loungani et al. 1990; Brainard and Cutler 1993; Loungani and Trehan 1997). First, it considerably extends the sample length to incorporate the last 20 years or so, an eventful period in US macroeconomic history. The extended sample confirms the main messages in earlier research—namely, that unemployment rates rise in the wake of persistent sectoral shocks, as proxied by the cross-industry dispersion of stock returns, and that these shocks account for a sizable fraction of unemployment rate fluctuations. Second, the share of unemployment fluctuations attributed to sectoral shocks increases substantially in moving from shorter duration to longer duration unemployment. In addition, long-duration unemployment exhibits very different response dynamics to innovations in the monetary policy and dispersion variables. These new results strengthen the case for a sectoral shifts interpretation of the empirical relationship between unemployment rates and the cross-industry dispersion of stock returns. Third, the paper provides evidence that the unusually large cross-industry dispersion of stock returns associated with the Great Recession helps explain why long-term unemployment has been such a prominent feature of its aftermath.

While not a direct test, our evidence provides support for theoretical work that assigns an important role to structural shocks. Phelan and Trejos (2000) show that permanent changes in sectoral composition can lead to aggregate downturns in a calibrated job creation/job destruction model of the U.S. labor market.

Section 2 deals with matters of measurement. The next two sections present the empirical evidence. In Sect. 3, we use bivariate regressions as in Romer and Romer (2004) to see how aggregate factors affect overall unemployment and long-term unemployment. Complementing this evidence, results from a VAR model are presented in Sect. 4. Section 5 concludes.

2 Measuring Sectoral Shocks and the Costs of Reallocation

Conditions in any given sector often evolve quite differently from the economy as a whole; witness the relative performance of the energy sector in the 1970s. Such differences in relative performance will typically be accompanied by a need to reallocate resources across sectors. This is likely to be an expensive, time consuming process. For instance, Lilien (1982) argues that labor reallocation is likely to take time because workers may have strong attachments arising from industry-specific skills and wage premiums associated with seniority. Search models suggest other reasons. In Phelan and Trejos (2000), reallocation is slow because the cost of creating a job is higher in the expanding sector.

Reallocation is likely to be costly in other ways as well. Lee and Wolpin (2006) estimate that the direct cost of an inter-sectoral move is 50–75 % of an individual's annual earnings. In a study of the Aerospace sector, Ramey and Shapiro (2001) find that firms recover just 28 % of the replacement cost of capital sold during a sectoral downturn. Here, we follow Lilien (1982), as well as many others, who measure the costs of reallocation in terms of what happens in the labor market. More specifically, we estimate the effect that sectoral shocks have on the unemployment rate and—importantly—on the duration of unemployment.

Measuring sectoral shocks is not a straightforward exercise either. As pointed out by Barro (1986) and Davis (1985), sectoral shocks are typically unobservable disturbances to technology and preferences, and are unlikely to come repeatedly from the same source. One solution is to turn to the stock market, in the expectation that any kind of shock specific to a particular industry is likely to show up in the relative performance of that industry's stock price. Importantly, permanent shocks to the industry are expected to have a larger impact on industry stock returns than temporary shocks.

Figure 3a and b provide some informal evidence in support. The top panel in Fig. 3a shows excess returns to homebuilders since the beginning of 2000. Excess returns turned negative in late 2005 and by early 2006 the (3 quarter average of the) excess return had fallen below −10 % per quarter. The middle panel shows the average duration of unemployment in the construction industry over the same period.[3] This is the average number of weeks for which workers whose last job

[3]Homebuilders are a subset of the construction industry, but the unemployment data are only available at relatively high levels of aggregation. We are currently engaged in constructing matching unemployment and stock market series.

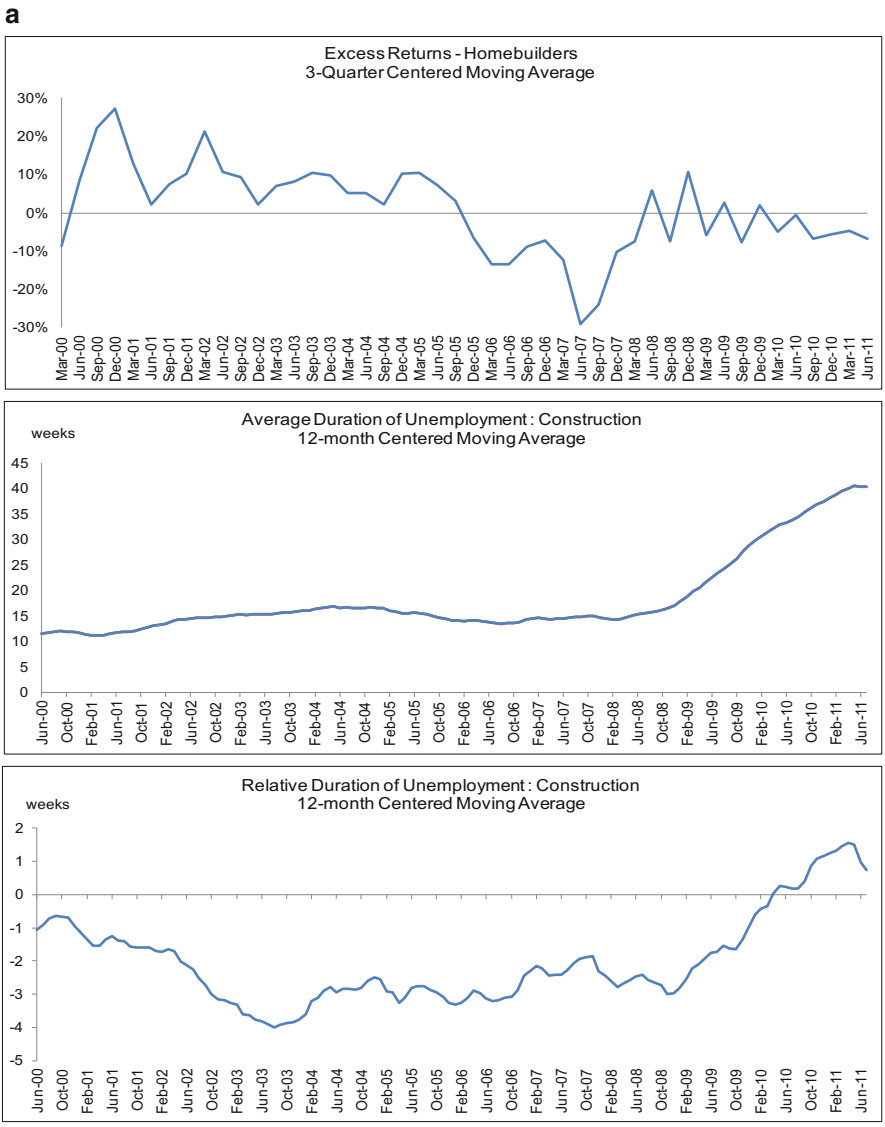

Fig. 3 (**a**) Excess returns and duration. (**b**) Excess returns and duration

was in the construction industry have been unemployed. Notice that duration in the construction sector does not begin to rise till early in 2008. The bottom panel shows duration in the construction industry relative to the aggregate economy. Here, again, it is the pronounced rise in relative duration since early 2008 that is noticeable.

The top panel of Fig. 3b shows excess returns in the commercial banking sector, while the bottom two panels show the average duration of unemployment in the

b

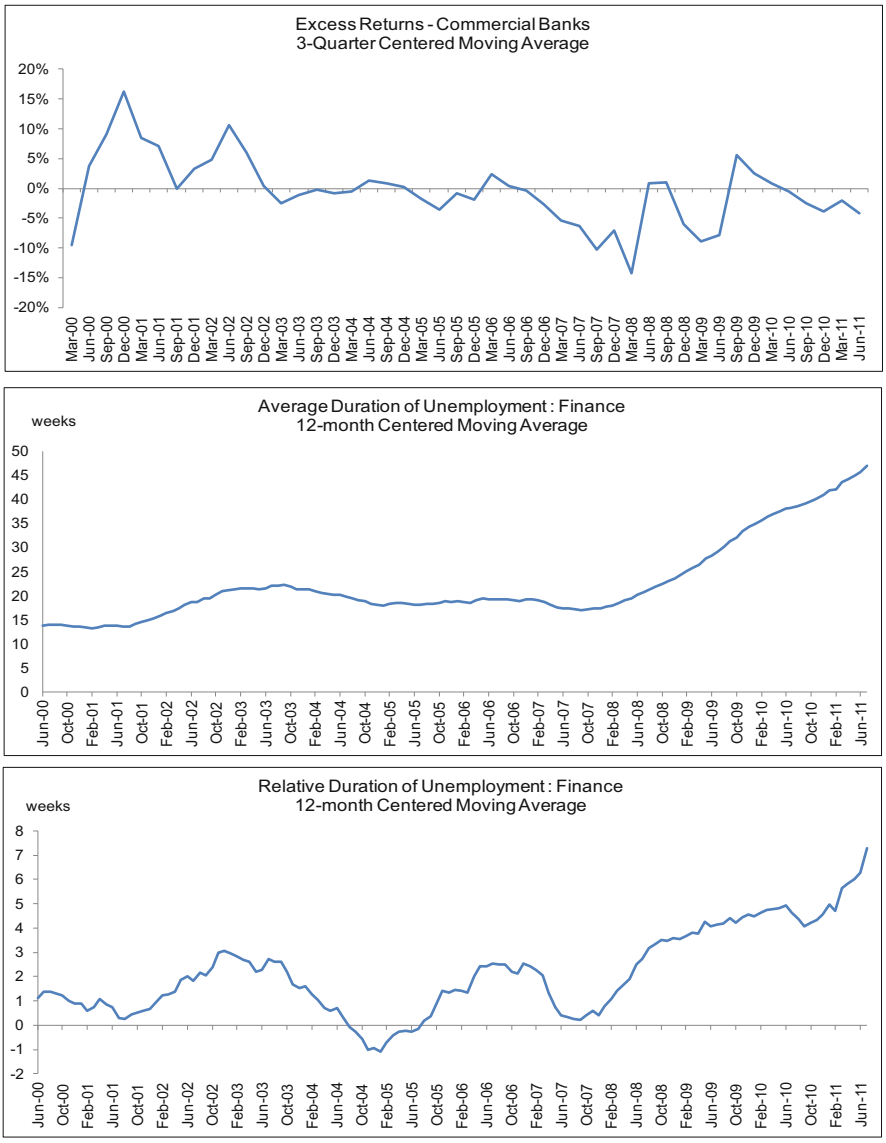

Fig. 3 (continued)

finance industry.[4] The decline in excess returns here happens somewhat later than that for homebuilders, though by the end of 2006 excess returns are below −5 %. Both average and relative duration move up noticeably in 2008 and beyond.

[4]Once again, commercial banks are a subset of the finance industry.

These figures provide some informal evidence that stock prices can be used as early signals about sectoral shocks. But we still have to determine which sectors have been hit by sectoral shocks at any point in time. Instead of continuing with an industry by industry examination of stock market returns and labor market developments, we will employ a dispersion index, following previous work by Loungani et al. (1990), Brainard and Cutler (1993), and others. The hypothesis is that the dispersion of stock returns across industries can be used as a proxy for shocks to the desired allocation of labor, i.e., as a measure of sectoral shifts. For instance, the arrival of negative news regarding the relative profitability of a particular industry is likely to be followed by an increase in stock price dispersion. In the long run, this news is likely to shift the economy's output mix away from the affected industry. This will necessitate a reallocation of resources, and the unemployment rate will rise as part of this process of reallocation of labor across sectors. Thus, an increase in stock price dispersion will be followed by an increase in the unemployment rate.

The stock market dispersion index follows Lilien (1982), who constructed a cross-industry employment dispersion index to proxy for the intersectoral flow of labor in response to allocative shocks. Subsequent researchers, most notably Abraham and Katz (1986), argued that employment dispersion may simply be reflecting the well-known fact that the business cycle has non-neutral effects across industries. An advantage of the stock price dispersion measure relative to Lilien's measure is that unlike employment changes, stock prices respond more strongly to disturbances that are perceived to be permanent rather than temporary. The industry stock price represents the present value of expected profits over time. If the shocks are purely temporary, the innovations will have little impact on the present value of expected profits and, hence, will have little impact on industries' stock prices. But persistent shocks will have a significant impact on expected future profits and will lead to large changes in industries' stock prices. Thus, a dispersion index constructed from industries' stock prices automatically assigns greater weight to permanent structural changes than to temporary cyclical shocks, and so will be less likely to reflect aggregate demand disturbances than a measure based on employment. Furthermore, it is these persistent shocks that are likely to cause productive resources, such as capital and labor, to move across industries.

For this paper, we update the stock market index used in earlier studies. The basic data consist of Standard and Poor's indexes of industry stock prices, providing comprehensive coverage of manufacturing as well as nonmanufacturing sectors of the economy. The sectoral shifts index is defined as

$$Dispersion_t = \left[\sum_{i=1}^{n} W_i (R_{it} - R_{mt})^2 \right]^{1/2}, \tag{1}$$

where R_{it} is the growth rate of industry i's stock price index, R_{mt} is the growth rate of the S&P500 (a composite index), and W_i is a weight based on the industry's

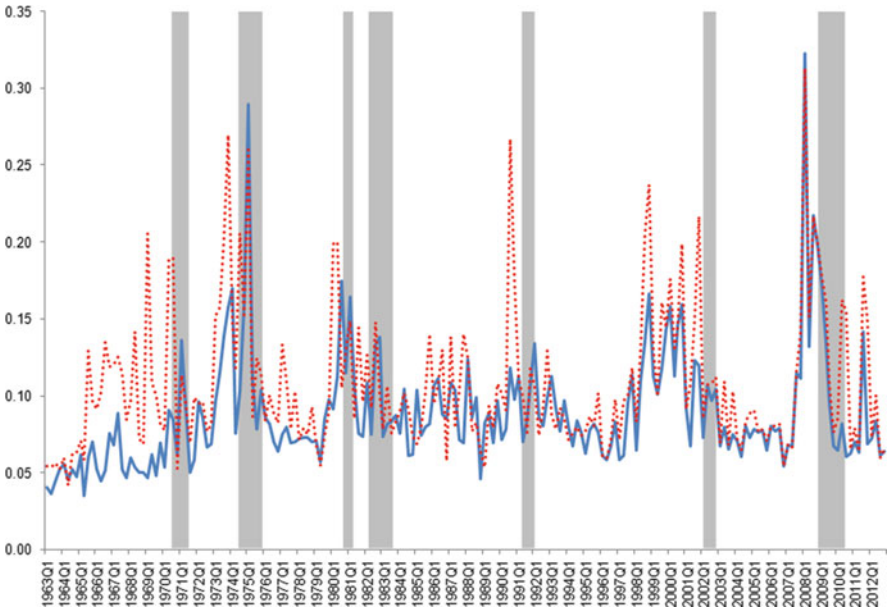

Fig. 4 Stock market returns dispersion index. *Solid line*: Return dispersion constructed by Brainard and Cutler (1993). *Dotted line*: Return dispersion constructed by Loungani et al. (1990)

share in total employment. Hence, the sectoral shifts index can be interpreted as the weighted standard deviation of industry stock returns.

Brainard and Cutler (1993) noted that some industry stock returns may be more cyclically sensitive, so aggregate shocks could increase the dispersion of returns. They introduced a modified measure that attempts to eliminate these cyclical effects by first regressing industry returns on market returns,

$$R_{it} = \beta_{0i} + \beta_{1i} R_{mt} + \varepsilon_{it} \tag{2}$$

Then the dispersion index is constructed from the 'excess returns,' that is, the unexplained portions (intercepts and residuals) from the regressions (2).

Figure 4 shows the behavior of the two measures of stock price dispersion index for the U.S. over the period 1963–2012. As is evident, the two measures are highly correlated, with a correlation coefficient of 0.82. The results in the remained of the paper are unaffected by which of the two indices is used. Since the Brainard and Cutler (1993) measure is preferable in theory, the empirical results shown are based on that measure.

Recession periods are typically associated with an increase in the dispersion index. However, the magnitude of the increase in the index during a particular recession is not necessarily reflective of the depth of the recession. The index increases by much more during the 1974–75 recession than it does during the

1982 recession, even though the latter recession was more severe in terms of output loss and the increase in the unemployment rate. This evidence suggests that changes in the relative profitability of industries during a contraction are not closely correlated with the size of the contraction, which is consistent with the interpretation that different recessions are marked by different mixes of sectoral and aggregate shocks. Similarly, the dispersion index has also recorded increases during expansionary periods. The dot-com boom experienced in the late 1990s provides a clear example, as stock prices in the information, technology, and communication sectors experienced much more rapid gains than the market average.

3 Alternative Explanations for Changes in Unemployment by Duration

The behavior of the unemployment rate can potentially be influenced by a variety of factors, both cyclical as well as structural, not all of which can be simultaneously included in a moderately sized VAR. Accordingly, in this section we compare the effects of the dispersion index on unemployment rates (of different durations) with the effects of other key macroeconomic variables, using a single-equation framework similar to Romer and Romer (2004) and Cerra and Saxena (2008). Specifically, we regress changes in the unemployment rate (ΔU) on its own lags as well as lagged values of the various "shocks". The lagged values allow for delays in the impact of the shocks on unemployment.[5] The estimated equation is

$$\Delta U_t = \alpha + \sum_{i=1}^{4} \Delta U_{t-i} + \sum_{i=1}^{4} Shock_{t-i} + \varepsilon_t \tag{3}$$

Four candidate shocks are evaluated using this framework. The first two are monetary and fiscal (tax rate) policy shocks, as identified by Romer and Romer (2004). Both these shocks are constructed so as to be exogenous to changes in output through the use of narrative records of the Federal Reserve Open Market Committee meetings, presidential speeches and Congressional reports. The third shock examined is related to oil prices, and is simply measured as the percentage change in the real price of oil. Finally, we look at the effect of changes in the stock price dispersion index. For each of these shocks, standard errors for the impulse response functions are estimated using a bootstrap method.[6]

The impact of a one standard-deviation change in the various shocks on the level of the unemployment rate is shown in Fig. 5. The unemployment rate increases

[5] A lag length of 4 quarters was found to be optimal.

[6] Specifically, 500 pseudo-coefficients are drawn from a multivariate normal distribution based on the estimates of the mean and variance-covariance matrix of the regression coefficient vector.

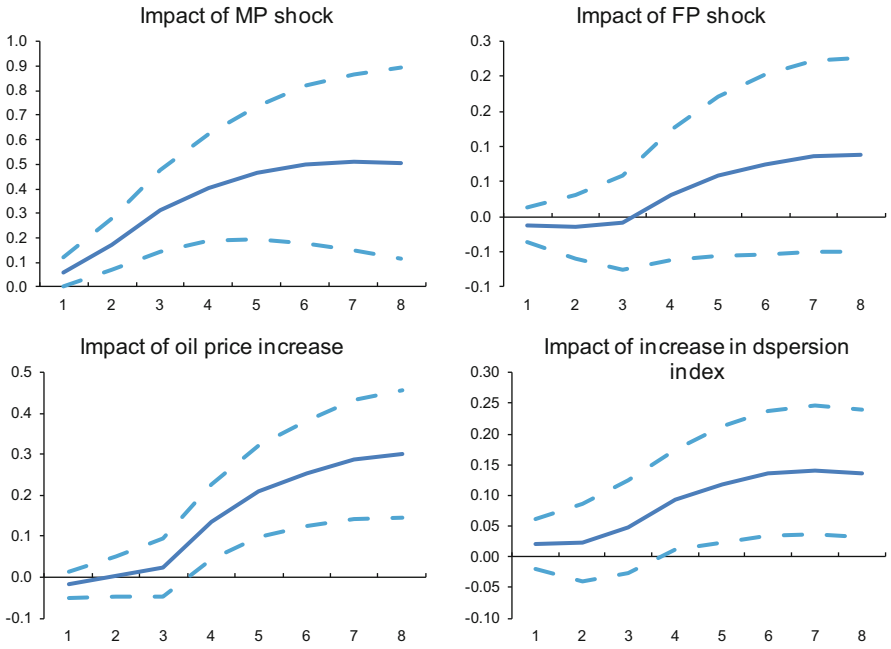

Fig. 5 Impulse response figures for unemployment (Univariate model)

following each shock, though the magnitude and significance of the responses vary across shocks. The response of unemployment to a monetary policy tightening is particularly large and persistent. A one standard deviation shock to monetary policy results in an increase in the unemployment rate of more than 0.5 % after 2 years. Shocks to fiscal policy, on the other hand, are small and insignificant. In contrast, increases in the real price of oil are associated with increases in the unemployment rate, with the peak impact occurring after 2 years. Finally, increases in the dispersion index are associated with a positive and significant change in the unemployment rate. A 1 standard deviation increase in the index results in an increase in the unemployment rate of about 0.2 % after a year and a half.

The long term unemployment rate (where duration exceeds 26 weeks) responds very differently to these shocks; see Fig. 6. The typical response of the long-term unemployment rate to either a monetary policy shock, a fiscal policy shock or an increase in the real price of oil is small in magnitude and of the *opposite* sign to the response of the unemployment rate observed in Fig. 5. These responses are counter intuitive, but not significant at the 90 % confidence interval. The response of the long-term unemployment rate to an increase in the dispersion index, however, is positive (as it is in Fig. 5) and statistically significant. A one standard-deviation increase in the index results in a gradual increase in the long-term unemployment rate, peaking at around 0.1 percentage points after 2 years. Given that the average long-term unemployment rate in the U.S. for the last 40 years is about 0.9 %, the

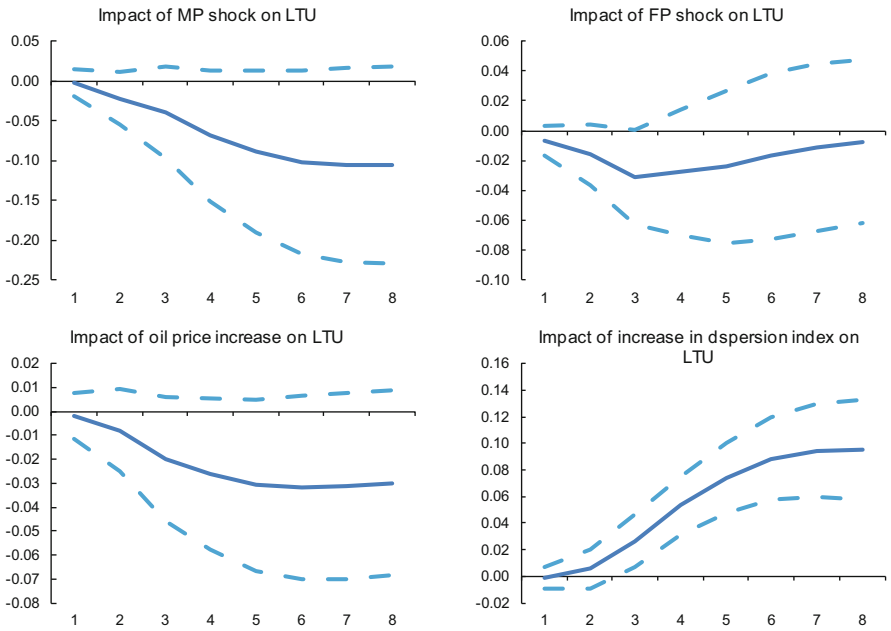

Fig. 6 Impulse response figures for long-term unemployment (Univariate model)

impact of changes in the dispersion index on long-term unemployment is relatively substantial.

The findings above highlight the importance of the dispersion index, relative to other standard macroeconomic factors, in explaining variations in unemployment—particularly, long-term unemployment. Given our results here, we will not carry over the fiscal policy and oil price variables into the next section. Before going further, it is worth pointing out that a regression of the dispersion index on lagged values of the monetary and fiscal policy shocks revealed that the index is not systematically correlated with these macro shocks.

4 Multivariate Specifications

In this section, we present the results from a VAR estimated on quarterly data from 1963:Q1 to 2012:Q4. The baseline model contains six variables, including the stock market dispersion index and the unemployment rate U_t. The other variables are real GDP growth, inflation, the federal funds rate, and the growth rate of the S&P500 index, R_{mt}. The inclusion of real GDP controls for the stage of the business cycle;

it also means that our model allows for a version of "Okun's Law." The variable measuring returns on the S&P500 index is included to rule out the possibility that the dispersion index explains unemployment because it is mimicking the behavior of the aggregate stock market. Finally, following Bernanke and Blinder (1992), the fed funds rate is included as a measure of monetary policy. The system is identified following the standard recursive ordering procedure. To avoid exaggerating the role of the dispersion index, we place it last in the estimation ordering. The other variables in the system are ordered as follows: GDP growth is placed first, followed by on the growth rate of the S&P 500, the unemployment rate, inflation and the fed funds rate. The lag length is set at 4 quarters.

4.1 The Effects of Sectoral Shocks

Figure 7 shows how unemployment responds to different shocks to the system, along with the associated 90 % confidence intervals.[7] The unemployment rate declines following a shock to output growth, with a one standard deviation increase in the output growth rate leading to a nearly 0.3 percentage point decrease in the unemployment rate after 1 year. Innovations to the fed funds rate, meanwhile, result in higher unemployment. Focusing on the response of unemployment to innovations in the dispersion index, we see that the unemployment rate gradually increases, with the response reaching a peak of above 0.25 percentage points 2 years after the shock. The impact of these identified shocks to the dispersion index—purged of the aggregate influence of GDP, total market return, inflation and monetary policy—on unemployment is higher than what we obtained from the regressions shown in the last section.[8]

The long-term unemployment rate (Fig. 8a) shows a hump shaped response to innovations in the dispersion index, just as the overall unemployment rate (Fig. 7). However, long-term unemployment reacts more gradually, reaching its peak closer to 3 years after the shock. The magnitude of the peak impact is again higher than what was found in the previous section. Long-term unemployment declines in response to output growth innovations, though just as with regards to dispersion shocks, the response is more delayed relative to the response of the overall unemployment rate. Long-term unemployment eventually increases following a shock to monetary policy, although the magnitude of the response is insignificant at the 90-percent confidence level. In contrast, the short-term unemployment rate (Fig. 8b) shows an insignificant response to innovations in the dispersion index,

[7] Standard errors are computed using the statistics based on the asymptotic distribution.

[8] We also estimate the VAR using levels of variables with quadratic trends instead of the log of difference of the variables. The impact of dispersion on unemployment is reduced somewhat but it remain statistically significant and the variance decomposition is not much changed.

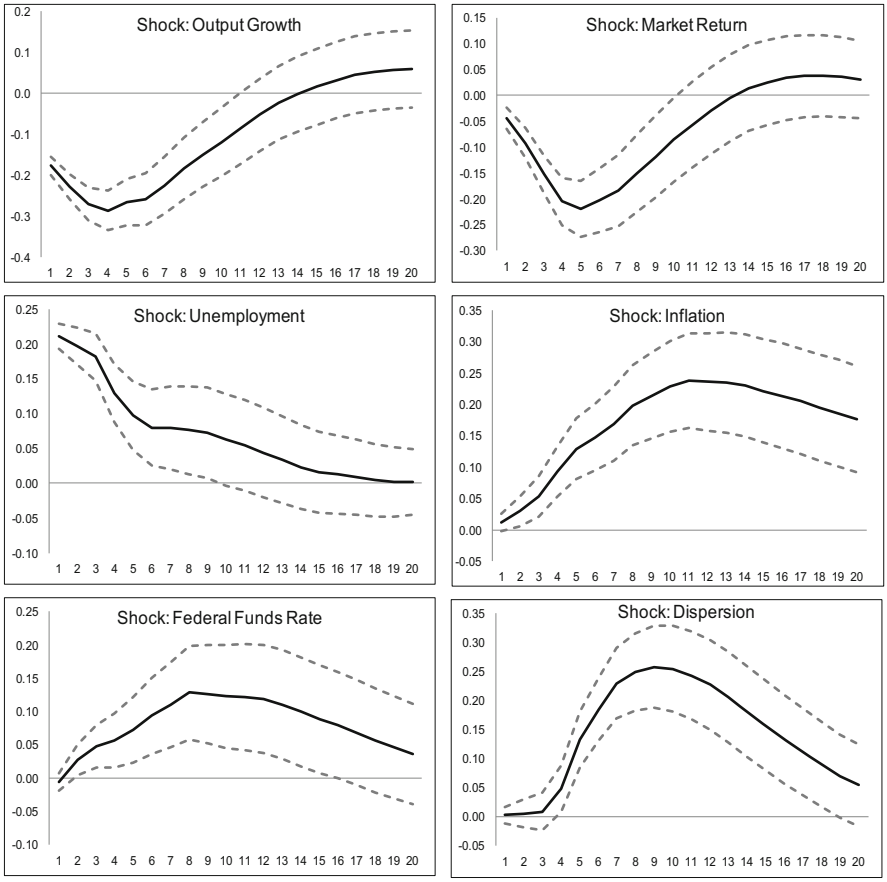

Fig. 7 Impulse response figures for long-term unemployment (VAR)

which supports the previous argument that the dispersion index mostly affects longer-duration unemployment.

A decomposition of the variance of unemployment forecast errors provides further evidence as to the importance of the dispersion index in explaining unemployment fluctuations. Tables 1 and 2 show the proportion of the forecast-error variance of overall unemployment and long-term unemployment, respectively, that is explained by the various shocks, given our identification scheme. At the 5-year horizon, close to a fourth of the variance of unemployment forecast errors is explained by innovations to the dispersion index. The proportion is much higher when we consider variations in long-term unemployment. Looking once again at the 5-year horizon, innovations to the dispersion index account for close to 40 % of the overall variance, making it more important than any other variable in the VAR, including the unemployment rate itself.

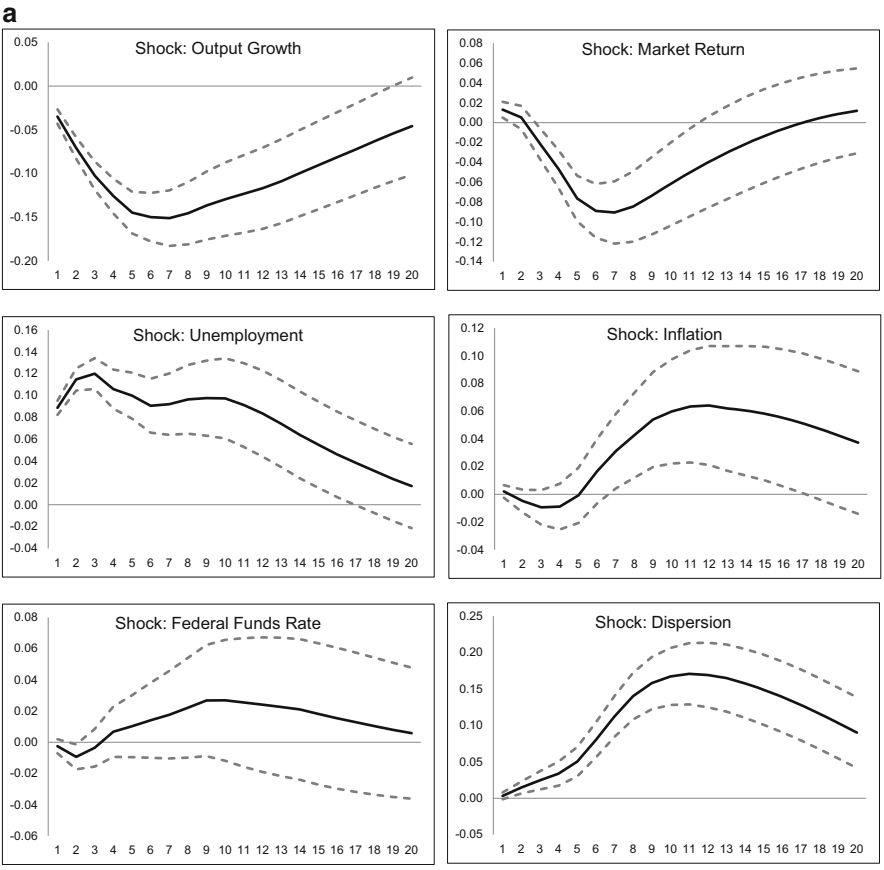

Fig. 8 (**a**) Impulse response figures for long-term unemployment (VAR). (**b**) Impulse response figures for short-term unemployment (VAR)

Figure 9 shows how the role played by dispersion shocks changes as we look at different durations of unemployment. For each duration of unemployment, the figure shows the proportion of the forecast error variance of the unemployment rate explained by innovations to the dispersion index at a 5 year horizon. In each case, the dispersion index is placed last in the ordering. The figure displays a striking pattern, where the proportion of the variation in unemployment explained by shocks to the dispersion index increases monotonically with the duration of unemployment. For short-term unemployment (less than 5 weeks), shocks to the dispersion index account for less than 5 % of the overall variation in the unemployment rate. But at the other end, where duration exceeds 26 weeks, dispersion shocks account for about 40 % of the variance of the forecast error.

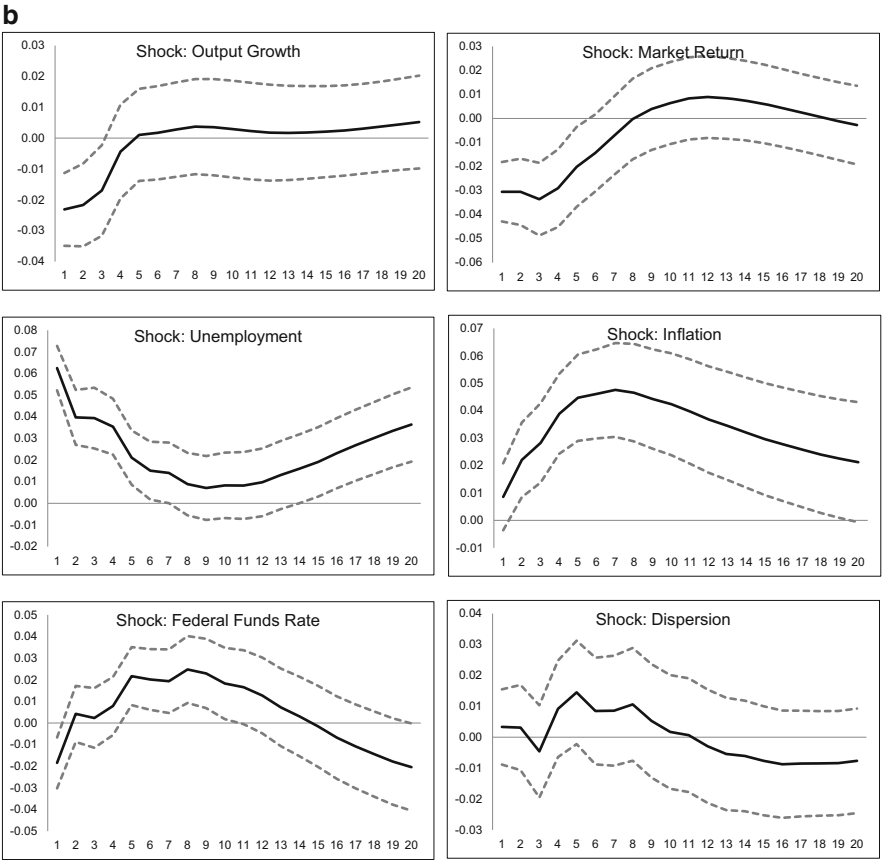

Fig. 8 (continued)

Table 1 Forecase-error variance decomposition for the unemployment rate

Horizon (Quarters)	Growth (%)	Market return (%)	Unemployment rate (%)	Inflation (%)	Fed funds rate (%)	Dispersion (%)
5	49.2	15.0	31.6	2.4	1.2	0.5
10	35.7	17.0	14.0	11.8	4.6	16.9
20	22.6	10.8	8.8	27.6	6.6	23.5

4.2 Sectoral Shocks and Long-Term Unemployment During the Great Recession

We now use the VAR estimated above to examine long-term unemployment during the Great Recession. Long term unemployment (defined as those who were unemployed for more than 26 weeks) constituted 16 % of total unemployment in

Table 2 Forecase-error variance decomposition for the long-term unemployment rate

Horizon (Quarters)	Growth (%)	Market return (%)	Long-term unemployment rate (%)	Inflation (%)	Fed funds rate (%)	Dispersion (%)
5	37.0	3.3	57.1	0.2	0.2	2.2
10	39.9	10.7	27.6	1.8	0.6	19.5
20	31.1	6.3	18.5	5.1	0.8	38.3

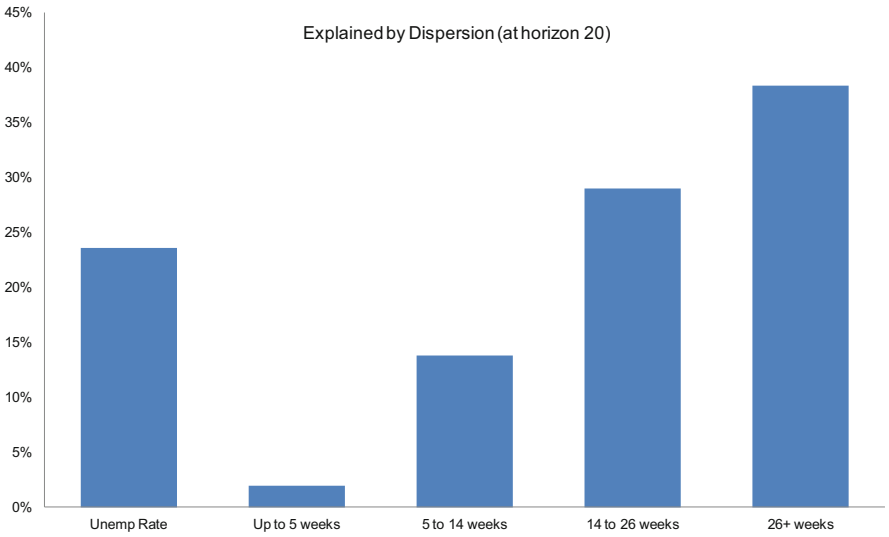

Fig. 9 Variance of unemployment

the fourth quarter of 2007 and 40 % in the fourth quarter of 2012. Notably, it has remained high despite a resumption of growth in output.

Figure 10 plots the long-term unemployment rate since the beginning of 2008, together with two forecasts. The base period for both forecasts is chosen to be the fourth quarter of 2007, the start of the recession as declared by the National Bureau of Economic Research. The forecast horizon extends to the fourth quarter of 2012, or 20 quarters. The line labeled "baseline projection" plots the conditional expectation of the long term unemployment rate over these 20 quarters as of 2007Q4. In other words, it is the VAR's forecast of the long-term unemployment rate as of 2007Q4. For the first year of the forecast horizon, long-term unemployment remained close to the baseline projection. Subsequently, however, long-term unemployment increased dramatically. At its peak in the first half of 2010, the long-term unemployment rate was more than 2 ½ percentage points higher than the baseline value.

The third line in the chart shows what the VAR's forecast of the long-term unemployment rate would have been if the orthogonalized dispersion shocks over the 2008Q1–2012Q4 period had been known at the end of 2007. Dispersion

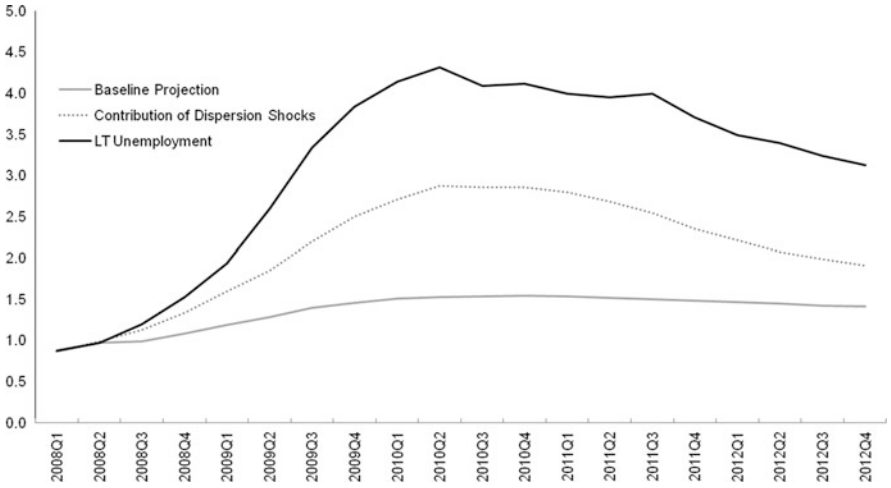

Fig. 10 Decomposition of long-term unemployment during the great recession

shocks turn out to be quite important in explaining the departure of the realized unemployment rate from the baseline forecast. From the beginning of 2009 up until the fourth quarter of 2012, shocks arising from the dispersion index accounted (on average) for more than 40 % of the difference between the actual long-term unemployment rate and its baseline projection. The contribution of shocks to GDP growth, on the other hand, was less than 15 % on average.

5 Conclusion

We have shown that structural shocks (as measured by an index of the cross section variance of stock prices) have a substantial impact on the unemployment rate in a sample that includes the Great Recession of 2007–2009 and the 'Not-So-Great' Recovery of 2010–12. Further, these shocks become more important as the duration of unemployment increases, a finding that accords with the intuition that such shocks should be associated with longer spells of search, as they cause workers to move across sectors.

An examination of the Great Recession shows that sectoral shocks account for close to half of the increase in the long duration unemployment rate that has taken place over this period. Once again, this accords with informal evidence about employment conditions in the construction sector and, to a lesser extent, in finance. In this, the Great Recession is similar to the recession of 1973–75, as sectoral shocks appear to have played a large role at that time as well.

We have also shown that our measure of cross section volatility does better the longer the duration of unemployment under consideration. This seems reasonable as

our measure is meant to capture shocks that cause reallocation across sectors, and such reallocation takes time.

Acknowledgements We thank our discussants, Steve Davis and Valerie Ramey, and other participants at the 2012 Annual Macro Conference of the Federal Reserve Bank of San Francisco. We thank Hites Ahir, Sam Choi and Jair Rodriguez for outstanding research assistance. We acknowledge useful comments from Daniel Aaronson, Larry Ball, Olivier Blanchard, Menzie Chinn, Robert Hall, Joao Jalles, Sam Kortum, Daniel Leigh, Akito Matsumoto, Gian Maria Milesi-Ferretti, Dale Mortenson, Romain Ranciere, Ellen Rissman, Jorge Roldos, David Romer, Kenichi Ueda, Ken West and participants at the Conference on "Globalization: Strategies and Effects" (organized by Bent Jesper Christensen and Carsten Kowalczyk in Kolding, Denmark), Conference on "Long-Term Unemployment" (organized by Menzie Chinn and Mark Copelovitch at University of Wisconsin, Madison), IMF, Paris School of Economics and the Midwest Economic Association meetings in Chicago.

References

Aaronson, D., Rissman, E. R., & Sullivan, D. G. (2004). Can sectoral reallocation explain the jobless recovery? *Economic Perspectives*, Federal Reserve Bank of Chicago.

Abraham, K., & Katz, L. (1986). Cyclical unemployment: Sectoral shifts or aggregate disturbances? *Journal of Political Economy, 94*, 507–522.

Barnichon, R., & Figura, A. (2010). *What drives matching efficiency? A tale of composition and dispersion*. Finance and Economics Discussion Series, Federal Reserve Board, 2011-10.

Barro, R. J. (1986). Comment on 'Do equilibrium real business theories explain postwar U.S. Business Cycles?'. *NBER Macroeconomics Annual, 1*, 135–139.

Benjamin, D., & Kochin, L. (1979). Searching for an explanation of unemployment in interwar Britain. *Journal of Political Economy, 87*(3), 441–478.

Bernanke, B., & Blinder, A. (1992). The Federal funds rate and the channels of monetary transmission. *American Economic Review, 82*(4), 901–921.

Black, F. (1987). *Business cycles and equilibrium*. New York: Basil Blackwell.

Black, F. (1995). *Exploring general equilibrium*. Cambridge: MIT Press.

Brainard, L. (1992). *Sectoral shifts and unemployment in interwar Britain*. NBER Working Paper No. W3980.

Brainard, L., & Cutler, D. (1993). Sectoral shifts and cyclical unemployment. *Quarterly Journal of Economics, 108*(1), 219–243.

Cerra, V., & Saxena, S. C. (2008). Growth dynamics: The myth of economic recovery. *American Economic Review, 98*(1), 439–457.

Clague, E. (1935). The problem of unemployment and the changing structure of industry. *Journal of the American Statistical Association, 30*(189), 209–214.

Davis, S. (1985). *Allocative disturbances and temporal asymmetry in labor market fluctuations*. Working Paper. University of Chicago Graduate School of Business.

Elsby, M., Hobijn, B., & Sahin, A. (2010, March). *The labor market in the great recession*. Brookings Panel on Economic Activity.

Estevao, M., & Tsounta, E. (2011). *Has the great recession raised U.S. Structural Unemployment?* IMF Working Paper.

Groshen, E., & Potter, S. (2003). Has structural change contributed to a jobless recovery? *Current Issues in Economics and Finance* (Vol. 9, No. 8). Federal Reserve Bank of New York.

Kirkegaard, J. F. (2009). *Structural and cyclical trends in net employment over U.S. Business Cycles, 1949-2009: Implications for the next recovery and beyond*. Peterson Institute for International Economics, WP 09-5.

Kocherlakota, N. (2010, August 17). Inside the FOMC, Speech at Marquette, Michigan.
Krugman, P. (2010, September 28). Debunking the structural unemployment myth. *New York Times*.
Lazear, E. P. (2012). The United States Labor Market: Status quo or a new normal?, Mimeo. http://www.kansascityfed.org/publicat/sympos/2012/el-js.pdf
Lee, D., & Wolpin, K. I. (2006). Intersectoral labor mobility and the growth of the service sector. *Econometrica, 74*(1), 1–46.
Lilien, D. (1982). Sectoral shifts and cyclical unemployment. *Journal of Political Economy, 90*, 777–793.
Loungani, P., Rush, M., & Tave, W. (1990). Stock market dispersion and unemployment. *Journal of Monetary Economics, 25*(June), 367–388.
Loungani, P., & Trehan, B. (1997). Explaining unemployment: Sectoral vs. aggregate shocks. Federal Reserve Bank of San Francisco *Economic Review*, pp. 3–15.
Nason, J., & Vahey, S. (2006). *Interwar U.K. unemployment: The Benjamin and Kochin hypothesis or the legacy of "just" taxes?* Federal Reserve Bank of Atlanta, Working Paper 2006-04.
Phelan, C., & Trejos, A. (2000). The aggregate effects of sectoral reallocations. *Journal of Monetary Economics, 45*(2), 249–268.
Ramey, V., & Shapiro, M. (2001). Displaced capital: A study of aerospace plant closings. *Journal of Political Economy, 109*, 958–992.
Romer, C. D., & David H. Romer. (1989). *Does monetary policy matter: A new test in the spirit of Friedman and Schwartz*. NBER Macroeconomics Annual, pp. 121–170.
Romer, C. D., & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review, 94*, 1055–1084.
Sahin, A., Song, J., Topa, G, & Violante, G. (2011). *Mismatch in the labor market: Evidence from the U.S. and the U.K*.
Shin, K. (1997). Inter- and intrasectoral shocks: Effects on the unemployment rate. *Journal of Labor Economics, 15*(2), 376–401.
Thomas, J. (1996). An empirical model of sectoral movements by unemployed workers. *Journal of Labor Economics, 14*(1), 126–153.
Toledo, W., & Marquis, M. (1993). Capital allocative disturbances and economic fluctuations. *Review of Economics and Statistics, 75*(2), 233–240.
Valletta, R. G. (1996, February 16). *Has job security in the U.S. declined?* Federal Reserve Bank of San Francisco Economic Letter 96-06.

U.S. Treasury Auction Yields Before and During Quantitative Easing: Market Factors vs. Auction-Specific Factors

Catherine L. Mann and Oren Klachkin

Abstract Using our unique data set of every U.S. Treasury auction from May 2003 to year-end 2012, we examine the relative importance of market factors vs. auction-specific factors in auction outcomes before and after the onset of quantitative easing in March 2009. We find that prevailing market conditions known in advance of the auction, such as the fed-funds rate, the value of alternative investments (S&P), and market volatility (VIX) are all significant for the auction high-yield. Information embodied in the previous auction of a specific maturity is correlated with the auction high-yield for Bills, implying that the Bills auctions have a forecastable component. Robustness analysis suggests that the Treasury auction market is close to efficient in that the best predictor of an auction high-yield is the market yield on the matched instrument the previous day. However, even so, auction-day innovations matter. Bid-cover is generally significant with higher bid-cover negatively correlated with the auction high-yield. There is some evidence that bidder types (Primary Dealers, Indirect bidders as a proxy for foreign investors, and Direct bidders) vary in their bids, but these results are fragile and depend on model specification. With regard to pre- vs. post-quantitative easing, market structural factors (FedFunds, S&P, VIX) do appear to be differentially correlated with the auction high-yield before and after onset of QE. On balance, the policy of quantitative easing implemented in the secondary market has affected the auction market for U.S. Treasury securities.

1 Introduction

The market for U.S. Treasury securities is large, global, and actively traded. Some $11 trillion of U.S. debt was outstanding in 2012, of which foreign investors held $5.2 trillion. Cross-border trading in long-term Treasury securities peaked at $20 trillion (annual rate) in 2011. Recent years, however, have introduced another

C.L. Mann (✉)
Brandeis University, Waltham, MA, USA
e-mail: CLMannEcon@gmail.com

O. Klachkin
Oxford Economics, Oxford, UK

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_11

important actor into the U.S. Treasury market: the Federal Reserve. From the beginning of the extraordinary monetary policy in March 2009 to end 2012 the Federal Reserve purchased $900 billion U.S. Treasury securities as part of its Large Scale Asset Purchase (LSAP) program, more commonly known as quantitative easing (QE); its holdings of U.S. Treasuries stood at $1.7 trillion at year-end 2012.

With public debt outstanding at about 70 % of nominal GDP, and foreign official investors accounting for 75 % of all foreign holdings, the sustainability of new issuance has taken the spotlight. Analysts are examining the auctions of U.S. Treasury securities for a real-time snapshot of demand, especially demand by foreign and official investors, for U.S. sovereign debt, as the 'canary in the coal mine."[1,2]

Using our dataset on every U.S. Treasury auctions from May 2003 to year-end 2012, we investigate factors underpinning the high yield at the auctions. Are auction high-yields mostly driven by prevailing market conditions known in advance of the auction, such as the fed-funds rate, the value of alternative investments, or market volatility and risk? Is the auction yield for a specific instrument best predicted the previous-day's market-yield for a similar-maturity instrument? Do day-of-auction specific factors, such as bid-cover ratio and type of bidder (such as foreign bidders) have any significant effect on the auction high-yield? Finally, does the auction market behave differently during the boom time through 2007 or after March 2009, when the Fed began its extraordinary monetary policy of intervention into the secondary-market for U.S. Treasury securities?

Using our unique data set of every U.S. Treasury auction from May 2003 to year-end 2012, we examine the relative importance of market factors vs. auction-specific factors in auction outcomes before and after the onset of quantitative easing in March 2009. We find that prevailing market conditions known in advance of the auction, such as the fed-funds rate, the value of alternative investments (S&P), and market volatility (VIX) are all significant for the auction high-yield. Information embodied in the previous auction of a specific maturity is correlated with the auction high-yield for Bills, implying that the Bills auctions have a forecastable component. Robustness analysis suggests that the Treasury auction market is efficient—the best predictor of an auction high-yield is the market yield on the matched instrument the previous day. With regard to auction-day innovations, bid- cover is generally significant with higher bid-cover negatively correlated with the auction high-yield.

[1]Marketable US debt held by the public, (US Treasury, *Monthly Statement of the Public Debt of the United States, December 31, 2012*, table 1). Nominal GDP, revised July 31, 2012, Bureau of Economic Analysis, *National Income and Product Accounts,* table 1. Figures for foreign holdings (table 1) and foreign official holdings (tables 10, 11) from U.S. Treasury, *Foreign Portfolio Holdings of U.S. Securities June 2012*, April 30, 2013. Federal Reserve holdings, Table H.4.1.

[2]During 2012, the Federal Reserve implemented the Extended Maturity Program, in which it sold shorter-term maturities Treasuries (0–3 years) for longer-term maturity Treasuries (6–30 year) while keeping the total dollar value the same. At the end of 2012, the Federal Reserve initiated so-called QE3, by resuming the purchase of long-term US Treasuries from the market at the pace of $45 billion per month.

There is some evidence that bidder types (Primary Dealers, Indirect bidders as a proxy for foreign investors, and Direct bidders) vary in their bids, but these results are fragile and depend on model specification. Market structural factors (FedFunds, S&P, VIX) appear to be differentially correlated with the auction high-yield before and after onset of QE. On balance, the policy of quantitative easing implemented in the secondary market has affected the auction market for U.S. Treasury securities.

We proceed as follows. The next section describes the U.S. Treasury auction process. Section 3 reviews recent literature on whether foreign investors and/or Federal Reserve purchases affect the yield on U.S. Treasury securities in secondary-markets, and the much more limited literature on U.S. Treasury auctions. Section 4 presents our U.S. Treasury auction-level dataset. Section 5 reports on the relationships between the auction high-yields and both market and auction-specific factors. Section 6 concludes.

2 An Overview of the U.S. Treasury Auction Process

The auction process for U.S. Treasury securities follows a formal and rigid structure. The Treasury Department first releases data on its auctions through a number of announcements and disclosures. Prior to an auction, a press release is distributed by the Treasury Department stating the maturity, type, and amount of Treasury security to be sold. The announcement also details the quantity of Treasuries the Federal Reserve has maturing on the issuance date; to keep constant the face value of securities held in the Federal Reserve's System Open Market Account, the central bank will purchase at least some of the securities at that auction. Furthermore they release all details regarding the amount investors are scheduled to purchase via the TreasuryDirect system. Thereafter, the bidding process opens for the remaining securities available for purchase at the auction.

The Treasury Department employs a uniform price (sometimes termed a modified Dutch) auction to sell its offerings. Bids submitted are delineated as either competitive or noncompetitive; competitive bids are placed according to the minimum yield the bidder is willing to accept for a quantity of securities while noncompetitive bids merely bid on the quantity with the yield received determined by the competitive bids. To determine the auction high-yield for the security, competitive bids and their respective yields are organized and accepted by increasing yield until the offering amount is reached.[3]

There are three bidder types: Primary Dealers, Direct bidders, and Indirect bidders. Primary Dealers are the twenty banks and securities broker-dealers that trade directly with the Federal Reserve and act as market makers for Treasury

[3]Competitive bids vastly out-weigh non-competitive bids: for example, at the October 14, 2010 auction competitive bids were $32 billion and non-competitive bids were $12 million.

securities and must bid on U.S. Treasury securities when Treasury securities are auctioned.[4]

Direct bidders have a direct trading relationship with the U.S. Treasury and are bidding for 'their own house accounts'. A Direct bidder can be a domestic U.S. money manager, bank or broker-dealers not classified as a Primary Dealer, or domestic mutual fund managers. Unlike Primary Dealers, Direct bidders are not required to bid on Treasury securities.

Lastly, the Indirect bidders are investors that do not bid via the previous two channels; rather Indirect bidders are 'customers placing competitive bids through a direct submitter, including Foreign and International Monetary Authorities placing bids through the Federal Reserve Bank of New York'.[5] Although foreign official entities can place bids with Primary Dealers, research by Michael Fleming at the Federal Reserve Bank of New York shows that the Indirect bidder has been a valid proxy for foreign and official bidders.[6,7] Metrics associated with Indirect bidders' demand for Treasuries are often used to gauge the willingness of foreign and official entities to buy U.S. Treasury securities.

Because the auction process and metrics are an important foundation for exogeneity of some of our variables in the empirical estimation, it is worthwhile discussing a stylized example, shown in Fig. 1.[8] The first important metric of any auction is the bid-cover ratio. This is the ratio of the total amount bid by all bidders relative to the sum of the amounts allocated to the various bidders. The bid-cover ratio is often used as a metric of overall demand for the issue.[9] Is the bid-cover ratio exogenous to the auction high-yield? Yes, at least weakly exogenous since the bid-cover ratio is the aggregation of individual bidders' reservation yield and quantity. The auction high-yield is determined by the set of bids that are covered, but the

[4]The specific institutions that are direct bidders can change over time. The current list can be found at: http://www.newyorkfed.org/markets/pridealers_current.html accessed May 25, 2012.

[5]Quoted from Public Debt News, footnotes.

[6]http://www.newyorkfed.org/research/current_issues/ci13-1.pdf. See Fleming (2007).

[7]According to Reuters, as of June 2011, China's official purchasing representative has unique and direct access to the auction process and may, therefore, no longer be classified in the Indirect bidder class. http://www.reuters.com/article/2012/05/21/us-usa-treasuries-china-idUSBRE84K11720120521. It has also been noted that China had, in the past, divided up its bids to various Primary Dealers to bid on its behalf so as to mask the specific magnitude of their bids. Rule changes in 2009 altered this behavior. Finally, during 2012, the mapping between Indirect and foreign bidders appeared to weaken. Based on investor-allocation data (which is made public with a delay from the auction date and therefore cannot be used in real time to assess the demand for foreign investors), the foreign share of investors has fallen whereas the hedge fund share of investors has risen. Both could be classified as Indirect bidders. Since our sample ends in 2012, using the Indirect bidder as a proxy for foreign investors would appear to remain valid.

[8]The details on the specific characteristics of the individual bids at any auction, and which specific institutions get the allocation from the auction are not publicly available.

[9]An auction where the bid-cover ratio is below 1 would be termed a 'failure' since the amount bid is less than the amount offered. There is no such auction failure in our sample.

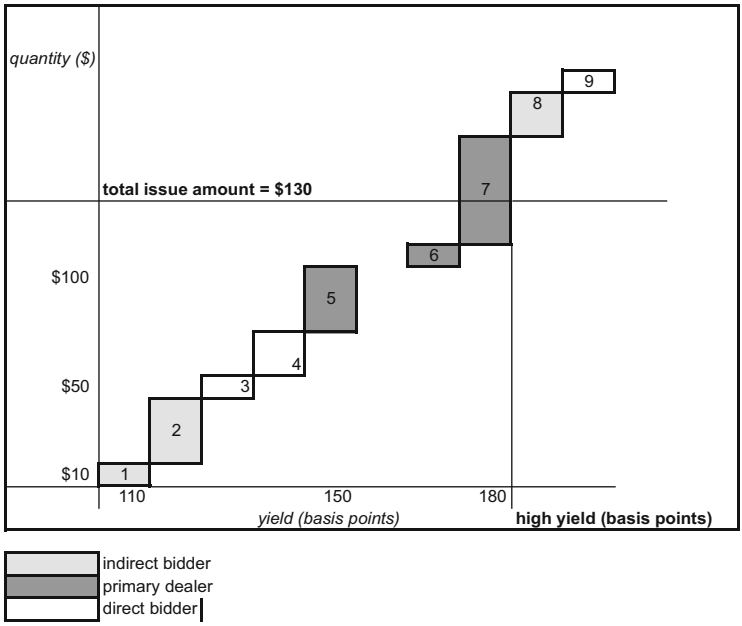

Fig. 1 Stylized example of U.S. Treasury auction

bid-cover ratio includes, as well, all the bidders whose reservation yield is 'too high' (e.g. above the auction high-yield).

An example of such an auction is shown in Fig. 1. $130 is the amount offered of the specific security. $190 is the amount bid by three bidder classes and nine individual bidders. The high-yield is 180 basis points, which is where the sum of the amount bid is the amount offered. Bidders 1–6 get the amount that they bid, as the high yield. Bidder 7 only gets $50 of what they bid. Bidders 8 and 9 do not get any of the issue. Bidders 8 and 9 are included in the bid-cover statistic, but do not affect the auction high-yield because their bids were above the high-yield. All sorts of bids (basis points, $) could generate the same high-yield, but yield a different bid-cover ratio, to the extent that many bids are placed above the auction high-yield.

Does bidder type matter? If all bidder types had the same preference for obtaining some of the issue at auction, the bidder types would be distributed evenly over the range of low to high bids: Some Direct bidders would bid high and some low, some Primary Dealers would bid high, and some low, and so on. Bidder type and auction high-yield would be uncorrelated.

However, in Fig. 1, this assumption is not true—in this example the Indirect bidders (bidders 1, 2) systematically offer to buy the issue at the lowest basis points. These bidders are assured of being allocated some of the issue, and they will be getting a higher yield than what they submitted. This bidder type puts downward pressure on the auction high-yield. So, bidder type could be important in determining the auction high-yield.

Further, information in the secondary market offers the opportunity for strategic bidding. Suppose that Primary Dealers, who must bid at the auction, do not want any of the particular security being issued at the auction, say, because they view the interest rate as inconsistent with inflation expectations. The secondary market for the specific maturity of security being auctioned gives the Primary Dealers information about what the likely high-yield at the auction will be, and the Primary Dealers could therefore bid on the 'high' side; hoping to be allocated relatively little of the auction. Their bids could put upward pressure on the high-yield compared to what it otherwise might be in an environment of homogeneous preferences among bidders.

3 Literature Context

This paper contributes to the literature on U.S. Treasury market in three ways. First, it fits into the literature on foreign demand for U.S. assets and interest rates, but it uses disaggregated auction data and Indirect bidders as a proxy for foreign demand, rather than secondary-market aggregates and cross-border net flows. Second, it fits into the literature on the impact of Federal Reserve purchases on U.S. Treasury yields, but it does so by evaluating whether coefficients in the auction high-yield regressions differ before vs. after the onset of Federal Reserve intervention into the secondary market. Third, it contributes to the literature on U.S. Treasury auctions, particularly considering market information and auction-specific information available prior to the auction, as well as auction-specific innovations on auction day.

3.1 The Literature on Foreign Demand for U.S. Assets, Including U.S. Treasuries, and Interest Rates

The literature often reaches different conclusions as to the importance of foreign demand for U.S. asset for U.S. interest rates. None of this literature uses data on individual securities. Warnock and Warnock (2005), using aggregate data and the 10-year U.S. Treasury interest rate, conclude that, foreign capital inflows reduce the yield on 10-year Treasuries by 150 basis points. On the other hand Rudebusch et al. (2006), focusing on trying to explain the conundrum of low long-term U.S. interest rates, use aggregate data and find that purchases of long-term Treasuries by foreign central banks between 2004 and 2005 had little explanatory power. Beltran et al. (2010) use econometric models that span these two previous works. They find that relaxing the econometric assumptions in Warnock and Warnock yields results that

support the findings of Rudebusch et al. and visa-versa.[10] Therefore, Beltran et al. conclude that the existing literature is not a reliable guide as to whether foreign official purchases affect U.S. interest rates.

Bernanke et al. (2011) find that international capital flows played an important role in lowering Treasury yields and returns on other U.S. assets, namely mortgages, in the years leading up to the 2008 financial crisis. Beltran et al. (2011) model foreign private and official purchases as endogenous, and examine the risk premium on Treasury securities as the metric of impact of foreign demand, an approach first considered by Sierra (2010), but find no impact of foreign demand. Sierra concludes that official and private foreign purchases have opposite effects on U.S. Treasury yields so that the two forces cancel out, yielding the same overall conclusion that foreign demand is not important factor in U.S. Treasury yields in the secondary market.

3.2 The Literature on the Federal Reserve Asset Purchase Programs and Interest Rates

This literature yields the more systematic conclusion that the Federal Reserve asset purchase programs have influenced interest rates in the secondary market. Doh (2010), Gagnon et al. (2010), and Neely (2010) using aggregate data, suggest that these various programs generally did reduce interest rates. Although, Stroebel and Taylor (2009) conclude that the purchase of mortgage-backed securities, in particular, did not have much impact on interest rates. In the most detailed examination, D'Amico and King (2010) use CUSIP data to find that the LSAP program reduced interest rates on the instruments (individual CUSIPs) that were purchased. Similar to D'Amica and King, we use a dataset with individual securities.

3.3 The Literature on Auction High-Yields

Research in various contexts has examined the relationship between matched maturity instruments of different vintages. This on-the-run/off-the-run literature is well reviewed by Pasquariello and Vega (2009), and concludes that bid-ask daily differentials of on-the-run and off-the-run instruments are narrowest immediately following an auction of the matched maturity instrument. Other researchers examine whether these differentials vary during times of financial turbulence. For example, research by Furfine and Remolona (2002), around the time of the Russian debt crisis

[10]Warnock and Warnock assume that foreign official inflows are exogenous, an assumption that Beltran et al. reject. Rudebusch et al. use trending data, which, upon examination by Beltran et al. is shown to yield spurious correlations.

in 1998, found that although the trading activity in already-issued securities varied around auction days, prices adjusted immediately between the newly on-the-run and the newly off-the-run security.

Few papers address the relationship between the auction high-yield and the information set of previously issued securities of the same maturity. Jegadeesh (1993) found a systematic 4 basis point difference between the auction yield and the maturity-matched instrument in the secondary market (and used that as a benchmark to consider whether Salomon Bros. had manipulated the market). Similarly, Hou et al. (2011) show that, in a several day window prior to and after an auction, the yield of the matched-maturity security is some 2–6 basic points below the auction yield, narrowing to zero on the auction day. These authors do not examine the time-series properties of this differential, nor whether it has explanatory power for future auctions of a similar maturity instrument.

4 Methodology and an Overview of the Auction Dataset

Our method is grounded in models of imperfect asset substitution, portfolio balance, and preferred habitat. The models allow for quantity supplied (or reduced as in the Federal Reserve LSAP) or demanded (e.g. by foreign investors) to impact price—that is, actors need not be atomistic in the marketplace. The portfolio balance model has a long history in foreign exchange markets, and has been used to assess the impact of foreign exchange intervention in the markets, with differing degrees of success. The portfolio balance model and imperfect asset substitution models were considered early in the life of the Euro for whether the type and amount of foreign asset purchases impacted the Euro's value.

The preferred habitat model is the foundation of D'Amico and King's work—their focus is on the shock reduction in supply of U.S. Treasury securities (UST) via the LSAP programs, whereas we parameterize bidders-at-auction as a shock increase in demand for UST. Given that the supply of UST offered at auction is known to the bidders, but who will bid and the characteristics of their bids is not, we conclude that the demand side prices the high-yield at the auction.

For our data, we construct a unique panel dataset of various metrics of every U.S. Treasury auction from May 2003 until year-end 2012. These metrics include the maturity of the security being sold at auction, the auction high-yield, the bid-cover ratio, the auction amount awarded to the three bidder types (Primary Dealers, Direct bidders, Indirect bidders). We also collected data on the secondary-market yields of comparable securities (matched by maturity) for the day before the auction. Auction data were obtained from the Bloomberg Profession Service while all secondary market yields were obtained from the Federal Reserve's H.15 database.[11]

[11]The H.15 data are reported as 'market yield, constant maturity, quoted on investment basis'.

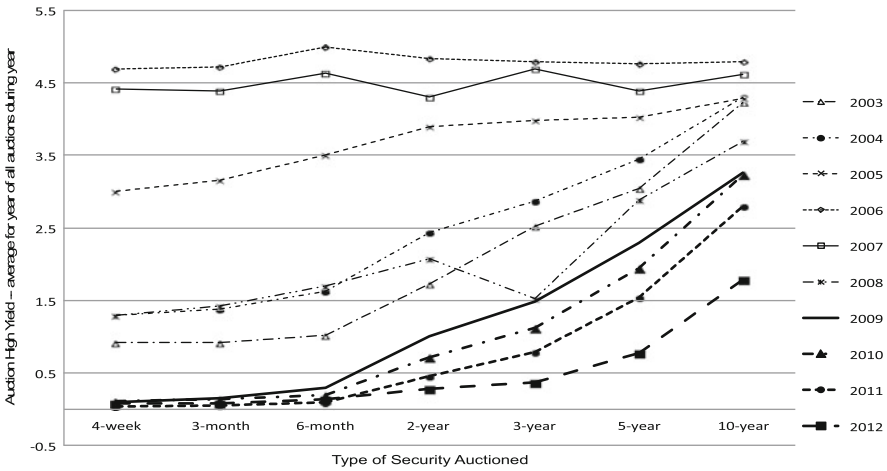

Fig. 2 Auction high-yield curve

As a backdrop to our analysis, we first review some basic indicators of recent U.S. Treasury auctions over our time period of 2003–2012, including the behavior of the yield curve and the bid-to-cover ratio, as derived from the auction-specific data, as well as the allocations of the auction issue to the bidder types.

Taking these auction-specific data, and aggregating each maturity over the year, Fig. 2 shows the evolution of the auction high-yield over the time period of our analysis, 2003–2012. Each line represents, for that year, the average of the high-yields obtained at auction of the respective maturity security. This auction-based yield curve has shifted down substantially, which, of course was the objective of both standard open market operations (early in the period) as well as the intervention by the Federal Reserve into the secondary markets (after 2009). This shifting down of the auction high-yield curve also reflects the market's perception of the state of the U.S. economy.

The bid-to-cover ratio (calculated from the auction-specific data as the average of all the auctions during the year for the specific maturity) increased over the period, most notably for auctions during 2009–2011, as shown in Fig. 3. The bid-cover ratios increased most dramatically for auctions of short-term UST (Bills, with maturity of less than 1 year) in 2009 and 2010. For medium-term maturity Bonds (2, 3 year) the increase in bid-cover is also apparent. Bid-cover for auctions at longer maturities (5, 10) flattened out after 2010, perhaps revealing investors' reticence to hold the longer-term UST securities on the grounds that future U.S. government borrowing implies either higher future interest rates (making today's longer-term, relatively lower coupon investments unprofitable), or potential future inflation (also making holding of low coupon, longer-dated securities unwise), or a combination of both.

The auction data do not reveal the amount of bid by each bidder type, but rather the amount awarded. Nevertheless, using the auction-specific data, we can view the

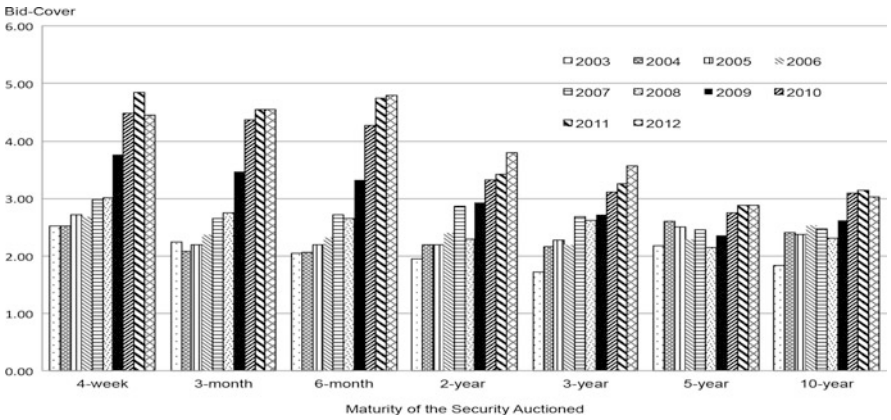

Fig. 3 Bid-Cover at auction, by maturity of security and year

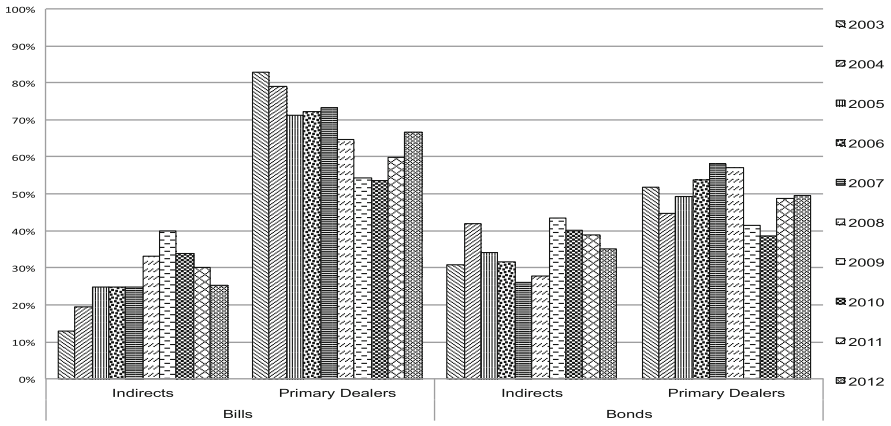

Fig. 4 Allocation of auction, by bidder type, maturity of security, and year

pattern of allocations over time. Figure 4 shows the share of the auction awarded to Primary Dealers and Indirect bidders (Direct bidders are the residual share.) Primary Dealers are awarded the lion's share, particularly of Bills auctions, but their share fell during the periods of greatest turmoil (2009). The share awarded Indirect bidders (the proxy for foreign investors) rose at the short-end during the turmoil, but has stayed high at the longer end in recent years. If the shares awarded in 2006–2007 were 'normal', it would appear that Indirect bidders are now larger awardees, shifting their portfolios toward the longer end of the maturity spectrum relative to their bids in 2006–2007. It may also be the case that the re-introduction of the 30-year instrument was taken-up particularly by the foreign official investors, thus lengthening the maturity of their portfolios.

The auction-specific data do not include information as to the nationality of the Indirect bidders. However, some insights can be gleaned from the Treasury

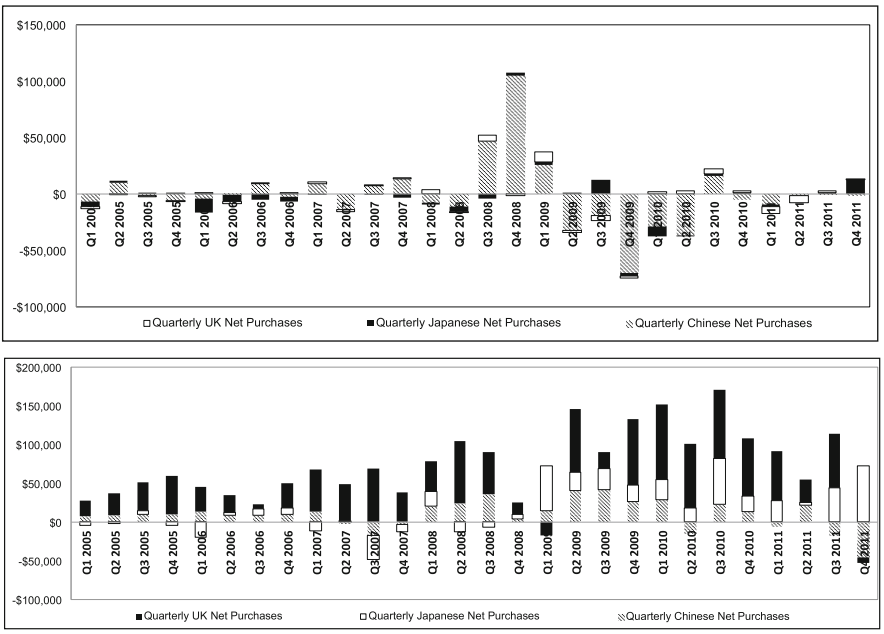

Fig. 5 Flow of UST to/from biggest U.S. Foreign Creditors (in US$ millions); Short-term (*top panel*) and long-term (*bottom panel*) (*Data source*: http://www.treasury.gov/resource-center/data-chart-center/tic/Pages/country-liabilities.aspx)

International Capital (TIC) system data. As is well known, China and Japan are the largest holders of U.S. Treasury securities, with an end of year (2012) stock at $1.2 trillion and $1.1 trillion, respectively. On the other hand, it is the flow of purchases of UST that might better reflect the Indirect bidders' award at an auction. As shown in Fig. 5 (for short-term UST in the top panel and long-term UST in the bottom panel) on a flow basis, using secondary market data, Chinese and Japan purchases continued apace, at least through 2010, but then China started to taper off. The United Kingdom is also a large player but its flow purchases are less regarded as for its own account and more as a conduit for other investors.[12]

5 Regression Model and Analysis

We turn now to our main objective, which is to model the auction high-yield. Our objective is to consider the relative importance of information available to bidders

[12]It is beyond the scope of this paper to attempt a matching of the secondary market data and the auction data, although this would be an interesting exercise.

going into the auction as compared to information revealed by the auction. We also want to address whether statistical relationships change for auctions during the boom time versus auctions after the onset of extraordinary monetary policy in March 2009.

Our econometric method is similar to D'Amico and King, but our objective is to explain the auction high-yield, not the secondary market yield. Factors relevant for the auction bidder and included in their information set going into the auction include overall financial market conditions and any information that could be gleaned from previous auctions of the specific maturity on offer. We can also consider an 'efficient market' specification whereby all information available to the investor prior to the auction is embodied in the interest rate on the matched maturity instrument prevailing in the market the day before the auction. Factors relevant for the high-yield that occur within the auction itself include bidder type and the bid-cover ratio.

$$\begin{aligned} \boldsymbol{i}_{t,j,} = {} & \mathrm{a}_{0,\boldsymbol{j}} + \mathrm{a}_1 \boldsymbol{FedFunds}_{t-1} + \mathrm{a}_2 \boldsymbol{S\&P}_{t-1} + \mathrm{a}_3 \boldsymbol{VIX}_{t-1} + \mathrm{a}_4 \boldsymbol{BidCover}_{t,j} \\ & + \mathrm{a}_5 \boldsymbol{Abs\$PD}_{t,j} + \mathrm{a}_6 \boldsymbol{Abs\$Indirects}_{t,j} + \mathrm{a}_7 \boldsymbol{Abs\$Directs}_{t,j} \\ & + \mathrm{a}_8 \boldsymbol{PrevAuctionDif}_{tj-1,j} + \left\{ \mathrm{a}_9 \boldsymbol{PrevDayMktYield}_{t-1,j} \right\} + \boldsymbol{e} \end{aligned}$$

5.1 Dependent Variable

- $\boldsymbol{i}_{\mathrm{t,j}}$: Auction high-yield for auction on day $\boldsymbol{t}$ of U.S. Treasury security with maturity $\boldsymbol{j}$
- t ranges from May-2003 to December 2012, a total of 1926 auctions
- j maturities include Bills (4-week, 3-month, 6-month) and Bonds (2, 3, 5, 10, 30-year).

5.2 Independent Variables: Information Set Prior to the Auction, and Innovations on Auction Day

- $\mathrm{a}_{0,\boldsymbol{j}}$; maturity fixed effects

Financial market conditions (lagged one day before the auction) are in the information set of the bidders as they come to the auction, and are common to all securities.

- $\boldsymbol{FedFunds}_{t-1}$: The Fed-funds rate anchors all the auction high-yields.
- A higher Fed-funds rate should be positively associated with the auction high-yield.
- $\boldsymbol{S\&P}_{t-1}$: The S&P index is a proxy for investment alternatives.

- A higher S&P should be negatively associated with the auction high-yield as investors require a higher yield at auction to induce them away from high-return equity markets.
- $\boldsymbol{VIX_{t-1}}$: VIX is a measure of risk.
- Risk and auction high-yield are likely negatively correlated, with higher risk inducing investors, at the margin, to bid more aggressively for the Treasury securities.

Innovations occur at the auction on day t and are specific to the auctioned maturity j.

- $\boldsymbol{BidCover_{t,j}}$: The total of bids relative to the issuance at auction dated t of security of maturity j.
- A higher bid-cover ratio should generally be negatively associated with the auction high-yield.
- $\boldsymbol{Abs\$PD_{t,j}}$: The dollar amount of securities of maturity j awarded to Primary Dealers at auction dated t.
- $\boldsymbol{Abs\$Indirects_{t,j}}$: The dollar amount of securities of maturity j awarded to Indirect bidders at auction dated t.
- $\boldsymbol{Abs\$Directs_{t,j}}$: The dollar amount of securities of maturity j awarded to Direct bidders at auction dated t.
- For these three allocation variables, the expected sign and even significance is unclear. If all the bids are distributed normal around the auction high-yield, then the amount allocated to any specific bidder type should be uncorrelated with the auction high-yield, and further, would be insignificant once controlling for bid-cover.

As discussed in the context of the example of the U.S. Treasury auction in Sect. 2, if bidder types do differ systematically, then the amount awarded to one vs. another bidder type could matter for the auction high-yield. And, if the three bidder types are systematically different in how they bid, then the coefficient sign could offer some insight into the type of bid relative to the auction high-yield. For example, suppose that all the Indirect bidders systematically bid 'low'; in this case, the allocated award to Indirect bidders would tend to have a negative sign, with a higher allocation to these bidders bringing down the auction high-yield. Alternatively, suppose the Primary Dealers tend to bid 'high', then a large award to them might be associated with upward pressure on the auction high-yield.

5.2.1 Auction-Specific Data in the Information Set of the Bidders

Previous research outlined in the literature review suggested that there may be a systematic difference between the auction high-yield and interest rate in the secondary market on the matched-maturity security the day before the action. Accordingly, we define two variables to investigate this hypothesis.

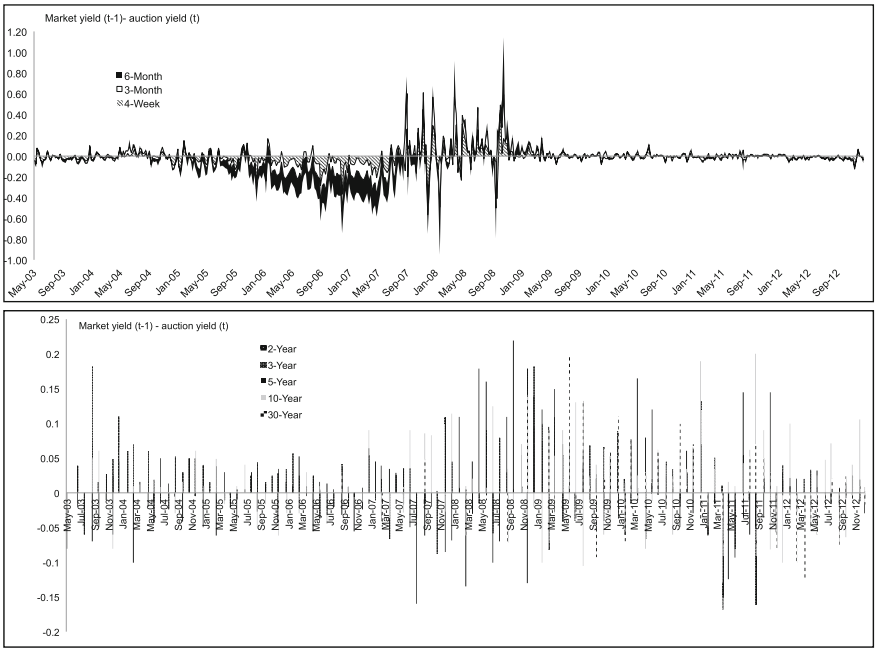

Fig. 6 Previous-day market yield vs. auction high yield for bills (*top panel*) and for bonds (*bottom panel*)

- ***PrevAuctionDif*** $_{tj-1j}$: The difference between the auction high-yield $\boldsymbol{i}_{\mathrm{t,j}}$ and the previous-days matched-maturity security j *from the previous auction of maturity j*, which is dated $\mathrm{tj}-1$ to differentiate the timing from the running calendar days. That is, the number of days before the auction dated time t depends on the frequency and time-table of the auction of maturity j.
- Figure 6 (for Bills and Bonds separately) suggests that this variable is not random, and therefore may contain information that could be exploited by auction participants.[13]
- ***PrevDayMktYield*** $_{t-1j}$: For a robustness check, and to consider whether the auction is an 'efficient market', the previous-day's secondary market yield for maturity j is included as a final explanatory variable. If the auction market is efficient in the classical sense, then this variable should have coefficient 1, and all other variables should be insignificant in the regression.

[13]Estimating the factors underlying the behavior of this variable, as the left-hand side, is put aside for another paper. The point of including this variable in the estimation of the auction high-yield is that the auction-specific information embodied in this variable is known by all bidders going into the auction for maturity j.

5.3 Estimation

Using our dataset on every U.S. Treasury auctions from May 2003 to year-end 2012, we investigate factors underpinning the high yield at the auctions. We are interested in the relative importance of auction yields as driven by prevailing market conditions known in advance of the auction relative to auction-specific information known before the auction. How relevant is information revealed by the auction? Finally, does the auction market behave differently during the boom and bust time versus after March 2009, when the Fed began its extraordinary monetary policy of intervention into the secondary-market for U.S. Treasury securities?

Preliminary regressions revealed statistical differences in Bills vs. Bond markets. So the regressions were estimated separately for these two maturities. Table 1 shows results for Bills and Bonds, before and after QE, and considers market variables known in advance (Fed Funds, S&P, VIX) and auction-specific information known in advance (the difference between the auction high-yield and previous day market yield for the last time an auction of a particular maturity took place—PrevAuctionDif—as described and shown in Fig. 6), and the information revealed by the auction (bid-cover and allocations to bidder type).

5.4 Market Variables in the Information Set

First, all the market variables are significant and of the correct sign, except for the Fed Funds rate during the QE period for Bonds. The high-yield at Bills auctions are more highly correlated with FedFunds rate, which makes sense. Higher S&P is negatively correlated with auction-yield yield, with a greater negative correlation with the Bond auction high-yield; again this makes sense given the substitution of longer-term assets. Higher volatility as measured by VIX is negatively correlated with both auctions, implying flight to Treasuries in a risky environment; the negative correlation is more notable for Bonds in the QE period.

5.5 Auction-Specific Variables in the Information Set

The information contained in the previous auction—how the high-yield at the auction varied relative to the market-yield the previous day for the matched instrument—is significant and positive for Bills in both periods. This suggests that auction-specific information that is known in advance can be used to forecast the auction high-yield for Bills. Thus, the auction bids could be strategic.

For Bonds, the information contained in the previous auction of the matched maturity instrument does not add to the information set of the bidders. These

Table 1 Estimation of auction high-yield before and during QE for bills and bonds

	Boom/Bust	QE	Boom/Bust	QE
	May 2003–March 2009	April 2009–December 2012	May 2003–March 2009	April 2009–December 2012
	Bill auctions		Bonds auctions	
In the information set				
Market variables				
Fed Funds	1.003***	0.7963	0.4287***	-1.5575**
	0.01149	0.0413	0.03721	0.7592
S & P	−0.001628***	−0.0001413***	−0.001528***	−0.004075***
	0.000123	0.0000204	0.0003635	0.0003431
VIX	−0.01622**	−0.001401	−0.01784***	−0.05629***
	0.001613	0.0003283	0.00401	0.006356
Auction-specific				
PrevAuctionDif	0.1614**	0.1815**	0.6786	−0.1948
	0.07916	0.08143	0.6156	0.4966
Auction-specific shocks				
Bid-Cover	−0.07222**	−0.2054***	0.132	0.05912
	0.03153	0.003735	0.1121	0.121
Abs$PD	0.007099*	0.0020365***	−0.06252***	0.02169
	0.003698	0.0005096	0.0161	0.01933
Abs$Indirect	0.008679*	0.001962***	−0.059***	0.02211*
	0.004875	0.0005673	0.01777	0.01263
Abs$Direct	0.04313	0.002749**	−0.03741*	0.0328*
	0.1575	0.001381	0.02205	0.01749
NOB	886	600	211	229
R2	0.9606	0.4688	0.7493	0.0059

OLS standard errors

***significant at 1 %; **significant at 5 %; *significant at 10 %

auctions are less frequent, and therefore any information in the previous auctions would be overtaken by information in the market variables.

5.6 Auction-Specific Information Revealed at the Auction

For Bills, but not for Bonds, the bid-cover ratio is negatively correlated with the auction high-yield. The bid-cover ratio is more strongly correlated in the QE period, which is consistent with the higher bid-cover ratios observed in the Figures above.

Considering the hypothesis of whether bidder type matters, the results are mixed. For Bills, in the pre-QE period, the sign is significant but positive for Direct bidders. Direct bidders do not have to bid at the auction. But, knowing the information from the previous auction, and bidding strategically, Direct bidders could potentially shade the auction high-yield up from where it might have been based on homogeneous and atomistic bids. In the QE period, the Primary Dealers and Indirects may be following this bidding strategy.

Considering the Bond auctions, the signs are the more-expected negative; a higher bid by a bidder type (that was accepted, since we only observe allocations, not bids) would normally be expected to reduce the auction high-yield. The amount awarded to all three types of bidders are statistically significant. The Primary Dealers appear to be relatively more negatively associated with the auction high-yield compared to the Indirect bidders, which is not consistent with the hypothesis that foreign investors bid 'lower' than the average. However, foreign bidders do appear to bid relatively lower than the Direct bidders. Once the QE period begins, none of the award magnitudes significant at the 1 % or 5 % level. Note also the very poor fit of the QE-period regression for Bonds, which may reflect the important role that the QE policy plays for the auction results even though QE is implemented in the secondary market.

Table 2 is a robustness check that considers whether the auction market is efficient in the sense that all information relevant to the auction is embodied in the previous-day's market yield for the matched-maturity instrument. This robustness check reveals, generally, that both maturities and time periods are nearly the classical efficient market: The coefficient on the previous-day's market yield for the matched instrument is close to 1.

In term of the auction-specific shocks, which are not in the information set and therefore could still be significant even in an efficient market, a higher bid-cover is negatively correlated with the auction high-yield, except for the pre-QE period for Bills. As in the previous specification, the allocations are all positively correlated with the auction high-yield for Bills in the QE-period, which is not expected except in the case of strategic bidding. For Bonds, in the QE-period, the allocation by Indirects, the proxy for foreign bidders, is statistically significant and negative, suggesting that these bidders bid 'low' in order to get the allocation.

Table 2 Robustness check based on 'Efficient Market'

	Boom/Bust	QE	Boom/Bust	QE
	May 2003–March 2009	April 2009–December 2012	May 2003–March 2009	April 2009–December 2012
	Bill auctions		Bonds auctions	
In the information set				
Auction-specific				
PrevDayMarketYield	0.9822***	0.9123***	0.9898***	1.00543***
	0.004254	0.01348	0.005777	0.006851
Auction-specific shocks				
Bid-Cover	0.03326**	−0.007972***	−0.03434***	−0.05237***
	0.01624	0.001469	0.01214	0.01315
Abs$PD	0.005324***	0.0006938***	0.000582	−0.004184*
	0.002051	0.0002074	0.001722	0.002414
Abs$Indirect	0.007845***	0.0006358**	−0.001244	−0.00385**
	0.002602	0.0002672	0.001898	0.00163
Abs$Direct	0.01868*	0.002306***	0.003717	−0.003208**
	0.009764	0.0006146	0.002314	0.002215
NOB	886	600	211	229
R2	0.9846	0.9459	0.9972	0.9982

OLS standard errors

***significant at 1 %; **significant at 5 %; *significant at 10 %

6 Conclusion

This paper uses a unique auction-by-auction data set to examine factors underpinning the auction high-yield for U.S. Treasury auctions over the time period 2003 to end 2012. We find that prevailing market conditions known in advance of the auction, such as the fed-funds rate, the value of alternative investments (S&P), and market volatility (VIX) are all significant for the auction high-yield. Information embodied in the previous auction of a specific maturity is correlated with the auction high-yield for Bills, implying that the Bills auctions can be forecast based on information from the previous auction, not just from market factors alone.

However, robustness analysis shows that all these lagged factors—both market and auction-specific—are embodied in the previous-day's market yield on the matched maturity instrument. In this sense, the Treasury auction market is efficient—the best predictor of an auction high-yield is the market yield on the matched instrument the previous day. The coefficient is not 1.0 however, particularly for Bills, leaving open the possibility for strategic auction behavior.

We also considered the role for auction-day information, such as bid-cover ratio and allocations by bidder type to influence the auction high-yield. Such auction-day information on bid-cover is generally significant and expected sign: Higher bid-cover is negatively correlated with the auction high-yield, particularly in the efficient-market specification. With regard to bidder allocations, the results are quite mixed. There is some indication that Indirect bidders bid 'low' in Bond auctions in the QE-period, but these results are fragile.

Finally, does the auction market behave differently during the boom/bust time through March 2009 versus during the period when the Fed is implementing quantitative easing? Market structural factors (FedFunds, S&P, VIX) do appear to be differentially correlated with the auction high-yield before and after onset of QE. Bid-cover appears to convey different information comparing the two time periods. Bidder allocations may differ pre-QE versus during-QE; although these results are less certain. On balance, there is some indication that the policy of quantitative easing in the secondary market has affected the auction market for U.S. Treasury securities.

References

Beltran, D., Kretchmer, M., Marquez, J. & Thomas, C. (2010, February). *U.S. Treasury yields and foreign holdings of U.S. Securities: An interim report*. Working paper. Board of Governors of the Federal Reserve System.

Beltran, D., Kretchmer, M., Marquez, J., & Thomas, C. (2011, May). *Foreign official holdings of U.S. Treasury Securities and U.S. Interest Rates: A lens on the 'Savings Glut'" Federal Reserve Board of Governors*. Print.

Bernanke, B. S., Bertaut, C., DeMarco, L. P., & Kamin, S. (2011, February). *International capital flows and the returns to safe assets in the United States, 2003-2007*. International Finance Discussion Paper No. 1014, Federal Reserve Board of Governors.

D'Amico, S., & King, T. B. (2010, September). *Flow and stock effects of large-scale treasury purchases*. Finance and Economics Discussion Paper Series 2010-52 Federal Reserve Board of Governors.

Doh, T. (2010). The efficacy of large-scale asset purchases at the zero lower bound. *Federal Reserve Bank of Kansas City Review*, second quarter, pp. 5–34.

Fleming, M. (2007, January). Who buys US Treasuries at auction? *Current Issues in Economics and Finance* (Vol. 13, No. 1). Federal Reserve Bank of New York.

Furfine, C. H., & Remolona, E. M. (2002, June).What's behind the liquidity spread? On-the-run and off-the-run US Treasuries in autumn 1998. *BIS Quarterly Review*, pp. 51–58.

Gagnon, J. E., Raskin, M., Remache, J., & Sack, B. (2010). *Large-scale asset purchases by the Federal Reserve: Did they work?* Federal Reserve Bank of New York Staff Report No. 441.

Hou, D., Fan, H. Zhang, J. (2011). *Anticipated and repeated shocks in liquid markets*. Draft dated February 21.

Jegadeesh, N. (1993). Treasury auction bids and the Salomon squeeze. *Journal of Finance, 48*(4), 1403–1419. September.

Neely, C. J. (2010, July). *The large scale asset purchases had large international effects*. Working Paper Series, 2010-018A, Federal Reserve Bank of St. Louis.

Pasquariello, P., & Vega, C. (2009). The on-the-run liquidity phenomenon. *Journal of Financial Economics, 92*(1), 1–24.

Rudebusch, G., Swanson, E., & Wu, T. (2006). The bond yield 'Conundrum' from a macro-finance perspective. *Special Issue of the Bank of Japan's Monetary and Economic Studies, 24*, 83–128.

Sierra, J. (2010, June). *International capital flows and bond risk premia*. Working Paper 2010-14, Bank of Canada.

Stroebel, J. C., & Taylor, J. B. (2009). *Estimated impact of the fed's mortgage-backed securities purchase program*. NBER Working Paper No. 15626.

Warnock, F. E., & Warnock, V. C. (2005, September). *International capital flows and U.S. interest rates*. Board of Governors of the Federal Reserve System. Federal Reserve System.

Exchange Rate Volatility, Euro Effect and the Two Margins of Trade: Evidence from Monthly Trade Data

Florian Johannsen and Inmaculada Martínez-Zarzoso

Abstract The aim of this paper is to provide further empirical evidence on the relation between exchange rate volatility, currency unions and trade. The novelties with respect to previous research are threefold. First, monthly trade and exchange rate data are used to take into account the short term effects of volatility on the bilateral exchange rate. Second, disaggregated trade data are used to deal with differences among industries and between final and intermediate goods. Finally, the existence of zero trade flows is taken into account by distinguishing between the extensive and the intensive margins of trade. Investigating the impact of exchange rate volatility and the Euro at the same time allows us to disentangle the effect of a common currency beyond the elimination of any variation in the exchange rate. Furthermore, the developments of the past years with the financial crisis and the EU enlargement to the East are taken into account, yielding additional findings and policy implications.

1 Introduction

The end of the Bretton Woods system in the early 1970s and the adoption of a floating exchange rate regime in 1973 raised the question of how the resulting increase in exchange rate volatility causes exchange rate risk and affects international trade and welfare. The EMU and the introduction of the Euro, associated with the abolition of several European currencies, led to a huge debate among economists concerning its effects on trade. Very recently, the global financial crisis as well as the catalyst of the debt crises and the massive central bank interventions especially in Europe and the U.S. increased exchange rate volatility again and brought the topic back on the agenda.

F. Johannsen • I. Martínez-Zarzoso (✉)
Georg-August Universität Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany
e-mail: martinei@uji.es

I. Martínez-Zarzoso
Universitat Jaume I, Castellón, Spain

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_12

In light of the recent events, especially the case of Europe and the Euro is worth a second glance. The question whether joining a currency union and thereby eliminating exchange rate volatility with the member countries boosts trade significantly is a very relevant question for many Central and Eastern European countries. The fact that countries like Poland have postponed their accession to the Euro is a strong indicator of the uncertainty whether or not the negative consequences of a currency union outweigh positive effects, especially on trade.

Early theoretical studies including Clark (1973) and Hooper and Kohlhagen (1978) predict negative effects of exchange rate volatility on trade, but their predictions are based on strong assumptions. When these assumptions are relaxed, results depend on a variety of factors. Some of them are whether the firms are active in several countries (Makin 1978), the speed of adjustment of the inflation rate to exchange rate movements (Cushman 1983, 1986), the degree of flexibility of the firms in adjusting inputs (Canzoneri and Clark 1984), or in changing target markets (Broll and Eckwert 1999), the degree of risk aversion of the firm (De Grauwe 1988; Viaene and de Vries 1992) and the types of shocks firms are exposed to (Barkoulas et al. 2002).

The empirical literature on the relationship between exchange rate volatility and trade does not present unambiguous evidence. Some studies find significant negative effects (e.g. Chit et al. 2010), some others positive effects (e.g. Klein and Shambaugh 2006), but most recent studies show mixed evidence (e.g. Hondroyiannis et al. 2008; Boug and Fagereng 2010; Eicher and Henn 2011).[1] Instead, evidence on currency unions and unilateral dollarization is much clearer. While early studies find large effects and trade to triple (Rose 2000; Frankel and Rose 2002), most recent studies report positive effects on trade of around 5–30 (e.g. Flam and Nordström 2007; Baldwin and Di Nino 2006; Eicher and Henn 2011). Other authors do not find significant effects (e.g. Berger and Nitsch 2008; Santos Silva and Tenreyro 2010). The usual argument of why effects for currency unions are more significant is that a currency union goes beyond the mere elimination of exchange rate variability and lowers transaction costs to a greater extent.[2]

Although almost all empirical studies are based on the gravity equation of trade, they differ significantly in methodology, panel of countries, time frame, volatility measure and degree of disaggregation of the trade data.

The aim of this chapter is to provide further empirical evidence on the relationship between exchange rate volatility, currency unions and trade. There are three main novelties with respect to previous research. First, higher frequency trade and exchange rate data are used to take into account the short term effects of volatility in the bilateral exchange rate. Second, disaggregated trade data are used to deal with differences among industries and between final and intermediate goods. Finally, and in contrast to previous studies, several econometric problems are taken into account.

[1] See literature surveys of Côté (1994), McKenzie (1999), Ozturk (2006), Bahmani-Oskooee and Hegerty (2007) and Auboin and Ruta (2011).

[2] Baldwin (2006) provides a good overview on the early literature.

Those include the existence of zero trade values, the distinction between extensive and intensive margins of trade and the simultaneous incorporation into the model of exchange rate volatility and currency effects. This later issue allow as disentangling the effect of a common currency beyond the elimination of any variation in the exchange rate with other members. Furthermore, the developments of the past years with the financial crisis and the EU enlargement to the east are taken into account, yielding additional findings and policy implications.

We argue that previous studies investigating the currency union effect in the early years of the Eurozone could be biased due to the boom in imports from other Eurozone members to the periphery countries. This boom, as we know today, was caused by consumption and housing bubbles that led to what is usually referred to as the European "debt crisis". Hence, trade effects for the early years, especially for final goods, could have been overestimated.

The rest of the chapter is structured as follows: Sect. 2 describes the empirical strategy and issues in the estimation. Section 3 presents the main results and finally, Sect. 4 concludes with a summary of the main findings and some policy implications.

2 Methodology

The empirical analysis is based on the standard gravity model of trade that was first developed by Tinbergen (1962). It is based on Newton's law of universal gravitation, according to which planets are mutually attracted in proportion to their physical mass and proximity. Transferred to the world of trade, physical mass is replaced with economic mass which is usually measured as GDP. Thus, trade between two countries is modelled as a function of their "economic mass" and the distance between them and has the following form:

$$X_{ij} = G A_i B_j \phi_{ij}, \tag{1}$$

where X_{ij} denotes the monetary value of exports from i to j, A_i and B_j denote all exporter all importer specific factors that influence production capacity in country i and demand in country j, respectively. G is a variable that does not depend on i or j such as the level of world trade liberalization. Finally, φ_{ij} represents the ease of exporter i to access to market j what is the inverse of bilateral trade costs.

2.1 Estimation Issues

The gravity equation of trade has seen numerous contributions and further developments in the past years. In particular, the work of Anderson and van Wincoop (2003) has been very influential. They show that for a well specified gravity equation trade

costs must be seen in relative terms to the rest of the world in order to model a country overall "resistance" to trade. This can be done by introducing "multilateral trade-resistance" (MTR) factors in the gravity equation. The basic idea is that two countries trade less with each other when they are surrounded by big economies than if they are surrounded by water, mountains or deserts and that the traditional gravity equation does not account for that.

We are dealing with MTR by introducing country-year dummies that control for time-varying exporter and importer effects. Because GDP variables for exporter and importer, usually employed in the gravity model to measure economic mass, vary only by year and country, we use the log of the cross-product of GDP instead.

Another important issue, especially when dealing with sectoral trade flows or trade between small countries, is the existence of zero trade flows. While previous trade theories were not capable of explaining the existence of zeros in trade flows, the monopolistic-competition model of heterogeneous firms developed by Melitz (2003) explains their existence with differences in productivity between firms. Helpman et al. (2008) specified a model that allows controlling for zero trade flows using a two-stage procedure. In the first stage, the extent of firms' entry into an export market (extensive margin), which is an unobserved variable in the standard gravity equation, is modelled using a Probit model given by:

$$\rho_{ij} = Pr\left(T_{ij} = 1\right) = \Theta\left(y_0 + \mathrm{v}_i + \sigma_j + \kappa VC_{ij} + \zeta FC_{ij}\right), \tag{2}$$

where the probability of exports from i to j depends on importer and exporter dummies v_i and σ_j, a bilateral variable denoting exporting costs, VC_{ij}, and fixed costs of entry denoted by FC_{ij}.

In the second stage the determinants of trade flows (in monetary value) are estimated using an augmented version of the gravity equation where tow elements of the first stage regression (the predicted value and the Inverse Mill's Ratio, IMR) are used to control for firm heterogeneity and sample selection bias. The later due to the existence of zero trade flows. Following this approach, the resulting equation then is:

$$X_{ij} = \beta_0 + \mathrm{I}_i + \mathrm{I}_j + \kappa VC_{ij} + \ln\left(e^{\delta\left(z_{ij}+\eta_{ij}\right)} - 1\right) + \beta_\eta \eta_{ij} + e_{ij}, \tag{3}$$

where X_{ij}, the dependent variable, denotes trade flows between i and j is and I_i and I_j denote the exporter and importer individual effects. The term in brackets is the share of firms that export to j, z is the predicted expensive margin obtained from the first stage and η_{ij} is the Inverse Mill's Ratio.

A requirement of this approach is an exclusion restriction: one variable that enters the first stage but not the second and that has no significant impact on the trade value, but on the probability to export.

Most authors choose a dummy whether or not to countries share the same religion as the excluded variable. In the European context, we do not consider this a good choice, as all countries share a Christian heritage and only some of their

trading partners differ in this respect.[3] In addition, the main religion of a specific country does usually not vary over time. Hence, in the case of the EU members, we assume the crossproduct of the time varying measures of corruption for exporter and importer to be an appropriate exclusion variable. The channel through which it affects trade is by rising insecurity and associated extra fixed costs for the exporting firm stemming from a higher probability of authorities or criminals trying to extort bribes in their homeland or export destination (Crozet et al. 2008).

For firms in countries with very low levels of corruption, this can be seen as a serious obstacle to start exporting as those countries are usually not used to this practices. But also positive effects for trade are conceivable: corrupt officials might allow firms to export or import even if their products do not meet technical, ethical, quality or safety standards. In overregulated countries this could lower fixed trade costs significantly (Rose-Ackerman 1999). Either way, by influencing fixed-costs rather than variable costs, corruption can be thought of as an additional barrier to trade, which should not have a significant impact on the value of trade once firms learn how to operate in a corrupt environment.

Corruption data are taken from the International Country Risk Guide (ICRG) published by the PRS Group and are a component of the Political Risk Dataset. It has a scale from zero (extremely high level of corruption) to six and assesses corruption within the political system.[4]

An additional possible source of estimation bias is the endogeneity of the decision to join a currency union. Frankel (2008) argues that endogeneity was not responsible for the extremely high estimates of early studies investigating currency union effects on trade. He presents estimates of similar magnitude to those found by a large number of early studies for the CFA zone, whose members have not decided to peg their currency to the Euro, but did so after France joined the Eurozone. The peg was not accompanied by other steps of integration that may have boosted trade and stands for an interesting natural experiment as the currency decision can be seen as exogenous.

In our case we assume endogeneity of the currency decision not to be a serious issue as past integration steps for Eurozone members are controlled for with a dummy for membership in the European Union (EU). Besides the common currency, Eurozone members have the same degree of trade facilitating integration as members of the EU. We assume that the decision to join the Eurozone is a political decision that is mostly driven by other factors different from those influencing the value of trade.

[3] Namely China, India and Turkey.

[4] In our dataset the crossproduct for both countries ranges from 2 to 36.

Table 1 Coverage. Non-EU members in italics

Countries				
Austria	Estonia	Ireland	Netherlands	Slovenia
Belgium	Finland	Italy	*Norway*	Spain
Bulgaria	France	*Japan*	Poland	Sweden
Cyprus	Germany	Latvia	Portugal	*Switzerland*
China	Greece	Lithuania	Romania	*Turkey*
Czech Republic	Hungary	Luxembourg	*Russia*	United Kingdom
Denmark	*India*	Malta	Slovakia	*USA*

2.2 Data

We have built a dataset with monthly bilateral trade for 35 countries over the period from January 1996 to December 2010. The countries included are the EU-27 countries and their mayor trading partners (Table 1).[5]

We use nominal monthly bilateral trade data disaggregated according to the BEC classification from Eurostat.[6] Unfortunately, Eurostat does not contain data on bilateral trade between two non-EU members; therefore data on trade flows between countries that have never been members of the EU is missing. The share of total EU-27 trade covered by our sample is well over 80 % (Table 3).

We assign the BEC sectors to three categories of goods, namely capital goods, intermediates and final goods, as recommended by the United Nations Department of Economic and Social Affairs (2007) (Table 2). The share for each BEC category in total trade for the whole sample is illustrated in Table 2 and the evolution over time of total trade for our three categories is shown in Fig. 1. Nominal GDP data are taken from the World Development Indicators database (WDI) at an annual level. To construct the bilateral exchange rates[7] and the volatility measure, we use daily nominal middle exchange rates reported by Datastream from the WM Company/Reuters.[8]

Different measures of exchange rate volatility have been proposed in the related literature. Most approaches have in common to measure the variance, but differ in the implementation. Examples are the standard deviation of the exchange rate or the moving standard deviation. Other measures, like ARCH and GARCH models, have gained popularity among researchers in recent years. In particular, A GARCH models the variance of the disturbance term for each period as a function of the

[5]Data for Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Luxembourg, Malta, Poland, Romania, Slovakia and Slovenia is missing for the years from 1996 to 1998.

[6]A thorough description of the BEC classification is available from the United Nations Department of Economic and Social Affairs (2007).

[7]The bilateral exchange rate measure is the average exchange rate of the past 6 months.

[8]This rate is the midpoint between the bid rate and the offered rate.

Table 2 BEC categories. Category 1 denotes capital, 2 intermediate and 3 consumption goods

BEC code	Description	Category	% of total trade
111	Food and beverages/primary/mainly for industry	2	0.57
112	Food and beverages/primary/mainly for household consumption	3	1.55
121	Food and beverages/processed/mainly for industry	2	0.45
122	Food and beverages/processed/mainly for household consumption	3	4.54
210	Industrial supplies n.e.s./primary	2	2.07
220	Industrial supplies n.e.s./processed	2	27.00
310	Fuels and lubricants/primary	2	3.61
321	Fuels and lubricants/processed/motor spirit	–	0.45
322	Fuels and lubricants/processed/other	2	2.63
410	Capital goods (except transport equipment)	1	14.32
420	Capital goods/parts and accessories	2	9.21
510	Transport equipment and parts and accessories thereof/passenger motor cars	–	6.68
521	Transport equipment and parts and accessories thereof/other/industrial	1	3.66
522	Transport equipment and parts and accessories thereof/other/non-industrial	3	0.43
530	Transport equipment and parts and accessories thereof/parts and accessor.	2	7.39
610	Consumer goods n.e.s./durable	3	3.32
620	Consumer goods n.e.s./semi-durable	3	4.93
630	Consumer goods n.e.s./non-durable	3	6.67
700	Goods not elsewhere specified	–	0.51

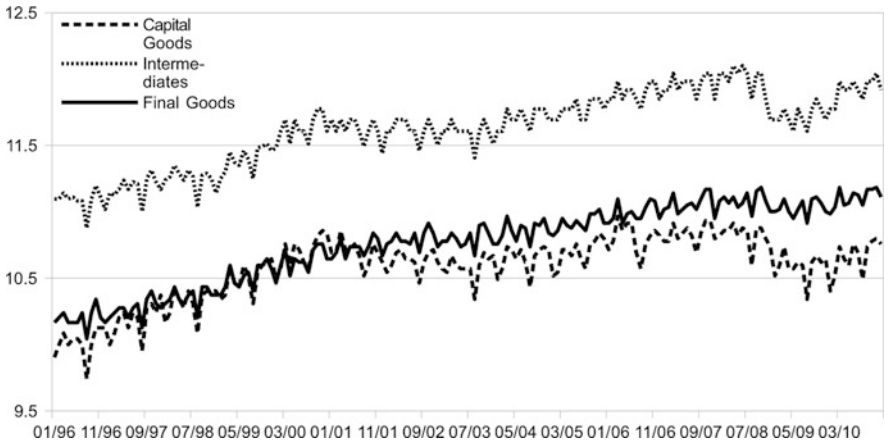

Fig. 1 Log of total trade value

Table 3 Share of total trade in the sample. Share is the average share of total trade value covered by our sample over all 16 BEC categories and 15 years for a single country

Eurozone		Other EU-Members		Non-EU Members	
Country	Share in %	Country	Share in %	Country	Share in %
Austria	87.07	Bulgaria	78.41	Switzerland	69.75
Belgium	91.62	Czech Republic	93.51	China	15.80
Cyprus	61.66	Denmark	83.75	India	19.60
Germany	80.58	Estonia	91.73	Japan	13.36
Spain	82.00	United Kingdom	82.91	Norway	72.39
Finland	85.16	Hungary	88.48	Russia	32.52
France	84.09	Lithuania	88.77	Turkey	49.57
Greece	71.01	Latvia	90.86	USA	19.04
Ireland	90.62	Poland	88.74		
Italy	80.78	Romania	85.09		
Luxemburg	96.18	Sweden	87.93		
Malta	64.49				
Netherlands	88.28				
Portugal	86.63				
Slovakia	93.69				
Slovenia	72.39				

errors in the previous periods. All measures have drawbacks, like for instance the high persistence of real exchange rate shocks when moving average representations are applied, or low correlation in volatility when ARCH/GARCH models are the measure of choice (Baum et al. 2004). The use of more sophisticated measures has however not changed the results significantly in the empirical literature on the impact of exchange rate volatility on trade (Ćorić and Pugh 2010).

Another important question is whether the volatility of the nominal or the real exchange rate or both should be included in the model. An advantage of the real exchange rate is, that it captures the true relative price of the good, however it also captures variation in the price levels, which is not desirable. Many studies use both exchange rates and compare the results. The differences they find are usually very small.[9]

Based on the recent literature, we have selected the standard deviation of the first difference of the logarithm of the nominal exchange rate, which has been used in various studies before (e.g. Clark et al. 2004):

$$Volatility_{ijt} = Std.dev.\left[\ln\left(e_{ijt,d}\right) - \ln\left(e_{ijt-1}\right)\right] d = 1 \ldots 130, \tag{4}$$

[9] A comprehensive comparison of the effects of real and nominal exchange rate volatility on exports was conducted by Cotter and Bredin (2011) finding that magnitude and direction are not changing, while timing effects can be different.

where $e_{ijt,d}$ denotes the daily bilateral exchange rate between countries i and j on business day t. This measure has the advantage of being equal to zero when the exchange rate follows a consistent trend, which apparently could be forecasted and consequently would not be a source of uncertainty.

To avoid bias from changes in price levels via spurious correlation, we use nominal exchange rates. The measure is constructed as a short-term volatility measure with bilateral exchange rates from the past 6 months. Departing from most previous studies, we construct the exchange rate volatility measure with daily exchange rates which allow more precise measures than "end of the month" values, as exchange rates sometimes tend to suffer more extreme movements at the end of each month. High persistence of exchange rate shocks is less of a problem as we only measure very short-term volatility of the past 6 months with high frequency data. In contrast to studies investigating long- or mid-run volatility, we investigate the effect of short term exchange rate volatility on trade by using a 6-month volatility measure. We assume that 6 months have 130 business days and thus construct the volatility measure accordingly.

3 Model Specification and Main Results

Estimations are conducted for three different categories of products: capital goods, intermediate goods and consumption goods. The idea is that these three groups differ significantly in terms of contracting patterns and that our variables of interest might affect trade flows in a different direction or/and to a different extent.

3.1 Model Specification

First, we are estimating FE and RE regressions with year-varying country fixed effects on the log of bilateral exports. Therefore, we are estimating the following equation:

$$\begin{aligned}\ln X_{ijkt} = {} & \beta_0 + \beta_1 \ln\left(Y_{it} * Y_{jt}\right) + \beta_2 \ln Distance_{ij} + \beta_3 EU_{ijt} + \beta_4 Euro_{ijt} \\ & \beta_5 Border_{ij} + \beta_6 Language_{ij} + \beta_7 Landlocked_{ij} + \beta_8 Island_{ij} + \beta_9 Colony_{ij} + \\ & \beta_{10} Volatility_{ijt} + \beta_{11} \ln ExRate_{ijt} + \beta_{12} Corruption_{ijt} + \kappa_k + \lambda_m + \alpha_{iy} + \nu_{jy} + \varepsilon_{ijkt},\end{aligned} \tag{5}$$

where the explained variable X_{ijkt} denotes nominal exports in sector k from the reporter country i to the partner country j at time t (month m in year y). The independent variable k controls for industry differences with dummy variables for each BEC category and λm controls for monthly seasonal effects with dummy variables for each month, m. The year-varying fixed effects α_{iy} and ν_{jy} are proxies for multilateral resistance factors.

Table 4 Variables

Variable	Description	Source
ln X_{ijkt}	Log of exports of good k from country i to j at time t in US$	Eurostat
ln Y_{ijt}	Log of the cross-product of nominal GDP of the countries i and j at time t in US$	CEPII
ln D_{istij}	Log of distance between capitals of country i and j in km	CEPII
EU_{ijt}	Dummy whether (1) or not (0) the countries i and j are both members of the EU at time t	CIA World Factbook 2011
$Euro_{ijt}$	Dummy whether (1) or not (0) the countries i and j have the Euro as a common currency and time t	CIA World Factbook 2011
$Border_{ij}$	Dummy whether (1) or not (0) the countries i and j share a common border	CEPII
$Language_{ij}$	Dummy whether (1) or not (0) the countries i and j share a common official language	CEPII
$Landlocked_{ij}$	Dummy whether none (0), one of the countries i and j (1), or both (2) are landlocked	CIA World Factbook 2011
$Island_{ij}$	Dummy whether none (0), one of the countries i and j (1), or both (2) are on an island	CIA World Factbook 2011
$Colony_{ij}$	Dummy whether (1) or not (0) the countries i and j ever had a colonial link	CEPII
$Volatility_{ijt}$	Bilateral volatility measure of the nominal exchange rate of the countries i and j at time t	WM Company/Reuters
ln Exch. $Rate_{ijt}$	Log of the bilateral nominal exchange rate of the countries i and j at time t	WM Company/Reuters
$Corruption_{ijt}$	Cross-product of the corruption measure of countries i and j at time t	Int. Country Risk Guide (ICRG)

The simultaneous inclusion of the measure of nominal exchange rate volatility and the dummy variable for mutual Euro membership allows us to capture convex effects as described by Baldwin (2006). Other variables are described in Table 4.

Next, in order take into account the existence of zero trade flows we are following the two stage approach proposed by Helpman et al. (2008). The first step estimation is a probit regression on the probability to export:

$$Pr\left(X_{ijkt}=1\right)=\Theta\left(\beta_0+\beta_1 \ln\left(Y_{it}*Y_{jt}\right)+\beta_2 \ln Distance_{ij}+\beta_3 EU_{ijt}+\beta_4 Euro_{ijt}+\beta_5 Border_{ij}+\beta_6 Language_{ij}+\beta_7 Landlocked_{ij}+\beta_8 Island_{ij}+\beta_9 Colony_{ij}+\beta_{10} Volatility_{ijt}+\beta_{11} \ln ExRate_{ijt}+\beta_{12} Corruption_{ijt}+\kappa_k+\lambda_m+\alpha_{iy}+\nu_{jy}+\varepsilon_{ijkt}\right), \tag{6}$$

The second step is estimated as a FE and RE regression including the linear prediction of exports down-weighted by its standard error (ZHAT) and the Inverse Mills Ratio (IMR). To fulfil the exclusion restriction, the variable corruption is excluded in the model:

$$\ln X_{ijkt} = \beta_0 + \beta_1 1\text{n}\left(Y_{it} * Y_{jt}\right) + \beta_2 \ln Distance_{ij} + \beta_3 EU_{ijt} + \beta_4 Euro_{ijt} + \beta_5 Border_{ij} + \beta_6 Language_{ij} + \beta_7 Landlocked_{ij} + \beta_8 Island_{ij} + \beta_9 Colony_{ij} + \beta_{10} Volatility_{ijt} + \beta_{11} \ln ExRate_{ijt} + \beta_{12} ZHAT + \beta_{13} IMR + \kappa_k + \lambda_m + \alpha_{iy} + \nu_{jy} + \varepsilon_{ijkt}. \quad (7)$$

3.2 Main Results

The extended gravity model is estimated for a sample of 35 countries over 15 years. Tables 5, 6, and 7 present the results for capital goods, intermediates and final goods separately.

The results show very robust negative effects of the volatility measure on trade for the current period and for all lags considered. While for final goods (Table 7) the effect is higher for the current value of the volatility variable than for the rest, for capital goods (Table 5) lags of the volatility variable present higher coefficients than the current value. Finally, for intermediates (Table 6) the current value and the first lag show higher elasticities than the second and third lags.

Mutual EU membership has a significant positive effect on the probability to trade and on the trade value. The coefficient of the EU dummy in the probit model is positive and statistically significant and range from 0.26 for capital goods and 0.34 for intermediates to 0.44 for final goods (Column 3 in Tables 5, 6, and 7). Thus, mutual EU membership increases the extensive margin of trade significantly, especially for industries producing final goods (Table 7).

The estimated coefficient for EU membership in the FE model, which should give an idea of the effect of EU membership on the intensive margin equals 0.09 for intermediate goods (Table 6, column 4), about 0.11 for capital goods (Table 5, column 4) and 0.2 for final goods (Table 7, column 4). Hence, the highest EU effect is found for final goods.

The results for the Euro effect are more ambiguous than for the EU effect, but still significant at conventional levels. The probability to trade is negatively affected with estimated coefficients around −0.25 for final goods and intermediates and −0.33 for capital goods. Trade value is affected negatively for capital goods with estimates around −0.10 and positively affected with estimates around 0.08 and 0.07 for intermediates and final goods, respectively.

In percentage points, the impact of the Euro on trade values lies, according to our estimations, at around 9 % for intermediates and 7 % for final goods. When not controlling for exchange rate volatility, the impact of the Euro on exports rises slightly (Tables 8, 9, and 10). For capital goods trade is around 11 % lower and around 8 % in the two stage approach.

Table 5 Regression results for capital goods

			1st Step	2nd Step	2nd Step
	FE	RE	Probit	FE	RE
ln GDP_{ijt}	0.447***	0.474***	0.462***	0.243***	0.467***
	−0.0395	−0.0128	−0.00222	−0.0402	−0.0127
ln $Distance_{ij}$	–	−1.206***	−0.861***	–	−1.206***
		−0.0543	−0.00653		−0.054
EU_{ijt}	0.115***	0.112***	0.260***	0.122***	0.118***
	−0.0161	−0.0161	−0.0103	−0.0161	−0.0161
$Euro_{ijt}$	−0.114***	−0.104***	−0.331***	−0.0708***	−0.0815***
	−0.0207	−0.0204	−0.0141	−0.0214	−0.021
$Border_{ij}$	–	0.606***	1.041***	–	0.659***
		−0.0976	−0.0337		−0.0984
$Language_{ij}$	–	0.194*	0.0408*	–	0.207*
		−0.112	−0.0243		−0.112
$Colony_{ij}$	–	0.539***	−0.138***	–	0.578***
		−0.118	−0.0281		−0.118
$Island_{ij}$	–	−0.698***	−0.247***	–	−0.723***
		−0.195	−0.00772		−0.194
$Landlocked_{ij}$	–	−1.989***	−0.0988***	–	−2.004***
		−0.251	−0.00744		−0.25
$Volatility_{ijt}$	−2.805***	−2.758***	−1.688***	−1.518***	−2.231***
	−0.514	−0.515	−0.368	−0.545	−0.542
L1.$Volatility_{ijt}$	−3.482***	−3.432***	−1.475***	−2.200***	−2.885***
	−0.612	−0.614	−0.363	−0.638	−0.636
L2.$Volatility_{ijt}$	−3.527***	−3.425***	−1.430***	−2.244***	−2.901***
	−0.509	−0.51	−0.358	−0.54	−0.537
L3.$Volatility_{ijt}$	−1.451***	−1.402***	−1.555***	−0.164	−0.735
	−0.531	−0.533	−0.364	−0.559	−0.556
ln $ExRate_{ijt}$	−0.301***	−0.296***	−0.00373	−0.299***	−0.295***
	−0.071	−0.0712	−0.0501	−0.0709	−0.0712
ln L1.$ExRate_{ijt}$	0.05	0.0578	−0.0924	0.05	0.0571
	−0.0634	−0.0636	−0.084	−0.0634	−0.0636
ln L2.$ExRate_{ijt}$	0.271***	0.240***	−0.0426	0.271***	0.241***
	−0.0734	−0.0732	−0.0816	−0.0734	−0.0732
ln L3.$ExRate_{ijt}$	0.0826	0.0751	0.0656	0.0804	0.0733
	−0.0547	−0.0549	−0.0471	−0.0547	−0.0548
$Corruption_{ijt}$	−0.0089***	−0.0088***	0.0156***	–	–
	−0.00308	−0.00304	−0.000661		
Zhat	–	–	–	0.0088***	0.0048***
			–	−0.00118	−0.00109
IMR	–	–	–	2.047***	5.926***
				−0.631	−0.509
Obs.	283,895	283,895	345,268	283,895	283,895
R2	0.194	0.697	–	0.194	0.698
RMSE	1.171	1.176	–	1.171	1.175

Reported R^2 is within R^2 for FE estimations and overall R^2 for RE estimations
Significance levels: *10 %, **5 % and ***1 %

Table 6 Regression results for intermediates

			1st Step	2nd Step	2nd Step
	FE	RE	Probit	FE	RE
ln GDP_{ijt}	0.682***	0.510***	0.390***	0.660***	0.516***
	−0.0179	−0.0105	−0.000915	−0.0143	−0.0105
ln $Distance_{ij}$	–	−1.544***	−0.708***	–	−1.562***
		−0.0489	−0.00278		−0.0488
EU_{ijt}	0.0896***	0.0873***	0.341***	0.0912***	0.0890***
	−0.00902	−0.00901	−0.00449	−0.009	−0.009
$Euro_{ijt}$	0.0942***	0.0894***	−0.257***	0.0785***	0.0705***
	−0.0116	−0.0115	−0.00613	−0.012	−0.0119
$Border_{ij}$	–	1.147***	1.096***	–	1.107***
		−0.0879	−0.0125		−0.0878
$Language_{ij}$	–	0.0904	0.117***	–	0.0731
		−0.102	−0.0102		−0.101
$Colony_{ij}$	–	0.284***	−0.137***	–	0.281***
		−0.106	−0.0115		−0.106
$Island_{ij}$	–	−0.570***	−0.193***	–	−0.582***
		−0.17	−0.00348		−0.17
$Landlocked_{ij}$	–	−2.071***	−0.243***	–	−2.092***
		−0.186	−0.00326		−0.186
$Volatility_{ijt}$	−2.435***	−2.416***	−1.144***	−2.927***	−3.003***
	−0.285	−0.285	−0.174	−0.299	−0.298
L1.$Volatility_{ijt}$	−2.560***	−2.522***	−0.868***	−3.033***	−3.090***
	−0.339	−0.339	−0.171	−0.351	−0.35
L2.$Volatility_{ijt}$	−1.865***	−1.833***	−0.792***	−2.334***	−2.397***
	−0.282	−0.282	−0.169	−0.296	−0.296
L3.$Volatility_{ijt}$	−0.617**	−0.594**	−2.426***	−1.086***	−1.153***
	−0.298	−0.298	−0.17	−0.312	−0.312
ln ExRateijt	−0.0911**	−0.0904**	0.0503**	−0.0911**	−0.0904**
	−0.0394	−0.0394	−0.023	−0.0394	−0.0394
ln L1.$ExRate_{ijt}$	0.0820**	0.0820**	−0.0324	0.0824**	0.0824**
	−0.0349	−0.0349	−0.0385	−0.0348	−0.0349
ln L2.$ExRate_{ijt}$	−0.0136	−0.0145	−0.0392	−0.0134	−0.0142
	−0.0409	−0.0409	−0.0374	−0.0409	−0.0409
ln L3.$ExRate_{ijt}$	0.0191	0.0195	0.00306	0.0186	0.0188
	−0.0301	−0.0301	−0.0216	−0.0301	−0.0301
$Corruption_{ijt}$	0.00117	0.0018	−0.0020***	–	–
	−0.00169	−0.00168	−0.000282		
Zhat	–	–	–	−0.0013***	−0.0016***
			–	−0.000338	−0.000329
IMR	–	–	–	2.545***	2.833***
				−0.182	−0.178
Obs.	1,045,992	1,045,992	1,381,072	1,045,992	1,045,992
R2	0.113	0.623	–	0.113	0.623
RMSE	1.243	1.244	–	1.243	1.244

Reported R2 is within R2 for FE estimations and overall R2 for RE estimations
Significance levels: *10 %, **5 % and ***1 %

Table 7 Regression results for final goods

			1st Step	2nd Step	2nd Step
	FE	RE	Probit	FE	RE
ln GDP_{ijt}	0.416***	0.449***	0.451***	0.234***	0.450***
	−0.0152	−0.00907	−0.00131	−0.0227	−0.00901
ln $Distance_{ij}$	–	−1.267***	−0.734***	–	−1.271***
		−0.042	−0.00381		−0.0417
EU_{ijt}	0.202***	0.201***	0.439***	0.202***	0.200***
	−0.0079	−0.00789	−0.00609	−0.00787	−0.00787
$Euro_{ijt}$	0.0648***	0.0681***	−0.241***	0.0741***	0.0718***
	−0.00999	−0.00993	−0.00865	−0.0104	−0.0103
$Border_{ij}$	–	0.857***	1.060***	–	0.865***
		−0.076	−0.0189		−0.0758
$Language_{ij}$	–	0.144*	−0.0883***	–	0.157*
		−0.0871	−0.0135		−0.0866
$Colony_{ij}$	–	0.338***	−0.129***	–	0.349***
		−0.0922	−0.0168		−0.0916
$Island_{jj}$	–	−0.244*	−0.0110**	–	−0.252*
		−0.147	−0.00456		−0.146
$Landlocked_{jj}$	–	−0.665***	−0.233***	–	−0.668***
		−0.161	−0.00424		−0.16
$Volatility_{ijt}$	−2.426***	−2.404***	−0.987***	−2.214***	−2.365***
	−0.234	−0.234	−0.2	−0.248	−0.247
L1.$Volatility_{ijt}$	−2.186***	−2.167***	−1.487***	−1.932***	−2.088***
	−0.282	−0.282	−0.199	−0.295	−0.294
L2.$Volatility_{ijt}$	−1.438***	−1.429***	−0.510***	−1.186***	−1.347***
	−0.235	−0.235	−0.197	−0.25	−0.249
L3.$Volatility_{ijt}$	−0.968***	−0.957***	−2.208***	−0.724***	−0.879***
	−0.247	−0.247	−0.199	−0.261	−0.26
ln $ExRate_{ijt}$	−0.276***	−0.276***	−0.268***	−0.274***	−0.275***
	−0.0347	−0.0347	−0.0297	−0.0347	−0.0347
ln L1.$ExRate_{ijt}$	−0.0592*	−0.0591*	−0.0931*	−0.0600*	−0.0599*
	−0.0311	−0.0312	−0.0499	−0.0311	−0.0312
ln L2.$ExRate_{ijt}$	0.0667*	0.0655*	−0.116**	0.0659*	0.0647*
	−0.036	−0.036	−0.0484	−0.036	−0.036
ln L3.$ExRate_{ijt}$	−0.0589**	−0.0589**	0.410***	−0.0602**	−0.0593**
	−0.0268	−0.0268	−0.0278	−0.0268	−0.0268
$Corruption_{ijt}$	0.0003	0.0006	0.0106***	–	–
	−0.00149	−0.00148	−0.00039		
Zhat	–	–	–	0.0014***	0.00071**
			–	−0.000362	−0.000353
IMR	–	–	–	5.923***	6.156***
				−0.417	−0.382
Obs.	879,509	879,509	1,035,804	879,509	879,509
R2	0.167	0.683	–	0.167	0.683
RMSE	1.006	1.007	–	1.006	1.007

Reported R2 is within R2 for FE estimations and overall R2 for RE estimations
Significance levels: *10 %, **5 % and ***1 %

Table 8 Fixed effects regressions for capital goods

	(1)	(2)	(3)	(4)	(5)	(6)
ln GDP_{ijt}	0.426***	0.504***	0.632***	0.575***	0.446***	0.457***
	−0.03	−0.0313	−0.0291	−0.0207	−0.0382	−0.0272
EU_{ijt}	0.119***	0.119***	0.138***	0.143***	0.113***	
	−0.016	−0.016	−0.0159	−0.0159	−0.016	
$Euro_{ijt}$	−0.112***	−0.112***	−0.100***	−0.0923***		−0.101***
	−0.0207	−0.0207	−0.0206	−0.0204		−0.0207
$Volatility_{ijt}$	−2.805***	−2.923***			−2.656***	−3.091***
	−0.514	−0.513			−0.513	−0.512
$L1.Volatility_{ijt}$	−3.485***	−3.520***			−3.286***	−3.879***
	−0.612	−0.612			−0.611	−0.61
$L2.Volatility_{ijt}$	−3.528***	−3.459***			−3.370***	−3.965***
	−0.509	−0.508			−0.508	−0.505
$L3.Volatility_{ijt}$	−1.468***	−1.455***			−1.323**	−1.505***
	−0.531	−0.531			−0.53	−0.531
ln $ExRate_{ijt}$	−0.301***		−0.332***		−0.302***	−0.300***
	−0.071		−0.0689		−0.071	−0.071
ln $L1.ExRate_{ijt}$	0.0499		0.0849		0.0499	0.0486
	−0.0634		−0.0612		−0.0634	−0.0634
ln $L2.ExRate_{ijt}$	0.271***		0.241***		0.272***	0.274***
	−0.0734		−0.073		−0.0734	−0.0734
ln $L3.ExRate_{ijt}$	0.0825		0.0933*		0.0823	0.0809
	−0.0547		−0.0544		−0.0547	−0.0547
Obs.	283,895	283,895	287,010	291,256	283,895	283,895
R2	0.194	0.194	0.198	0.2	0.194	0.194

Significance levels: * 10 %, ** 5 % and ***1 %

When the model is estimated excluding exchange rate movements, the estimated Euro effects remain almost unchanged, whereas the exchange rate volatility coefficients change only slightly.

All coefficients of the other variables included in the gravity equation yield the expected signs. Estimates are always significant and positive for the GDP cross product and negative and significant for the distance between capitals. Controls for contiguity always yield significant positive estimates and the coefficient of the variable island is negative and significant. While the control variable for common official language shows mixed results, former colonial ties have a negative impact on the probability to export, but a positive on the value exported (Table 11).

The excluded variable in the second stage, which is expected to have an impact only on the probability to trade, but not on the value, does a considerably good job. Our bilateral corruption measure has an insignificant impact on trade value and a significant impact on the probability. Only for capital goods, the impact on the value was statistically significant, but very low. When testing our results for robustness, we find that neither reducing the time period (Table 12, column 1–3), nor excluding

Table 9 Fixed effects regressions for intermediates

	(1)	(2)	(3)	(4)	(5)	(6)
ln GDP_{ijt}	0.506***	0.560***	0.473***	0.606***	0.493***	0.631***
	−0.0325	−0.0131	−0.0159	−0.0104	−0.0199	−0.0174
EU_{ijt}	0.0891***	0.0890***	0.101***	0.103***	0.0936***	
	−0.00898	−0.00898	−0.00888	−0.00887	−0.00897	
$Euro_{ijt}$	0.0939***	0.0939***	0.104***	0.100***		0.101***
	−0.0116	−0.0116	−0.0115	−0.0115		−0.0116
$Volatility_{ijt}$	−2.435***	−2.450***			−2.559***	−2.659***
	−0.285	−0.284			−0.284	−0.284
L1.$Volatility_{ijt}$	−2.560***	−2.534***			−2.722***	−2.866***
	−0.339	−0.338			−0.338	−0.337
L2.$Volatility_{ijt}$	−1.865***	−1.875***			−1.993***	−2.203***
	−0.282	−0.282			−0.282	−0.28
L3.$Volatility_{ijt}$	−0.615**	−0.613**			−0.738**	−0.648**
	−0.298	−0.298			−0.297	−0.298
ln $ExRate_{ijt}$	−0.0911**		−0.115***		−0.0910**	−0.0904**
	−0.0394		−0.0383		−0.0394	−0.0394
ln L1.$ExRate_{ijt}$	0.0820**		0.0892***		0.0820**	0.0809**
	−0.0349		−0.0337		−0.0349	−0.0349
ln L2.$ExRate_{ijt}$	−0.0136		−0.0217		−0.0139	−0.0122
	−0.0409		−0.0407		−0.0409	−0.0409
ln L3.$ExRate_{ijt}$	0.0191		0.0192		0.0195	0.0185
	−0.0301		−0.0299		−0.0301	−0.0301
Obs.	1,045,992	1,045,992	1,057,399	1,073,052	1,045,992	1,045,992
R2	0.113	0.113	0.116	0.119	0.113	0.113

Significance levels: * 10 %, ** 5 % and ***1 %

big non-European countries (Table 12, columns 4–6) from the sample significantly changes results for the EU or Euro dummies. Nevertheless, the volatility variables lose significance and their general impact is less clear-cut. This is probably due to the fact that exchange rates in the full sample are more volatility before 1999 for countries that later joined the Euro and in general between EU-27 and countries with more or less free floating exchange rates like U.S., Russia or India.

4 Conclusions and Policy Implications

In contrast to most previous studies, we do not find unambiguous results for exchange rate volatility. Instead, we find evidence for a significant negative impact, admittedly small in size. We find that sectors react differently with regard to the time frame and size of the impact. While mutual EU membership promotes trade via the extensive and intensive margin for most goods, Euro membership does so only via

Table 10 Fixed effects regressions for final goods

	(1)	(2)	(3)	(4)	(5)	(6)
ln GDP_{ijt}	0.417***	0.352***	0.639***	0.538***	0.439***	0.456***
	−0.0127	−0.014	−0.0134	−0.0132	−0.0109	−0.0128
EU_{ijt}	0.202***	0.202***	0.209***	0.209***	0.205***	
	−0.00787	−0.00787	−0.00778	−0.0078	−0.00785	
$Euro_{ijt}$	0.0648***	0.0647***	0.0705***	0.0722***		0.0813***
	−0.00999	−0.00999	−0.00996	−0.00992		−0.00997
$Volatility_{ijt}$	−2.426***	−2.492***			−2.502***	−2.896***
	−0.234	−0.234			−0.234	−0.233
$L1.Volatility_{ijt}$	−2.186***	−2.193***			−2.288***	−2.846***
	−0.282	−0.282			−0.282	−0.281
$L2.Volatility_{ijt}$	−1.438***	−1.420***			−1.518***	−2.189***
	−0.235	−0.235			−0.235	−0.233
$L3.Volatility_{ijt}$	−0.967***	−0.979***			−1.045***	−1.001***
	−0.247	−0.247			−0.247	−0.247
ln $ExRate_{ijt}$	−0.276***		−0.291***		−0.276***	−0.275***
	−0.0347		−0.0338		−0.0347	−0.0347
ln $L1.ExRate_{ijt}$	−0.0592*		−0.0582*		−0.0591*	−0.0615**
	−0.0311		−0.0301		−0.0311	−0.0311
ln $L2.ExRate_{ijt}$	0.0667*		0.0654*		0.0665*	0.0713**
	−0.036		−0.0359		−0.036	−0.036
ln $L3.ExRate_{ijt}$	−0.0589**		−0.0447*		−0.0587**	−0.0614**
	−0.0268		−0.0267		−0.0268	−0.0268
Obs.	879,509	879,509	889,410	902,978	879,509	879,509
R2	0.167	0.166	0.17	0.171	0.167	0.166

Significance levels: *10 %, **5 % and ***1 %

the intensive margin and not for capital goods. This could provide some evidence for a pronounced specialization process taking place in the Eurozone at the industry level after the introduction of the Euro, which results in countries exporting goods from a lower number of industries, but a higher overall value. According to our results, the effect is slightly stronger for intermediate than for final goods. However, further research is needed to confirm this statement.

The results for the extensive margin are in contrast to findings obtained by other authors (e.g. Bergin and Lin 2012; Baldwin and Di Nino 2006), who mostly found positive Euro effects on the extensive margin. This is probably due to the lower level of disaggregation and higher frequency of the trade data used in this study. Negative effects on a monthly level do not necessarily mean that positive effects on a yearly and product level are not possible. It would mean that while the Euro forced a strong specialization process with pronounced seasonality on the industry level, the number of different products traded between members of the Eurozone on a yearly basis has increased.

Table 11 Beta coefficients

	Capital goods				Intermediates				Final goods			
			2nd Step	2nd Step			2nd Step	2nd Step			2nd Step	2nd Step
	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE
ln GDP_{ijt}	0.364	0.385	0.198	0.38	0.492	0.368	0.476	0.372	0.346	0.374	0.195	0.374
ln $Distance_{ij}$		−0.374		−0.374		−0.429		−0.434		−0.401		−0.403
EU_{ijt}	0.019	0.019	0.02	0.02	0.013	0.013	0.013	0.013	0.034	0.034	0.034	0.034
$Euro_{ijt}$	−0.013	−0.012	−0.008	−0.009	0.01	0.009	0.008	0.007	0.008	0.008	0.009	0.008
$Border_{ij}$		0.062		0.067		0.106		0.103		0.088		0.088
$Language_{ij}$		0.015		0.016		0.006		0.005		0.011		0.012
$Colony_{ij}$		0.037		0.04		0.018		0.017		0.024		0.024
$Island_{ij}$		−0.107		−0.111		−0.078		−0.079		−0.039		−0.04
$Landlocked_{ij}$		−0.352		−0.355		−0.322		−0.326		−0.119		−0.119
$Volatility_{ijt}$	−0.007	−0.006	−0.004	−0.005	−0.005	−0.005	−0.006	−0.006	−0.006	−0.006	−0.006	−0.006
$L1.Volatility_{ijt}$	−0.008	−0.008	−0.005	−0.007	−0.005	−0.005	−0.006	−0.007	−0.006	−0.006	−0.005	−0.005
$L2.Volatility_{ijt}$	−0.009	−0.008	−0.005	−0.007	−0.004	−0.004	−0.005	−0.005	−0.004	−0.004	−0.003	−0.004
$L3.Volatility_{ijt}$	−0.004	−0.003	0	−0.002	−0.001	−0.001	−0.002	−0.003	−0.003	−0.003	−0.002	−0.002
ln $ExRate_{ijt}$	−0.306	−0.301	−0.304	−0.3	−0.083	−0.082	−0.082	−0.082	−0.289	−0.289	−0.287	−0.287
ln $L1.ExRate_{ijt}$	0.051	0.059	0.051	0.058	0.074	0.074	0.075	0.075	−0.062	−0.062	−0.063	−0.063
ln $L2.ExRate_{ijt}$	0.277	0.245	0.277	0.246	−0.012	−0.013	−0.012	−0.013	0.07	0.069	0.069	0.068
ln $L3.ExRate_{ijt}$	0.085	0.077	0.082	0.075	0.017	0.018	0.017	0.017	−0.062	−0.062	−0.064	−0.063
$Corruption_{ijt}$	−0.021	−0.02							0.001	0.002		
Zhat			0.028	0.015			−0.007	−0.009			0.007	0.004
IMR			0.005	0.014			0.012	0.014			0.009	0.009

Table 12 Robustness checks. RE regressions as in column (2) in Tables 5, 6 and 7

	(1)	(2)	(3)	(4)	(5)	(6)	n	(8)	(9)	(10)
	Sample 1999–2010			Without big four			Adding countries to sample (capital goods)			
Sectors	1	2	3	1	2	3	1	1	1	1
ln GDP_{ijt}	0.474***	0.529***	0.477***	0.448***	0.489***	0.390***	0.477***	0.461***	0.460***	0.459***
	−0.0134	−0.0109	−0.00954	−0.0112	−0.00964	−0.0084	−0.0121	−0.0119	−0.0119	−0.0129
ln $Distance_{ij}$	−1.201***	−1.549***	−1.265***	−1.266***	−1.575***	−1.315***	−1.246***	−1.222***	−1.217***	−1.297***
	−0.0554	−0.0499	−0.0437	−0.0482	−0.0473	−0.0415	−0.0488	−0.0498	−0.0502	−0.0524
EU_{ijt}	0.130***	0.0958***	0.204***	0.0564***	0.078***	0.136***	0.089***	0.119***	0.127***	0.129***
	−0.0162	−0.00902	−0.00797	−0.0165	−0.0098	−0.00846	−0.0165	−0.0163	−0.0161	−0.0163
$Euro_{ijt}$	−0.0761***	0.134***	0.107***	−0.103***	0.116***	0.068***	−0.104***	−0.103***	−0.105***	−0.0841***
	−0.0244	−0.014	−0.012	−0.0211	−0.0127	−0.0108	−0.0211	−0.0209	−0.0206	−0.0208
$Border_{ij}$	0.601***	1.153***	0.860***	0.539***	1.255***	0.816***	0.501***	0.581***	0.603***	0.492***
	−0.0998	−0.0897	−0.079	−0.0885	−0.0869	−0.0766	−0.0896	−0.0888	−0.09	−0.0946
$Language_{ij}$	0.193*	0.103	0.152*	0.304***	0.00207	0.179**	0.298***	0.249**	0.213**	0.322***
	−0.115	−0.104	−0.0906	−0.105	−0.104	−0.0908	−0.101	−0.102	−0.104	−0.108
$Colony_{ij}$	0.548***	0.290***	0.338***	0.172	0.247**	0.143	0.363***	0.543***	0.533***	0.819***
	−0.121	−0.109	−0.0959	−0.122	−0.12	−0.106	−0.113	−0.108	−0.109	−0.117
$Island_{ij}$	−0.815***	−0.578***	−0.156	0.149	0.104	1.210***	−0.912***	−1.034***	−1.027***	−1.293***
	−0.199	−0.172	−0.152	−0.186	−0.167	−0.147	−0.184	−0.185	−0.186	−0.193
$Landlocked_{ij}$	−1.510***	−2.275***	−1.116***	−1.247***	−1.575***	0.872***	−2.123***	−1.721***	−1.655***	−1.358***
	−0.264	−0.193	−0.169	−0.247	−0.191	−0.166	−0.248	−0.237	−0.235	−0.263
$Volatility_{ijt}$	−0.0637	−1.349***	−1.162***	−0.236	−1.024**	−0.968***	−0.25	−0.452	−0.329	−0.154
	−0.764	−0.424	−0.341	−0.769	−0.455	−0.355	−0.776	−0.738	−0.731	−0.748
$L1.Volatility_{ijt}$	−0.699	1.202**	−0.00509	−0.51	1.176**	−0.277	−0.561	−1.59	−1.709*	−1.744*
	−0.985	−0.55	−0.443	−0.999	−0.596	−0.465	−1.008	−0.97	−0.959	−0.987

(continued)

Table 12 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	n	(8)	(9)	(10)
	Sample 1999–2010			Without big four			Adding countries to sample (capital goods)			
Sectors	1	2	3	1	2	3	1	1	1	1
$L2.Volatility_{ijt}$	2.008**	0.0105	1.635***	1.626	−0.682	1.761***	1.814*	1.427	1.373	1.564
	−1.007	−0.56	−0.454	−1.04	−0.616	−0.482	−1.049	−0.996	−0.983	−1.016
$L3.Volatility_{ijt}$	−0.547	0.153	−0.225	−0.48	0.970**	0.436	−0.518	−0.807	−0.55	−0.555
	−0.753	−0.415	−0.339	−0.795	−0.464	−0.365	−0.801	−0.746	−0.737	−0.76
$\ln ExRate_{ijt}$	0.878	−0.408	−0.195	0.363	−0.524	−0.0683	0.93	0.279	0.155	−0.0256
	−0.582	−0.323	−0.286	−0.655	−0.386	−0.336	−0.633	−0.49	−0.471	−0.508
$\ln L1.ExRate_{ijt}$	−3.464***	0.645	−0.183	−2.301	0.699	−0.325	−2.837**	−1.863*	−1.632	−1.437
	−1.255	−0.694	−0.616	−1.445	−0.851	−0.739	−1.399	−1.08	−1.041	−1.116
$\ln L2.ExRate_{ijt}$	3.299***	0.00908	0.838	1.647	−0.133	0.535	1.439	1.226	1.3	1.531
	−1.179	−0.65	−0.579	−1.434	−0.846	−0.735	−1.392	−1.058	−1.022	−1.096
$\ln L3.ExRate_{ijt}$	−0.831*	−0.256	−0.802***	0.376	0.0121	−0.448	0.549	0.249	0.0475	−0.386
	−0.499	−0.274	−0.245	−0.64	−0.378	−0.329	−0.621	−0.466	−0.45	−0.485
$Corruption_{ijt}$	−0.0083**	0.00143	0.00143	0.0028	0.007***	0.012***	0.0023	−0.00394	−0.0058*	−0.0112***
	−0.00332	−0.00184	−0.00164	−0.00345	−0.00202	−0.00176	−0.00349	−0.00325	−0.00316	−0.00311
Year > 1998	Yes	Yes	Yes	–	–	–	–	–	–	–
USA	Yes	Yes	Yes	–	–	–	Yes	Yes	Yes	Yes
Russia	Yes	Yes	Yes	–-	–	–	–	Yes	Yes	Yes
Japan	Yes	Yes	Yes	–-	–	–	–	–	Yes	Yes
China	Yes	Yes	Yes	–	–	–	–	–	–	Yes
Turkey	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	–
Obs.	258,383	950,750	800,811	226,992	837,354	702,413	242,764	256,432	270,343	270,092
R2	0.694	0.619	0.68	0.719	0.649	0.704	0.715	0.709	0.705	0.708
RMSE	1.186	1.254	1.026	1.112	1.249	0.999	1.148	1.162	1.166	1.183

Significance levels: *10 %, **5 % and ***1 %

The introduction of controls for firm heterogeneity and sample selection bias does not change the results. Nonetheless, extensive and intensive margin are affected very differently by our variables of interest. When dropping most observations with higher volatility from the sample, the impact of exchange volatility on trade is less clear while the impact of mutual EU and Euro membership remains robust.

Policy implications stemming from our results are manifold. Policymakers should keep in mind, that currency unions come at great costs with regard to the flexibility of the domestic monetary policy and positive trade effects may be very limited and do not exist for all types of goods. The elimination of exchange rate volatility can also be achieved by a fixed peg. Although we find trade effects to be small, it still may be the best choice to avoid negative impacts as experienced currently in Eurozone and grants greater flexibility.

The question whether stabilizing the exchange rate is a desirable objective for policymakers is unclear and it is also unclear to which extent the real exchange rate is a variable that policymakers should be able to influence or actually can influence, besides establishing a currency union, a fixed peg or Dollarization (Eichengreen 2007; Rodrik 2008).

In the light of the current economic and political crisis in Europe, our results provide evidence that a common currency may reduce investments and thus trade in capital goods within the currency union. Together with other imbalances, like current account imbalances or real exchange rate misalignments, this may lead to a loss in competitiveness that can not be compensated by a devaluation of the domestic currency by a single member.

Acknowledgements Financial support received from the Ministry of Science and Innovation is gratefully acknowledged (ECO 2010-15863). The authors also would like to thank the participants of the conference Globalization: Strategies and Effects held in Koldingfjord, Denmark, as well as the editors and a referee for their helpful comments and suggestions.

References

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review, 93*(1), 170–192.

Auboin, M., & Ruta, M. (2011, October). *The relationship between exchange rates and international trade: A review of economic literature*. WTO Working Paper, 17.

Bahmani-Oskooee, M., & Hegerty, S. W. (2007). Exchange rate volatility and trade flows: A review article. *Journal of Economic Studies, 34*(3), 211–255.

Baldwin, R. (2006). *The Euro's trade effect*. ECB Working Paper Series, 594.

Baldwin, R., & Di Nino, V. (2006). *Euros and zeros: The common currency effect on trade in new goods*. NBER Working Paper Series, 12673.

Barkoulas, J. T., Baum, C. F., & Caglayan, M. (2002). Exchange rate effects on the volume and variability of trade flows. *Journal of International Money and Finance, 21*(4), 481–496.

Baum, C. F., Caglayan, M., & Ozkan, N. (2004). Nonlinear effects of exchange rate volatility on the volume of bilateral exports. *Journal of Applied Econometrics, 19*(1), 1–23.

Berger, H., & Nitsch, V. (2008). Zooming out: The trade effect of the euro in historical perspective. *Journal of International Money and Finance, 27*(8), 1244–1260.

Bergin, P. R., & Lin, C.-Y. (2012). The dynamic effects of currency union on trade. *Journal of International Economics, 87*, 191–204.

Boug, P., & Fagereng, A. (2010). Exchange rate volatility and export performance: A cointegrated VAR approach. *Applied Economics, 42*(7), 851–864.

Broll, U., & Eckwert, B. (1999). Exchange rate volatility and international trade. *Southern Economic Journal, 66*(1), 178–185.

Canzoneri, M. B., & Clark, P. B. (1984). *The effects of exchange rate variability on output and employment*. International Finance Discussion Papers, 240.

Chit, M. M., Rizov, M., & Willenbockel, D. (2010). Exchange rate volatility and exports: New empirical evidence from the emerging East Asian economies. *The World Economy, 33*(2), 239–263.

Clark, P. B. (1973). Uncertainty, exchange risk, and the level of international trade. *Economic Inquiry, 11*(3), 302–313.

Clark, P. B., et al. (2004). *Exchange rate volatility and trade flows—some new evidence*. International Monetary Fund Occasional Paper, 235.

Ćorić, B., & Pugh, G. (2010). The effects of exchange rate variability on international trade: A meta-regression analysis. *Applied Economics, 42*(20), 2631–2644.

Côté, A. (1994). *Exchange rate volatility and trade: A survey*. EconWPA, 9406001.

Cotter, J., & Bredin, D. (2011). *Real and nominal foreign exchange volatility effects on exports—the importance of timing*. Working papers 201001, Geary Institute, University College Dublin.

Crozet, M., Koenig, P., & Rebeyrol, V. (2008). *Exporting to insecure markets: A firm-level analysis*. Working Papers 2008-13, CEPII Research Center, Paris, France.

Cushman, D. O. (1983). The effects of real exchange rate risk on international trade. *Journal of International Economics, 15*(August), 45–63.

Cushman, D. O. (1986). Has exchange risk depressed international trade? The impact of third-country exchange risk. *Journal of International Money and Finance, 5*(3), 361–379.

De Grauwe, P. (1988). Exchange rate variability and the slowdown in growth of international trade. *IMF Staff Papers, 35*(1), 63–84.

Eichengreen, B. (2007). *The real exchange rate and economic growth*. Commission on Growth and Development Working Paper, 4, World Bank.

Eicher, T. S., & Henn, C. (2011). One money, one market: A revised benchmark. *Review of International Economics, 19*(3), 419–435.

Flam, H., & Nordström, H. (2007). *Explaining large euro effects on trade: The extensive margin and vertical specialization*. Manuscript, Institute for International Economic Studies, Stockholm University.

Frankel, J. (2008). *The estimated effects of the Euro on trade: Why are they below historical effects of monetary unions among smaller countries?* NBER Working Papers, 14542.

Frankel, J., & Rose, A. K. (2002). An estimate of the effect of common currencies on trade and income. *The Quarterly Journal of Economics, 117*(2), 437–466.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics, 123*(2), 441–487.

Hondroyiannis, G., et al. (2008). Some further evidence on exchange-rate volatility and exports. *Review of World Economics, 144*(1), 151–180.

Hooper, P., & Kohlhagen, S. W. (1978). The effect of exchange rate uncertainty on the prices and volume of international trade. *Journal of International Economics, 8*(4), 483–511.

Klein, M. W., & Shambaugh, J. C. (2006). Fixed exchange rates and trade. *Journal of International Economics, 70*(2), 359–383.

Makin, J. H. (1978). Portfolio theory and the problem of foreign exchange risk. *Journal of Finance, 33*(2), 517–534.

McKenzie, M. D. (1999). The impact of exchange rate volatility on international trade flows. *Journal of Economic Surveys, 13*(1), 71–106.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695–1725.

Ozturk, I. (2006). Exchange rate volatility and trade: A literature survey. *International Journal of Applied Econometrics and Quantitative Studies, 3*(1), 85–102.

Rodrik, D. (2008, Fall). The real exchange rate and economic growth. *Brookings Papers on Economic Activity*, *39*(2), 365–439.

Rose, A. K. (2000). One money, one market: Estimating the effect of common currencies on trade. *Economic Policy, 15*(30), 7–46.

Rose-Ackerman, S. (1999). *Corruption and government: causes, consequences, and reform*. Cambridge: Cambridge University Press.

Santos Silva, J. M. C., & Tenreyro, S. (2010). Currency unions in prospect and retrospect. *Annual Review of Economics, 2*(1), 51–74.

Tinbergen, J. (1962). *Shaping the world economy; Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

United Nations Department of Economic and Social Affairs. (2007). Future revision of the Classification by Broad Economic Categories (BEC).

Viaene, J.-M., & De Vries, C. G. (1992). International trade and exchange rate volatility. *European Economic Review, 36*(6), 1311–1321.

Sovereign Debt and Austerity in the Euro Area: A View from North America

Anne O. Krueger

Abstract A major issue that has arisen in the Euro area has been the appropriate policy response to a fiscal/sovereign debt crisis in a Euro zone member country. The debate has been particularly heated with respect to the responses of the troika—the International Monetary Fund, the European Union, and the European Central Bank—to the difficulties encountered by Greece in 2010 and subsequently. In this paper, I explore the relationship between "austerity" and other policies required by the troika, and the extension of financial support for a heavily indebted country. The essential argument is that, unless measures are taken to make sovereign debt sustainable over the medium or longer term, financial support for a country within a currency union cannot enable a resumption of growth. The necessary measures normally include fiscal adjustments which are politically unpopular. But the necessary "austerity" is less painful for a country than the abrupt adjustment that would be forced upon the country in the absence of external financial support. For those providing financial support, fiscal and other adjustments are a prerequisite for enabling the debtor to resume growth.

1 Introduction

Since World War I, there has been strong dissatisfaction with the exchange rate regime prevailing in the international financial system. The effort to return to the gold standard in the 1920s was costly, and formed a large part of the Keynesian dissatisfaction with that regime. Efforts after the Second World War to adopt a fixed but adjustable exchange rate system seemed reasonably successful, at least for industrial countries, for about two decades, until it was abandoned in the early

An earlier version of this paper was presented at a Conference in Honor of Ronald McKinnon in January 2015 at Stanford University

A.O. Krueger (✉)
School of Advanced International Studies, Johns Hopkins University, Washington, DC, USA

Stanford Center for International Development, Stanford University, Stanford, CA, USA
e-mail: annekrueger@jhu.edu

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_13

1970s. Some countries moved to floating exchange rates, but managed floats were more prevalent until the 1990s. Until then, most countries' regimes also included restrictions on capital flows, although they began to be relaxed as early as the late 1960s. As that happened, "capital account crises" became more frequent.

Against that background, the European economies were gradually integrating. Starting with the Coal and Steel Community and the Treaty of Rome, the European countries in the Common Market achieved closer and closer ties. During the same period, Europe also increasingly integrated with the rest of the world, as removal of non-tariff barriers and lowering of tariffs as negotiated under the General Agreement on Tariffs and Trade (GATT) stimulated trade throughout the world.

The European economic success was widely recognized. Long after it was clear that postwar reconstruction had been achieved, growth rates within the European Common Market, which later became the European Union (EU), remained well above earlier historical levels. This success is credited, among other things, with spurring the desire of Eastern European countries to reform their economies to join the EU and later integrative arrangements. Certainly, the Common Market countries became the biggest trading bloc in the world. By the time the EU was formed (as a step toward further integration), its trading volume with the rest of the world was second only to that of the United States.

Although there were crises from time to time over issues such as the common agricultural policy and the degree of subsidiarity, the political will to integrate seemed always to overcome the economic disagreements that arose. When the Euro was adopted as a common currency in 1999–2001,[1] it was intended as a further step in the process of integration. Although misgivings about a single currency among disparate countries were voiced, those supporting the introduction of a common currency carried the day by a large margin.

Only in 2010, when the Greek tragedy started to unfold, did it begin to be recognized that there might be deeper problems in the Euro area. It is not feasible in a single paper to discuss all of these problems. But as events have unfolded, a common theme in the press, domestic parliaments, and elsewhere has been criticism of the "austerity" imposed upon the Greeks in return for the large loans the government received from the three institutions known as the troika, the EU, the European Central Bank (ECB), and the International Monetary Fund (IMF).

It is the purpose of this paper to analyze the relationship between the "austerity" and the financial support received by Greece from the troika. Section 2 provides a brief sketch of international monetary history as it pertains to balance of payments and sovereign debt difficulties, and a short review of the evolution of economists' ideas regarding these phenomena. Section 3 focuses on some aspects of thinking regarding "an optimum currency area", as they provide a framework for analysis of the Greek problem, and the extent to which the Euro zone appeared to meet the criteria for such an area. Attention then turns to the Greek crisis, in Sect. 4,

[1]The Euro was first introduced as a currency in 1999, but the phase-out of national currencies was not completed until 2002.

starting with a snapshot of the situation as it existed early in 2010, and then asking a counterfactual question: What if there had been no financial support from the troika? Finally, Sect. 5 addresses the issues associated with the formulation of a reform program for a country such as Greece, confronting incipient difficulties, and draws some lessons.

2 Background Issues

Over many centuries, a number of countries' governments have been confronted with balance of payments problems. Some have faced circumstances in which debt-servicing obligations had built up to an extent that sovereigns could borrow no more and had to react as they could not service their debt. Others had borrowed to finance current account deficits (either because of monetary and fiscal excesses or because of deteriorating terms of trade) until they could no longer obtain credit. At that point it was recognized that domestic economic activity would drop (or was already dropping) unless additional sources of foreign exchange were forthcoming and/or domestic policies were altered.[2] Others recognized that their international payments position was so difficult that domestic economic activity would drop (or was dropping) unless additional sources of foreign exchange were forthcoming.

To be sure, there were circumstances in which countries simply could not continue debt servicing, as shown in Reinhart and Rogoff (2009), and debt was unsustainable. Until the mid-1940s, these crises were addressed on an ad hoc basis by the debtor government and its creditors. During that period (at least prior to the Great Depression), most industrial countries were on the gold standard, and when crises erupted, sovereigns adjusted primarily through tightened monetary and fiscal policy. Taussig (1921, pp. 464f.) famously showed that short-term borrowing and lending between countries in fact smoothed over most difficulties, although that was against the backdrop of shifts in monetary and fiscal policy.

After World War II, the IMF was established. As already noted, it was thought that the gold standard had unduly restricted governments' abilities to adjust monetary and fiscal policy to suit domestic needs, particularly in times of recession or prospective downward shifts in the level of economic activity.[3] According to its Articles of Agreement, the Fund was to promote international financial stability as a means of enabling the smooth flow of international trade. In order to do that, it was to oversee fixed-but-adjustable exchange rates, to lend for short-term shortfalls in foreign exchange and in support of financial stability, and also to provide financial support in cases where underlying macroeconomic imbalances had to be corrected.

[2] See Reinhart and Rogoff (2009) for a history of sovereign debt crises.

[3] John Maynard Keynes was foremost among those advocating an altered international monetary regime. He was a British delegate to the Bretton Woods conference which negotiated the Articles of Agreement of the International Monetary Fund. See Robbins et al. (1990) for an account.

The IMF became the entity usually approached by the authorities in countries where international financial difficulties were encountered.[4]

The delegates at Bretton Woods assumed that there would be a fixed exchange rate system, termed a "fixed but adjustable" system, with countries approaching the Fund when they ran into difficulties. Most countries adjusted their exchange rates through devaluation only when confronted with a strong need to do so.

Although the intent of the founders of Bretton Wood was that any temporary shortfall in foreign exchange (which they assumed would normally come about because of terms of trade changes or cyclical factors) should be financed, in practice it turned out that when shortfalls arose, macroeconomic policies (exchange rate, monetary, and fiscal policies) had generally been inconsistent with a sustainable balance of payments position and policy corrections were needed. Usually, it was anticipated that a period of 2–3 years, with exchange rate and monetary and fiscal adjustments, would suffice to enable the country to resume normal economic activity with a sustainable external position. Occasionally debt was unsustainable and had to be restructured.

In the initial postwar years, all but four countries in the world used exchange controls, so when difficulties occurred they were normally the result of current account deficits and inability to finance debt-servicing obligations or needed imports. Debt servicing obligations, in turn, consisted almost entirely of short-term trade credits (largely owed to private creditors) or long-term official obligations. Although some industrial countries approached the IMF because of balance of payments difficulties in the early postwar years, by the 1970s IMF support went almost entirely to developing countries (almost all of which were heavily reliant at that time on commodity exports).

In most cases, development efforts had been stretched with fiscal deficits and loose monetary policies, but it was normally a cyclical drop in commodity prices of major exports (oil for some countries, minerals or agricultural products for others) that triggered a crisis as current account deficits had been incurred even when the terms of trade were favorable.

In almost all cases, a fixed nominal exchange rate had been maintained for an extended period of time while the domestic price level was rising at a rate above that of inflation in other countries also pegged to the U.S. dollar. This came about largely because of expansionary fiscal and monetary policies. The real exchange rate had thus been appreciating and resources had been shifting toward the production of non-tradable goods and services while consumption had shifted toward imported goods, with increasingly large current account deficits.[5]

[4]It is important to remember that the short-term lending was supposed to be to tide countries over *temporary* shortfalls in their ability to finance their outstanding balances.

[5]In some developing countries, good fortune in the form of high commodity prices offset inflationary pressure and there was no need for adjustment, at least until commodity prices fell again. A few countries experienced balance of payments difficulties without prior excessively expansionary policies because of sharp and unexpected drops in their terms of trade. But in the majority of cases, prior expansionary policies were the major contributor to difficulties, and even

An approach to the IMF usually led to an agreed-upon (between the IMF and the country in difficulty) policy reform program to be carried out by the country and short-term (2–3 years) loans by the IMF to the country to enable imports (which had normally been severely restricted as the crisis approached) to expand and permit expansion of economic activity as the policy changes envisaged in the program took time to take hold. For later purposes, a continuing irony should be noted: short-term lending was needed to restart economies, which of course increased indebtedness, while too much indebtedness was part of the underlying problem.

The programs adopted by agreement between the IMF and the debtor country usually entailed tightening of fiscal and monetary policy and a devaluation of the exchange rate. In cases where there were large, identifiable bottlenecks to economic activity, the program called for a removal, or at least a reduction in the severity, of those bottlenecks.

As the postwar Bretton Woods system evolved, it quickly became a "gold-exchange standard" system, with the U.S. dollar as the key currency. Until the early 1970s, the system worked well enough to support global growth rates never previously seen except after wars or famine. After the early postwar years, most industrial countries had gradually relaxed and later abandoned capital controls. They continued to maintain fixed nominal exchange rates pegged to the dollar and macroeconomic policies under which they adjusted if there were changes in their international situations. The last industrial country with the support of the IMF to devalue in the last century was the United Kingdom in 1972.

As tariff and nontariff barriers fell under the auspices of the GATT, the entire industrial world's trade grew rapidly and, with it, international monetary and financial interdependence. Most developing countries, however, continued to rely on exchange controls under fixed or heavily managed exchange rate regimes for several decades longer.

For developing countries, exchange controls enabled them in the short run to pursue domestic macroeconomic policies often with resulting inflation rates that were not consistent with the real exchange rate that would have enabled sustained financing in the long run. Unless a fixed exchange-rate country with inflation rates above the world average (or with growth of demand for imports of goods and services well in excess of the supply of exports at an increasingly overvalued real exchange rate) experienced a favorable shock (a sharp increase in its terms of trade), the distortions to the economy associated with import shrinkage (as exchange controls limited imports) became larger. Either an inability to finance needed imports or to meet debt-servicing obligations eventually led to an approach to the IMF.

The two circumstances often merged as a developing country would borrow to sustain its development efforts as long as it could and would be confronted with the need for change only when markets closed to any further increase in sovereign

in cases where a drop in the prices of major exports was the proximate factor leading to difficulties, macroeconomic adjustment was called for.

debt. Until the sharp increase in private lending to developing countries in the late 1970s and 1980s, the scope for expanding capital inflows was relatively limited. Thereafter, as discussed below, current account deficits were sustained for extended periods in some countries, until private financing dried up.

The agreed-upon program usually included an adjustment to the exchange rate and some degree of tightening of fiscal and monetary policy. When deemed necessary to achieve the change in the current account needed and restore economic growth, other measures were also included, such as the relaxation or abandonment of exchange controls or increases in prices of goods sold by state-owned enterprises (SOEs), so that the government would no longer need to finance large losses.

IMF financing was always short-term. Economic policy reforms were intended to reduce excess demand (which had spilled into the current account deficit) through fiscal and exchange rate adjustment and simultaneously enable acceleration of growth. In cases where the reduction in excess demand and the acceleration of growth could not be expected to enable resumption (or continuation of voluntary debt service), debt restructuring was undertaken. Until the 1980s most debt of developing countries was owed to official creditors, and restructuring was undertaken through the Paris Club (the group of creditors who met under the auspices of the French Treasury). As private lending increased in importance, negotiations had to take place between the sovereign and its creditors.

By the 1990s, many developing countries had shifted to floating (or at least reasonably flexible) exchange rates and begun to follow less inflationary macroeconomic policies. As that happened, private capital flows, both short and long-term, to developing countries rose dramatically.

In the 1960s, a few developing countries starting with South Korea and Taiwan had opened up their economies, abandoned exchange controls, and incurred capital account deficits (largely by borrowing from private banks) to finance investments in excess of domestic savings. They had grown sufficiently rapidly and the rate of return on investment was sufficiently high so that debt servicing was not a problem. In the longer run, however, some other countries began experiencing difficulties in servicing their obligations incurred with capital account financing of deficits. Learning from the experience of those rapidly growing "Asian tigers", as they came to be called, private banks were willing to lend to other middle-income developing countries as their policy frameworks improved.

In the short run, the heightened availability of private capital flows meant that countries encountering temporary current account imbalances were much more readily able to finance them than had been the case in earlier years. These difficulties took two forms: On the one hand, in many cases, current account deficits reflected an expansionary monetary policy where the inflationary impact was offset by the influx of imports (predominantly for government or private consumption). On the other hand, some countries had relied on current account deficits to finance additional investments when domestic savings were below desired investment. When the additional investment had adequate real rates of return, economic growth was sufficient to meet financial obligations. But either when investments financed by foreign capital had been less productive than anticipated and debt-servicing

obligations could not be met or there was a "sudden stop" in capital inflows because foreigners were concerned about the country's ability to service debt or because global conditions changed (such as in 2008 with the financial crisis), a "capital account crisis" could threaten, or ensue, if appropriate measures were not taken.

Regardless of the cause of the difficulties, when capital inflows dropped sharply, major macroeconomic adjustments were called for. As in the case of earlier "balance of payments crises", the IMF was called upon to provide short-term financing during the adjustment process, and it did so once it was reasonably assured that the country would regain macroeconomic balance under the adjustment program.

As world economic growth continued, however, it became evident that different groups of countries were experiencing growth at very different rates. While there were many factors contributing to growth differentials, it came to be recognized that governments' "structural economic policies" could sometimes significantly repress economic growth.

Sometimes, those policies were closely related to macroeconomic imbalances and themselves contributed to them. Such items as subsidies (especially for agriculture and energy) constituted drains on budgets and crowded out more productive private investments. Sometimes SOEs, which were prevalent in some countries, experienced large financial losses that had to be covered out of general government revenues. Whether this was the result of inefficiency in the enterprise or price controls imposed by the government, the losses had to be addressed if fiscal balances were to be restored.

Some practices led to distortions that inhibited economic growth without an immediate fiscal impact. Over time, however, slower growth itself came to be seen as a central problem and when governments attempted to "stimulate" the economy, the removal or relaxation of the distortions was necessary to attain a more satisfactory growth rate. Labor market restrictions, credit rationing (to favored economic activities to avoid government subsidies), bureaucratic red tape, and a host of other measures were, in some instances, serious bottlenecks to economic activity. This was all the more the case when a country had accumulated large foreign debts: without a satisfactory growth rate, the challenge of reducing debt servicing burdens enough to restore creditworthiness and accelerate growth over the longer term was almost insurmountable.

Although the "tequila" crisis in Mexico at the end of 1994 was the first headline capital account crisis, it was the Asian crisis of 1998 that awakened the world to the problems associated with capital account crises. It was probably true that private creditors did not fully understand the differences between "emerging markets" such as Mexico and the Asian crisis countries. Some degree of over-optimism was partly responsible for lending. Either way, some emerging markets had been enabled to access private capital markets to finance underlying imbalances in the domestic macroeconomy.

One aspect of the early "capital account" crises should be noted: after the European crises of the early 1990s, as discussed below, a debate began over whether fixed but adjustable exchange rates were a realistic possibility. As markets had learned that a fixed exchange rate gave private investors the option of a one-sided

bet (as data would indicate whether there was accumulation or loss of reserves) on the exchange rate, it seemed likely that speculators could and would force exchange rate changes with losses to the Central Bank defending the currency. Even when countries were in surplus, inflationary pressures could, albeit more slowly, force an adjustment.

As more and more countries were moving toward policies of greater flexibility in their exchange rates,[6] European integration was proceeding apace. As barriers to transactions across national borders were reduced or eliminated, businesses found disruptions associated with exchange rate movements more and more difficult to address.

Gradually, attention turned to achieving fewer exchange rate movements between EU currencies. The ECU (European Currency Unit) had been adopted, but by the early 1990s led to severe disruptions in the foreign exchange markets. That led to increased support for a common currency, resulting in the creation of the Euro in the late 1990s, as noted above. As is well known, the Euro was regarded as successful even through the global financial crisis of 2007–08,[7] but that view was profoundly shaken by the events that followed, starting with the Greek difficulties.

There have, then, been three major foci of attention over the decades since World War II. First came the increased and more nuanced discussion of the fixed-but-adjustable exchange rate regimes. Next came the period of flexible exchange rate regimes, when trade and capital accounts were increasingly open. Finally, efforts to integrate led to a move toward exchange rate fixity among countries with close trading ties.

2.1 Evolving History of Thought

The history of thought has mirrored the three stages outlined above.[8] As seen earlier, when the IMF Articles were formulated, it was generally thought that the gold standard had unduly restricted domestic economic and fiscal policies. But the implications of the fixed-but-adjustable exchange rate, with the one-sided bet it gave private speculators when governments were defending their currencies, were not

[6]Few countries eschewed all intervention in the foreign exchange market, but many restricted their interventions to "smoothing" abrupt changes, or otherwise buying or selling foreign exchange in limited amounts (see Reinhart and Rogoff 2004).

[7]To be sure, there were some who were skeptics at the time the Euro was introduced, both in Europe and North America. Danish voters rejected the Euro in part because of doubts about its durability.

[8]This is not to say that there were no other important issues. These included questions such as the damage done by quantitative restrictions, the role of multinational corporations in trade, the determinants of trading patterns, and so on. However, none of these was/is central to analysis of the Greek difficulties.

fully recognized. During the first quarter century of the IMF, academic debate raged over whether exchange rates should be fixed or flexible.[9]

A second issue focused on the supply of international reserves. Given the rapid growth of international trade (more than 10 % annually during the 1950s and 1960s), it was clear that the demand for reserves was increasing at a more rapid rate than the rate at which additional gold would be mined. Countries increased their US dollar holdings as a share of reserves when reserves were rising rapidly along with trade volumes. Attention was therefore given to the problems of a key currency, which the U.S. dollar had clearly become.[10] It came to be recognized that a key currency country would inevitably have to incur current account deficits to enable other countries to obtain the foreign exchange reserves desired as economies grew. Policy makers also became convinced that an additional form of reserve was needed, and it was agreed that the IMF could create and issue Special Drawing Rights (SDRs). Many argued that, if exchange rates were flexible, the need for SDRs would vanish and key currency problems would largely disappear.

But for the second and third stages, another question was raised and analyzed. As early as the mid-1950s, Robert Mundell[11] turned around the question of what the exchange rate regime should be: what was the optimal currency area? This question focused debate on the issue along an entirely different line, to which attention turns below.

By 1973, issues associated with the one-sided bet, the current account deficits of the dollar as key currency, and perhaps with the intellectual arguments themselves, led to the abandonment of the fixed-but-adjustable exchange rate system, at least among the major industrial countries. To the surprise of many, flexible exchange rates did not obviate the need for foreign exchange reserves. In addition, a consensus formed around what came to be known as "the impossible trinity": Over the longer term, a country could not have a fixed exchange rate, choose its own domestic macroeconomic policies, AND have an open capital account. At least one of the three had to give. If a country adopted a fixed exchange rate regime and had an open capital account, its monetary and fiscal policy had, over the long run, to adapt to the needs of external balance. If, on the other hand, flexible exchange rates and an open capital account were adopted, then domestic fiscal and monetary policy could be adapted to meet the needs of the domestic economy.

From the impossible trinity,[12] it followed that if a country was confronted with a crisis, the proximate cause was that it had adopted inconsistent exchange rate, macroeconomic, and capital account policies. This important insight, which certainly holds over the long run, underlay a lot of thinking (and policies) regarding balance of payments and capital account crises.

[9] The classic paper was and is Friedman (1953).

[10] See Kenen (1983).

[11] Mundell (1961).

[12] See Obstfeld et al. (2005).

In many instances, the authorities did try to maintain a nominal exchange rate that resulted in a current account deficit financed by a capital inflow, often in the form of borrowing. As debts to the rest of the world built up, lenders began questioning the creditworthiness of the country and, if credit access was withdrawn or only offered at increasing spreads, either macroeconomic adjustment had to be undertaken or a financial crisis would ensue in the absence of large positive shocks.

But as the Asian and other financial crises were being addressed within this analytical framework, Europe was embarking on further integrative steps toward a common currency. As the EU had evolved, closer economic integration led to increasing discomfort when exchange rates between ECU members were adjusted.

Although discussion of a common currency had gone on for decades, it was only in the mid-1990s, after the European currency maelstrom of 1992, that a large number of EU members decided that individual currencies should be abandoned and a common currency, the Euro, adopted. Under the Maastricht Treaty, countries joining the Euro area would undertake to maintain their inflation rates at no more than 2.5 % above the lowest rate in the area, their sovereign debts at less than 60 % of GDP, and their fiscal deficits within 3 % of GDP.[13] These limits were set in recognition that countries adopting a common currency could not have vastly differing macroeconomic policies, although it was not clear exactly how these limits would be enforced[14] or whether they would be sufficient for the task of maintaining internal balance within the Euro area.

3 Criteria for an Optimum Currency Area

Mundell's (1961) contribution was to raise the question of the criteria for a currency area. He focused on the need for mobility of factors of production within a currency area (which presumably called for flexibility in factor prices), a mechanism for risk sharing (presumably through fiscal transfers), and some degree of synchronization of business cycles, perhaps by sharing the same commodity composition of trade. These continue to be recognized as desirable attributes of a currency union.

However, even prior to those desiderata, there was a necessary condition for currency union: fiscal and monetary policies could not diverge to any significant degree for any sustained periods of time. These issues were raised within the context of the Euro area by McKinnon (2002).[15]

Mundell, in his seminal paper, and McKinnon (1963), were the first to analyze the question of what\criteria should determine an optimum currency area. They

[13] See Issing (2008).

[14] There was supposed to be a financial penalty for exceeding the limits, but the first two Euro zone countries to violate them were France and Germany, and fines were not imposed.

[15] A few argued that entering into a currency union would prevent such disparities from arising and provide discipline for currency union members.

initially focused their analyses on the trading patterns that would be best for an optimum currency area. Later contributions on the trade side focused on the degree to which production patterns between partner countries differed and the degree to which production structures were diversified.

But when it was agreed that the Euro should become the common currency, macroeconomic criteria came sharply into focus. McKinnon (2002) provided a good overview of thought at about the time of the Euro's introduction. He listed three characteristics that made a common currency desirable: (1) fairly close trading ties so that there was considerable economic interdependence (preferably free trade, of course, but while that is preferable in general, something very close to it is necessary for a single currency area); (2) reasonably developed financial markets so that agents in each country could diversify risk by owning assets, especially equities, in foreign firms producing goods and services consumed domestically (and it was thought that this would be especially important for low-income countries, that were more specialized in production); and (3) sufficiently strong public finances in each member.

McKinnon regarded these criteria as being met by the countries in the Euro zone. Since there was relatively free trade within the EU, the first criterion was clearly met.[16] He pointed to the "explosive growth" in private Euro-denominated bonds issued after 1999 (the start of the Euro) as indicating that currency risk had greatly increased the attractiveness of cross-border holdings within the Euro area, and that this ability to reduce risk contributed to further benefits from membership in the Euro zone.

The third criterion was virtually unanimously agreed: public finances had to be under control before a common currency could be adopted. It would have been unreasonable, for example, to advocate a currency union between, say, Chile and either Argentina or Brazil when the Chilean inflation rate was near zero and that in Argentina and Brazil was in double or triple digits.[17] Like many others, McKinnon pointed to the Maastricht criteria as reason for believing that the third criterion was met. Strong public finances, in turn, implied to McKinnon a fairly stable price level for each member.

[16]Some had argued that countries specialized in the production and export of only a few commodities would not be fit candidates for a common currency and should tie to a currency such as the dollar. McKinnon rejected this argument, pointing out that citizens in a specialized country would have even greater need for financial integration so that they could reduce the risks associated with swings in the price of their exports by diversifying their financial portfolios.

[17]A few argued that entering into a currency union would prevent such disparities from arising and provide discipline for currency union members.

3.1 Lessons Learned

No one today would argue with the three criteria. But from experience, there are two additional criteria, to which less attention was paid at the time the Euro was launched or in its first few years of life. One was in interpretation of the criteria and one was in ignoring the political economy aspects of the criteria.

It is on these issues that I focus here. It will be recalled that the Maastricht treaty provided that signatories should: (1) prevent fiscal deficits in any year from exceeding 3 % of GDP[18]; (2) have an inflation rate that does not exceed the rate in the member country with the lowest inflation by more than 2.5 percentage points; and (3) prevent sovereign debt from rising above 60 % of GDP.

At first sight, the three limits would appear to inhibit any Euro zone country from letting its public finances become too weak for membership, as McKinnon thought. But there are several interrelated problems. A first is whether initial conditions for joining the Euro zone are appropriately set. A second is how countries present their fiscal accounts. A third is the need for structural fiscal balance. A fourth is reckoning what "counts" as part of the government's fiscal accounts. A fifth is the potential problems that could arise if a future government chose to break the rules. A sixth is the problem of enforcement. And a seventh is the possibility of misrepresentation of the fiscal accounts.

Turning to initial conditions, one needs only to think of German unification and the choice of a conversion rate for East German Deutschmarks to recognize that setting an inappropriate exchange rate upon entry into currency union (in this case between the two earlier Germanys) can result in major problems. In the East German case, it is estimated that the productivity (and per capita income) difference was greater than five to one. But the exchange rates were unified as if the differential were only two to one. In Germany's case, the unification of fiscal accounts and the political willingness to finance transfers to the former East Germany softened the adjustment challenge, but it was still very difficult.

In the formation of a customs union, the need for finding realistic currency conversion rates is clearly essential. On the whole, this appears not to have been much of a problem for the Euro zone countries, although Germany after Euro adoption initially encountered difficulties and then undertook sufficient structural reforms to result in German current account surpluses with the rest of the Euro zone.[19]

[18] As stated by Otmar Issing (see Issing 2008, pp. 192–96), who was a member of the preparatory committee for the Euro, members of the committee believed that a 3 % fiscal deficit would be sufficient to offset any downward shift in demand in any of the member countries. With hindsight, this had to be based on the view that there would be fiscal surpluses in good years.

[19] At the time of the launch of the Euro, there were concerns voiced in several countries, including Greece, that the rate of conversion from local currencies to Euros had been inappropriate and resulted in increases in domestic prices. Even if that was the case, the magnitude of the "mistake" was relatively very small compared to the later buildup of imbalances.

A second transition problem, and one that has frequently been noted, is that the borrowing costs for the weaker members of the Euro zone fell dramatically upon entry. While at first sight it would appear that this was a benefit, a problem was created: As debt was rolled over, with spreads over German bonds all but disappearing from their earlier levels of 300–400 basis points, parliaments were able to approve expenditure increases of several percentage points of GDP in the initial years after Euro adoption with no change in their prospective fiscal balances. The evidently low borrowing costs induced expenditure increases, which were later difficult to reverse, and fiscal deficits rose in years when there should have been structural surpluses.

Turning now to the Maastricht criteria, most economists would agree that a desirable fiscal stance is one in which there is "structural" balance. That is, when economic activity is below potential output so there is an output gap, fiscal deficits might be incurred but these would be offset by surpluses in years when the output gap was negative. The Maastricht treaty did not specify "structural" balance, which is defined as the rates of taxation and expenditure that would result in overall fiscal balance in a "normal" year.

Determining what is "normal" for fiscal balance is in itself a difficult job. But in addition, there was nothing about "structural" in the Maastricht Treaty or its implementation. A consequence of the Maastricht rules was that a country could budget for a fiscal deficit of 3 % and/or reach the 60 % debt ceiling in good times, but then find that deficit increases or debt increases (or both) were warranted in a downturn. Moreover, the proposed enforcement mechanism was weak and was in any event not implemented when France and Germany, the two largest Euro zone countries, both exceeded the 3 % limit in good times in the middle of the decade.

But things were worse than that! In the first place, policy-makers are usually optimists when estimating future revenue streams and pessimists when estimating government expenditures. And even those estimates are based on rosy views of the likely growth of economic activity over the budgetary period. And when new taxes are introduced, or expenditures cut, in an effort to reduce fiscal imbalances, again excessive optimism prevails.

Second, defining what is on-budget and what is off-budget is a matter for each country to decide for itself. Future pension liabilities are partly or largely off budget in many countries. State-owned enterprise budgets are often off-budget, but their deficits can loom large, and financing from the central government is frequently used to cover revenue shortfalls. In some instances, domestic banks are persuaded, cajoled, or required to buy government paper, and that, too, carries dangers.

Nonetheless, SOE accounts are usually separately recorded and SOE deficits then do not enter into calculation of central government deficits. In some countries, governments have passed rules under which all current expenditures must be financed out of current revenues, but investment may be financed by borrowing. The question of what is counted as investment then can become a major issue—Education? Health? Infrastructure maintenance?

Finally, another concern is that even in circumstances where the administration in power is carrying out "sound" macroeconomic policies, there is not necessarily a guarantee that future administrations will do the same.

4 Austerity and Lending

While the above-noted concerns are always there, one implication was not widely spelled out: if a country has an open capital account (as it virtually must in a currency union), monetary policy is determined outside the country's borders and the exchange rate is fixed, the only instruments left to achieve macroeconomic balance and growth in times of financial difficulties are reductions in the fiscal deficit and structural reform.

By the time of the Maastricht Treaty, the world had learned, over the preceding half century, that an appropriate response to current-account and sovereign debt difficulties was an IMF package in which the country agreed to fiscal and monetary tightening, exchange rate adjustment, and structural reforms, and received a short-term loan from the IMF to smooth the necessary adjustment.

In the Greek case, it was fiscal excesses that led to the crisis. The fiscal deficit had been large AND the fiscal accounts had been misreported! It is estimated that the Greek primary deficit—that is, the fiscal deficit before any interest payments—was over 15 % of GDP when in 2010 the new government announced a much greater magnitude of the prospective fiscal deficit than previously anticipated, Greece immediately lost market access. While the underlying fault was an unsustainably large fiscal deficit, the immediate problem was one of financing the government. To be sure, had the misrepresentation not been reported when it was, the Greek economy would nonetheless have been "found out" eventually, and deeper problems underlay the misreporting behavior.

When Greece could not borrow (and loans were needed not only to cover the current excess of government expenditures and owed interest payments over revenues but also to finance the rollovers of existing debt) the situation was dire because Greece was in the Euro zone, the government could not resort to the printing press. Cutting government expenditures sufficiently to be able to pay civil servants, pensions, and meet other expenses would have entailed draconian measures. Since the fiscal year was already under way, cuts of some 15 % or more in government expenditures or raising revenues for the rest of the fiscal year would have been necessary.

There is no way in the very short run to raise tax revenues sufficiently rapidly to offset any significant part of such a shortfall. One of the first rules for any government is to insure that its military and its civil servants are paid. Almost as important is insuring the continuing flow of pensions and other entitlements. This could not have been done without external financial support or debt restructuring. In discussions of austerity (in Greece and elsewhere in the Euro zone), it is generally forgotten that pain would have been far greater had financial support from the ECB, the EU, and the IMF not been forthcoming.

Once the troika began examining the Greek plight early in 2010 it rapidly became evident that Greece's economy had been sustained by rising fiscal deficits for the preceding decade (financed by borrowing at low interest rates because of Euro zone membership) and that the underlying economic structure was inimical to growth.

It was clear to most observers that major structural reforms were needed even with fiscal consolidation if Greece was to achieve a satisfactory rate of economic growth. Needed reforms included reducing the size of the bureaucracy, raising tax revenue by reducing the vast underreporting of income, and removing[20] a myriad of rules and regulations that encouraged firms to move abroad and discouraged new business startups.

While the Greek situation was an extreme case (although some governments have been enabled to be equally profligate at least for a period of time when not in a currency union by resorting to forcing the banking system to buy sovereign debt and enduring high and rising inflation), the essential point is that when a government is faced with major financing difficulties, two things are generally called for: (1) adjustments in economic policies to address the underlying causes of the fiscal difficulties; and (2) some degree of short-term financial support to reduce the pain during the inevitable adjustment process. In order to achieve a satisfactory longer-term prospect, structural reforms may be (and in the Greek case were and are) needed to facilitate growth.

In addition, large fiscal deficits are often a reflection of a government's efforts to sustain economic activity and even growth when the underlying structure of the economy is detrimental to growth. That was certainly the case in Greece.

Questions can certainly be raised about the speed and magnitude of the needed fiscal adjustment, whether the balance between increasing tax revenue and reducing expenditure was appropriate, whether the structural reforms were of appropriate form and duration, and so on.

In Greece's case, there were major delays in legislating and implementing the structural changes agreed upon with the troika. This is clearly reflected in the World Bank's Ease of Doing Business rankings, IMF delays in releasing tranches of loans, and other reports. Yet, without structural reforms of sufficient magnitude, Greek prospects for future economic growth could not have significantly improved.

[20]See the World Bank Group (2014) for data. In the 2008 Report, before the onset of the crisis, Greece was ranked 100th in the overall Ease of Doing Business index. In 2010, Greece was ranked 109th, just below Yemen. In 2014, the ranking was 65th, but some crucial aspects had deteriorated. For example, Greece was ranked 158th in enforcing contracts (it had been ranked 87th in 2008). Even in the ease of obtaining electricity for a new business, Greece was ranked 80th in 2014.

It can be argued, of course, that loan restructuring was also needed. Indeed, a restructuring was undertaken in 2012, although that still left debt at the end of 2014 at 177 % of GDP.[21] But the ability of any country to service its debt and restore an appropriate ratio of debt to GDP is crucially dependent on the rate of economic growth.[22]

A major difficulty for the IMF and the international community in confronting the problems of an overly indebted country is that financial support reduces the incentives for governments to undertake the necessary expenditure, revenue, and structural adjustments. To the extent financing is forthcoming from the international community, the urgency of adjustment is reduced.

That dilemma certainly has confronted the troika in its negotiations with Greece. It is widely reported that, in the Greek case, many of the badly needed measures (such as increasing enforcement of existing tax laws and cutting redundancies in government employment and state owned enterprises) either did not happen or were significantly watered down and delayed. As already noted, until 2014, the primary surplus in Greece was negative so that on net all interest payments were financed by additional borrowing.[23] Less austerity would have meant larger increases in debt to the official sector. Moreover, it can plausibly be argued that the appropriate policy response may well have been a much more drastic initial shift in the fiscal stance, with a larger "internal devaluation", as happened in Latvia's response to its crisis in 2008.

5 Conclusions

A first lesson from the Greek (and other) Euro zone experiences is clearly that the benefits achieved during years of fiscal excesses are smaller than the costs of correcting the resulting underlying imbalances when that posture can no longer be financed.

To lend to Greece without some assurances of meaningful structural reforms would have been a recipe for continuing and indeed worsening fiscal performance and related debt issues. Yet, achieving a means of doing this where the lenders were

[21] Most of the haircut on debt was borne by the private sector, so that about 85 % of remaining debt is to the official sector, including bilateral and multilateral (the EU, the ECB, and the IMF) creditors.

[22] It will be recalled that the rate at which the debt/GDP ratio changes is equal to the initial debt/GDP ratio times the interest rate minus the growth rate. The higher the growth rate, the more rapidly the debt/GDP ratio will decline for any given interest rate.

[23] Insofar as contagion was/is an issue, it centers primarily on the fact that the private banks in some Euro area countries (most notably France and Germany) held large quantities of Greek and other south European sovereign debt. Had Greece defaulted, it was and is clear that a considerable part of the equity of some of the banks in the creditor countries would have been significantly reduced. Their holdings were greatly reduced after 2010.

not blamed for the country's ills and did not use temporary bridge financing to avoid reforms is challenging.[24]

When a macroeconomic crisis does occur with high indebtedness, there is little choice other than to impose at least a measure of austerity, which is bound to be painful. The burden of proof falls on showing that the benefits from slower adjustment outweigh the costs of greater indebtedness. Those benefits would need to include a more rapid rate of economic growth at later dates in order to finance the additional debt so incurred. Evidence to date suggests that more rapid adjustment is, in most cases, a quicker path to restored growth. That said, there is a question as to the appropriate speed of adjustment. Arguing about the appropriate speed, rather than whether austerity was wrong, would be beneficial.

There are a large number of issues raised by the Euro zone experience that have not been dealt with here. These include the need for a banking union in a common currency area, political economy issues, questions about unsustainable debt, and the role of the international community in addressing crises. The need for banking union has clearly been recognized and is beyond the scope of this paper.

As to political economy, serious questions arise as to the ramifications of austerity policies in crisis countries. When cuts must be made in government expenditures or taxes raised, there is certain to be political resistance. The types of adjustment required are even more unpopular when it appears that it is foreigners who are imposing these measures.

Sustainability of debt is a serious issue. As already seen, it is estimated that Greece's sovereign debt is about 177 % of GDP at the time of writing (late 2014) and the average rate of interest on it is in excess of 5 %. It would require the nominal rate of GDP growth to be in excess of 8 % if Greece were to have any prospect of lowering the debt burden over the longer run. Finding an appropriate mechanism for debt restructuring without huge costly delays could benefit creditors and debtors alike.

These problems arise within the Euro zone. There are similar issues for individual countries confronting financial crises. But there is yet another issue for the Euro zone vis-à-vis the international community. The IMF was intended to relate to its individual countries and their needs. For it, the members of the Euro zone are neither wholly independent countries nor is the Euro zone only one country. This raises serious questions of appropriate policy for the international community. In the Greek case, officials in the EU and ECB initially intended to "go alone," but then discovered that they needed the core competence of the IMF—the ability to provide reasonably accurate fiscal estimates for existing policies and possible changes in them—in order to proceed. The IMF's normal practice is to lend to countries in need on the basis of an agreed upon adjustment program, releasing tranches of the

[24]Some assert that, even without fiscal imbalances, a crisis would have arisen because of the (largely policy-induced) structural rigidities in the Greek economy. That is not inconsistent with the view that the crisis itself was fiscal: It was fiscal excesses that prevented earlier falls in GDP.

loan on completion of a review to determine whether performance is sufficiently in conformity with what was agreed upon.

With the troika, other Euro zone members were not entirely a disinterested party, in significant part because of fears of "contagion." In the Euro zone crisis, as the IMF was called in as a junior partner in order to analyze the macro and microeconomic structure and recommend the necessary measures to restore sustainability, there were strong pressures for the IMF to accede to a program that was deemed inadequate to the task. There are reports that IMF staff (correctly, with hindsight) believed that debt was not sustainable prior to the initial "bailout". Yet, the European partners insisted upon support. It is an open question whether the Greek authorities would have followed the program more closely, with beneficial effects on Greek economic performance, had they not been aware of pressures on the IMF.

How the IMF can function, if it is left in the junior partner role, without the ability to withhold further financing if agreed upon conditions are not met, is a crucial question, and one to which little attention has yet been paid.

References

Friedman, M. (1953). The case for flexible exchange rates. In *Essays in positive economics* (pp. 157–203). Chicago: University of Chicago Press.

Issing, O. (2008). *The birth of the Euro*. Cambridge: Cambridge University Press.

Kenen, P. B. (1983). *The role of the dollar as an international currency* (Group of thirty occasional paper 13). New York: Consultative Group on International Economic and Monetary Affairs.

McKinnon, R. I. (1963). Optimum currency areas. *The American Economic Review, 53*, 717–725.

McKinnon, R. I. (2002). Optimum currency areas and the European experience. *Economics of Transition, 10*, 343–364.

Mundell, R. A. (1961). A theory of optimum currency areas. *American Economic Review, 51*, 657–665.

Obstfeld, M., Shambaugh, J. C., & Taylor, A. M. (2005). The Trilemma in history: Tradeoffs among exchange rates, monetary policies, and capital mobility. *Review of Economics and Statistics, 87*, 423–438.

Reinhart, C. M., & Rogoff, K. S. (2004). The modern history of exchange rate arrangements: A reinterpretation. *The Quarterly Journal of Economics, 119*, 1–48.

Reinhart, C. M., & Rogoff, K. S. (2009). *This time is different: Eight centuries of financial folly*. Princeton, NJ: Princeton University Press.

Robbins, L., Meade, J. E., Howson, S., & Moggridge, D. E. (1990). *The wartime diaries of Lionel Robbins and James Meade, 1943-45*. Basingstoke: Macmillan.

Taussig, F. W. (1921). *Principles of economics* (Vol. 1). New York: Macmillan.

World Bank Group. (2014). *Doing business: Measuring business regulation*. http://www.doingbusiness.org.

Part IV
International Integration and Migration

Possible Policy Responses to a Dark Side of the Integration of Regions and Nations

Oded Stark

Abstract In this chapter we study policy responses to an increase in post-merger distress. We consider the integration of regions and nations as a merger of populations which we view as a revision of social space, and we identify the effect of the merger on aggregate distress. The chapter is based on the premise that the merger of groups of people alters their social landscape and their comparators. Employing a specific measure of social distress that is based on the sensing of relative deprivation, a merger increases aggregate distress: the social distress of a merged population is greater than the sum of the social distress of the constituent populations when apart. In response, policies are enacted to ensure that aggregate distress and/or that of individuals does not rise after a merger. We consider two publicly-financed, cost-effective policies designed so as not to reduce individuals' incomes: a policy that reverses the negative effect of the merger on the aggregate level of relative deprivation, bringing it back to the sum of the pre-merger levels of aggregate relative deprivation of the two populations when apart; and a policy that is aimed at retaining the relative deprivation of each individual at most at its pre-merger level. These two policies are developed as algorithms. Numerical examples illustrate the application of the algorithms.

This chapter builds on and develops ideas presented in a lecture delivered at the Aarhus University Conference on "Globalization: Strategies and Effects," Koldingfjord, November 2011. The warm words and helpful comments of a referee, the advice and guidance of Carsten Kowalczyk, and the wise suggestions and able assistance of Marcin Jakubek are gratefully acknowledged.

O. Stark (✉)
ZEF, University of Bonn, Walter-Flex-Strasse 3, 53113 Bonn, Germany

University of Warsaw, Warsaw, Poland

Georgetown University, Washington, DC, USA
e-mail: ostark@uni-bonn.de

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_14

1 Introduction

We study policy responses to an increase in aggregate social distress brought about by the integration of regions and nations, which we view as a merger of populations and the revision of social space and the comparison set. Specifically, we look at the merger of populations as a merger of income vectors; we measure social distress by aggregate relative deprivation; and we maintain that (except in the special case in which the merged populations have identical income distributions) a merger increases aggregate relative deprivation. Given this increase, we assess how a budget-constrained policy-maker can reverse the increase by means of least-cost post-merger increases in individual incomes.

When populations merge, the social environment of the individuals who constitute the merged population changes: people who were previously outside the individuals' social domain are brought in. Mergers of populations occur in many spheres of life, at different times and places. They arise as a result of administrative considerations or naturally, they are imposed or chosen. Conquests bring hitherto disparate populations into one, provinces consolidate into regions, adjacent villages that experience population growth melt into one town, schools and school classes are joined, firms concentrate production from two plants in one, branches of a bank amalgamate, East Germany and West Germany become united Germany, European countries integrate. A typical driving force of integration is a presumption of economic gain.[1] With the help of specific examples, Stark (2010) and Stark et al. (2012b) raise the possibility that the revision of social space associated with the integration of societies can chip away at the sense of wellbeing of the societies involved. If integration brings in its wake social distress, then a compensating or higher economic gain is required to make integration desirable. Put differently, for economic integration to be worthwhile for the merged populations, the anticipated boost in productivity needs to be high enough to offset the strain on the individuals in the merged population.

In Section 2 we present measures of individual and aggregate relative deprivation and we claim that the aggregate relative deprivation of merged populations is larger than or equal to the sum of the pre-merger levels of the aggregate relative deprivation of the constituent populations (a superadditivity result). In Section 3 we study policy responses to the increase in post-merger discontent. Section 4 provides discussion and conclusions.

[1] For a discussion from a macroeconomic perspective of the benefits from the integration of nations and regions, see Rivera-Batiz and Romer (1991). For empirical assessments of the economic consequences of European integration see, for example, Henrekson et al. (1997), and Beckfield (2009).

2 A Measure of Deprivation and the Superadditivity of Aggregate Relative Deprivation with Respect to the Merger of Two Populations

We measure the distress of a population by the sum of the levels of distress experienced by the individuals who constitute the population. We refer to this sum as the aggregate relative deprivation (*ARD*) of the population. We measure the distress of an individual by the extra income units that others in the population have, we sum up these excesses, and we normalize by the size of the population. This approach tracks the seminal work of Runciman (1966) and its articulation by Yitzhaki (1979), and Hey and Lambert (1980); a detailed description is in Stark and Hyll (2011). In our definition of relative deprivation we resort to income-based comparisons, namely, an individual feels relatively deprived when others in his comparison group earn more than him. To concentrate on essentials, we assume that the comparison group of each individual consists of all members of his population.

Formally, for an ordered vector of incomes in population P of size n, $x = (x_1, ..., x_n)$, where $x_1 \leq x_2 \leq ... \leq x_n$, we define the relative deprivation of the i-th individual whose income is x_i, $i = 1, 2, ..., n$, as

$$RD(x_i, x) \equiv \frac{1}{n} \sum_{j=i}^{n} (x_j - x_i). \tag{1}$$

To ease the analysis that follows, an alternative representation of the relative deprivation measure is helpful.

Lemma 1 Let $F(x_i)$ be the fraction of those in the population P whose incomes are smaller than or equal to x_i. The relative deprivation of an individual earning x_i in population P with an income vector $x = (x_1, ..., x_n)$ is equal to the fraction of those whose incomes are higher than x_i times their mean excess income, namely,

$$RD(x_i, x) = [1 - F(x_i)] \cdot E(x - x_i \mid x > x_i). \tag{2}$$

Proof We multiply $\frac{1}{n}$ in Eq. (1) by the number of the individuals who earn more than x_i, and we divide $\sum_{j=i}^{n} (x_j - x_i)$ in Eq. (1) by this same number. We then obtain two ratios: the first is the fraction of the population who earn more than the individual, namely $[1 - F(x_i)]$; the second is the mean excess income, namely $E(x - x_i \mid x > x_i)$. □

The aggregate relative deprivation is the sum of the individual levels of relative deprivation

$$ARD(x) = \sum_{i=1}^{n} RD\,(x_i, x) = \sum_{i=1}^{n} \frac{\sum_{j=i}^{n} (x_j - x_i)}{n}. \tag{3}$$

ARD(x) is our index of the level of "distress" of population *P*. (For several usages of this measure in recent related work see Stark 2010; Fan and Stark 2011; Stark and Fan 2011; Stark and Hyll 2011; Stark et al. 2012a; Stark et al. 2012b.)

We now consider two populations, P_1 and P_2, with ordered income vectors $x^1 = \left(x_i^1\right)$ and $x^2 = \left(x_i^2\right)$ of dimensions n_1 and n_2, respectively. Total population size is $n = n_1 + n_2$. The ordered income vector of the merged population is denoted by $x^1 \circ x^2$, and is the n-dimensional income vector obtained by merging the two income vectors and ordering the resulting n components from the lowest to the highest.[2]

In the following claim we state that the difference $ARD\left(x^1 \circ x^2\right) - ARD\left(x^1\right) - ARD\left(x^2\right)$ is in fact non-negative: a merger increases aggregate relative deprivation or leaves it unchanged. Namely, if we conceptualize the merger of two income vectors as an addition operator, then *ARD* is a superadditive function of the income vectors. (A function H is superadditive if for all x, y it satisfies $H\,(x+y) - H(x) - H(y) \geq 0$.)

Claim 1 Let P_1 and P_2 be two populations with ordered income vectors x^1 and x^2, and let $x^1 \circ x^2$ be the ordered vector of merged incomes. Then

$$ARD\left(x^1 \circ x^2\right) - ARD\left(x^1\right) - ARD\left(x^2\right) \geq 0.$$

Proof A proof for the case of the merger of two populations with two members each is in Stark (2010); a proof for the case of the merger of two populations of any size is in Stark (2013). □

Example 1 Consider the merger of populations P_1 and P_2 with income vectors $x^1 = (1, 2)$ and $x^2 = (3, 4)$, respectively. The pre-merger levels of aggregate relative deprivation are $ARD\left(x^1\right) = 1/2$ and $ARD\left(x^2\right) = 1/2$. In the merged population with income vector $x^1 \circ x^2 = (1, 2, 3, 4)$, we have that $ARD\left(x^1 \circ x^2\right) = 5/2 > 1 = ARD\left(x^1\right) + ARD\left(x^2\right)$. This example vividly illustrates further why a formal proof of the superadditivity result is needed. Even in the simple case in which the two populations do not overlap and a relatively poor, two-person population $x^1 = (1, 2)$ merges with a relatively rich, two-person population $x^2 = (3, 4)$, the overall relative deprivation effect cannot be pre-ascertained. In such

[2]The operator ∘ is commutative and associative on the set of ordered vectors, and satisfies the closure property.

a case, it is quite clear that upon integration members of the poorer population are subjected to more relative deprivation, whereas members of the richer population other than the richest are subjected to less relative deprivation. Because one constituent population experiences an increase of its *ARD* while another experiences a decrease, whether the *ARD* of the merged population is higher than the sum of the *ARD*s of the constituent populations cannot be determined without formal analysis. Put differently, in a setting in which others could only bring negative externalities, a smaller population will always experience less aggregate relative deprivation. But in a setting such as ours when others joining in can confer both negative externalities (of 3 and 4 upon 1 and 2) and positive externalities (of 1 and 2 upon 3), it is impossible to determine without proof whether the expansion of a population will entail a reduction in aggregate relative deprivation or an increase.[3]

Because throughout we have kept incomes unchanged, the incomes of the members of a constituent population are not affected by its merger with another population: in our setting, a merger changes the social comparisons space that governs the sensing and calculation of relative income (relative deprivation), but it leaves absolute incomes intact. If we assume that individuals' wellbeing depends positively on absolute income and negatively on the experienced relative deprivation, a merger leads to a deterioration of the aggregate wellbeing of at least one of the merged populations.

We next ask how a government that is concerned about the increase of the aggregate level of social distress will be able to respond in a cost effective manner. Governments must be well aware that an increase in social distress could translate into social unrest, and there have been plenty of episodes, historical and current, to remind governments of the short distance between social distress and social protest.

3 Policy Responses to the Post-Merger Increase in *ARD*

The unwarranted repercussions of a merger on the wellbeing of populations and individuals invite design and assessment of policies aimed at counteracting the increase in *ARD* or in individuals' *RD*.

We study publicly-financed, cost-effective policies that are constrained not to reduce individuals' incomes. We consider two targets of governmental policy aimed at reversing the deleterious effect of merger:

[3]To see the variation in the externality repercussion even more starkly, note that when 3 joins 1 and 1, he confers a negative externality on the incumbents; when 3 joins 5 and 5 he confers neither a negative externality nor a positive externality on the incumbents; and when 3 joins 4 and 5, he confers a positive externality on incumbent 4.

Bringing down the aggregate level of relative deprivation to a level equal to the sum of the pre-merger levels of the aggregate relative deprivation in the two populations when apart; we refer to this problem as Π_1.

Seeing to it that no individual in the integrated population senses higher relative deprivation than the relative deprivation he sensed prior to the merger; we refer to this problem as Π_2.

Naturally, the government is keen to minimize the cost of implementing its chosen policy, which it enacts subject to the condition that in the process, no income is allowed to decrease.[4] Under the first policy, individual levels of relative deprivation may increase, decrease, or remain unchanged. Under the second policy, individual levels of relative deprivation cannot increase. This added constraint implies that the budgetary allocation needed to solve the second problem will be larger than the corresponding one needed to solve the first problem.

The cost of the solutions to these two problems can be interpreted as lower bounds on the additional income that the process of economic (income) growth has to yield in order to retain the aggregate relative deprivation or the individual levels of relative deprivation at their pre-merger levels.

3.1 Solving Problem Π_1

Clearly, the basic requirement of problem Π_1 can be satisfied by a trivial solution: lifting the incomes of all the individuals to the highest level of income in the merged population. In general, such a solution will not, however, be optimal.[5] It will be possible to achieve optimality by choosing carefully a subset of individuals for whom the marginal increase in incomes yields the highest marginal decrease in aggregate relative deprivation.

Consider the subset in the merged population of the individuals who earn the lowest income; we denote this subset by Ω. We now analyze what happens when marginally and by the same amount we increase the incomes of the individuals in Ω, where marginal increase refers to such an increase that the incomes of these

[4]We resort here to this last condition because of an implicit assumption that an individual's utility depends positively on his income and negatively on his relative deprivation. Because we do not know the exact rate of substitution between decrease in relative deprivation and decrease in income, we do not know how much income we could take away from an individual whose relative deprivation decreased in the wake of the merger. Therefore, to guarantee that the utility of an individual will not be decreased in the process, we impose the requirement that incomes cannot be lowered. Put differently, seeing to it that the individual's post-merger relative deprivation is not higher than his pre-merger relative deprivation while holding the individual's income constant constitutes a sufficient condition for retaining the individual's wellbeing at its pre-merger level.

[5]There are, however, specific cases where this solution is optimal such as when, for example, the merged populations consists each of one individual, with one individual earning less than the other.

individuals will not become higher than the income of any individual outside the set Ω.

First, suppose that the set Ω consists of just one individual out of the n members of the merged population, and that the government appropriates a sum ε to increase his income, where ε is small enough to satisfy our definition of a marginal increase in income. Using Eq. (2), this individual's relative deprivation decreases by $\frac{n-1}{n}\varepsilon$ because the mean excess income of the fraction of $\frac{n-1}{n}$ individuals earning more than him is reduced by the amount ε. At the same time, as this individual's income was, and continues to be, the lowest in the population, this disbursement does not increase the relative deprivation of any other individual and, therefore, the change in aggregate relative deprivation is

$$-\Delta ARD = \frac{n-1}{n}\varepsilon. \tag{4}$$

We next show that Eq. (4) is the highest marginal decrease in *ARD* achievable upon spending ε on a single individual. We do this by contradiction. Suppose that we were to increase by ε not the income of the lowest-earning individual but the income of an individual earning $x_i > x_1$. Then, the relative deprivation of this i individual will decrease as a result of his income getting closer to the incomes of the individuals earning more than him, but the relative deprivation of those individuals who earn less than him will increase. Namely, when $\overline{n}_i$ ($\tilde{n}_i$) is the number of individuals earning strictly more (less) than x_i, the marginal change in aggregate relative deprivation will be

$$-\Delta ARD = \frac{\overline{n}_i}{n}\varepsilon - \frac{\tilde{n}_i}{n}\varepsilon = \frac{\overline{n}_i - \tilde{n}_i}{n}\varepsilon, \tag{5}$$

because the mean excess income of the fraction of $\frac{\overline{n}_i}{n}$ individuals earning more than x_i falls by the amount ε, yet at the same time, the relative deprivation of each of the $\tilde{n}_i$ individuals earning less than x_i increases by $\frac{\varepsilon}{n}$. Because $\tilde{n}_i \geq 1$ and $\overline{n}_i < n$, comparing Eqs. (4) and (5) yields

$$\frac{\overline{n}_i - \tilde{n}_i}{n}\varepsilon < \frac{n-1}{n}\varepsilon.$$

Thus, channeling the transfer to an individual who is not the lowest income recipient in the population yields a lower decrease in aggregate relative deprivation than increasing the income of the individual who earns the lowest income.

Second, we allow the set Ω to expand to include more than one individual. We denote by $|\Omega|$ the size of the set Ω. Suppose again that the government appropriates the sum ε to increase the earnings of each member of the subset Ω by $\frac{\varepsilon}{|\Omega|}$. The fraction of the individuals who are earning more than members of the Ω set is equal to $\frac{n-|\Omega|}{n}$, and the mean excess income of these individuals falls by $\frac{\varepsilon}{|\Omega|}$. Therefore, each of the members of Ω will experience a decrease in relative deprivation equal to $\frac{n-|\Omega|}{n}\frac{\varepsilon}{|\Omega|}$. Again, because no individual experiences an increase in his relative

deprivation, this disbursement yields a change in aggregate relative deprivation

$$-\Delta ARD = |\Omega| \frac{n-|\Omega|}{n} \frac{\varepsilon}{|\Omega|} = \frac{n-|\Omega|}{n}\varepsilon.$$

As in the case of the set Ω consisting of a single individual, this is obviously the optimal use of ε for any subset of the merged population.

Drawing on the preceding protocol, we present the optimal solution to problem (policy response) Π_1 in the form of an algorithm as follows.

Algorithm A_1:

1. Include in the set Ω all the individuals who earn the lowest income in the merged population.
2. Proceed to increase simultaneously the incomes of the members of the set Ω, until either
   a. the aggregate relative deprivation is brought down to the pre-merger level or
   b. the incomes of the members of the set Ω reach the income of the first individual(s) who is (are) not a member (members) of this set, in which case start from step 1 once again.

It is easy to ascertain the optimality of Algorithm A_1: at each step, we increase the incomes of those individuals who earn the lowest, therefore the decrease in aggregate relative deprivation is the most effective, and the relative deprivation of no individual increases in the process. Heuristically, we start "pumping" incomes from the bottom, and we simultaneously gauge the aggregate relative deprivation response. The two processes move in tandem, and in opposite directions. The pumping from below is ratcheted up the hierarchy of the individuals, and it ceases when aggregate relative deprivation reaches the level at which it was prior to the merger.

Example 2 Consider the merger of populations P_1 and P_2 with income vectors $x^1 = (1, 2)$ and $x^2 = (3, 4)$, respectively. The pre-merger levels of aggregate relative deprivation are $ARD\left(x^1\right) = 1/2$ and $ARD\left(x^2\right) = 1/2$. Because in the merged population with income vector $x^1 \circ x^2 = (1, 2, 3, 4) \equiv z^0$ we have that $ARD\left(x^1 \circ x^2\right) = 5/2 > 1 = ARD\left(x^1\right) + ARD\left(x^2\right)$, the government seeks to lower the aggregate relative deprivation of the merged population back to $1/2 + 1/2 = 1$. Applying Algorithm A_1, we first include in the set Ω the individual earning 1, and we increase his income. Upon his income reaching the income of the next individual who earns 2, we obtain the income vector $z^1 = (2, 2, 3, 4)$, with $ARD\left(z^1\right) = 7/4$. Thus, giving the individual earning 1 an additional unit of income is insufficient to bring down aggregate relative deprivation to its pre-merger level. We therefore add the next individual (the individual whose pre-merger income was 2) to the set Ω, and we proceed to further increase the incomes of each of the two individuals who now constitute the set Ω and whose incomes are, for now, 2 each. At the point where these two incomes are elevated to 11/4 each,

we obtain $z^2 = (11/4, 11/4, 3, 4)$ with $ARD\left(z^2\right) = 1$. Thus, in order to bring the aggregate relative deprivation in the merged population to the sum of the pre-merger level, we have to transfer 7/4 units of income to the individual earning 1, and 3/4 units of income to the individual earning 2, which gives 10/4 as the total cost of implementing the policy.

3.2 Solving Problem Π_2

In order to solve problem Π_2, we first present a simple link between the levels of relative deprivation and the levels of income in a population.

Lemma 2 If an individual has the *i*-th highest income in a population, he has the *i*-th lowest level of relative deprivation in the population.[6]

Proof It is easy to see that individuals earning the highest level of income have zero relative deprivation, which is the lowest possible level, whereas the order of the other individuals in the relative deprivation hierarchy is obtained from the two relationships

$$RD\left(x_j, x\right) > RD\left(x_k, x\right) \text{ for } x_j < x_k$$

and

$$RD\left(x_j, x\right) = RD\left(x_k, x\right) \text{ for } x_j = x_k. \square$$

Lemma 2 tells us that the relative deprivation of an individual is inexorably related to his rank in the income hierarchy. The procedure of solving problem Π_2 builds on the simple fact that the hierarchy of the levels of relative deprivation mimics in reverse the hierarchy of incomes.

The following algorithm solves problem Π_2.

Algorithm A_2:

1. Starting with the post-merger income vector $x^1 \circ x^2$, we construct a vector *w* by arranging the elements of the $x^1 \circ x^2$ vector in descending order with respect to the pre-merger levels of relative deprivation. (If two or more individuals have the same pre-merger level of relative deprivation, we place leftmost the one with the lower income.)
2. We pick the individuals one at a time according to their placement in the *w* vector starting from the rightmost end and proceeding leftwards. If an individual has

[6]By *i*-th highest we mean an ordering that allows for (co-)sharing a position, that is, in a population with incomes (1, 2, 2, 3), the individual earning 3 has the 1st highest income, the individuals earning 2 have the 2nd highest incomes, and the individual earning 1 has the 3rd highest income.

higher relative deprivation than prior to the merger, we increase his income to the minimal level that brings down his relative deprivation to the pre-merger level. If the relative deprivation of an individual is the same as or is lower than prior to the merger, we do not raise his income.

To establish the rationale and optimality of Algorithm A_2, we implement 1 above by re-numbering the incomes $x_1, ..., x_n$ in a descending order, namely forming a vector $w = (w_n, ..., w_2, w_1)$, such that the leftmost individual earning w_n is the individual who had the highest pre-merger level of relative deprivation, and the rightmost individual earning w_1 is the individual whose pre-merger level of relative deprivation was the lowest.

The optimality of Algorithm A_2 hinges on the property that an individual's relative deprivation never increases as a result of changes made after his "turn" has come, given that we are proceeding leftwards in the w vector. To see this, we denote the vector of incomes after i steps, $1 \leq i < n$, with i incomes $w'_i, ..., w'_1$ being dealt with, as $w^i = (w_n, ..., w_{i+1}, w'_i, ..., w'_1)$. When we proceed then to the next income w_{i+1}, one of two possibilities arises.

First, the current relative deprivation of the individual with income w_{i+1}, $RD\left(w_{i+1}, w^i\right)$, can be lower or equal to the relative deprivation that he had prior to the merger; in such a circumstance, we do not increase his income. Therefore, the relative deprivations of other individuals, in particular those with incomes to the right of this individual, $w'_i, ..., w'_1$, do not increase. The second possibility is that the current relative deprivation of the individual with income w_{i+1}, $RD\left(w_{i+1}, w^i\right)$, is higher than his pre-merger relative deprivation. In such a circumstance, we increase his income to the level w'_{i+1}, which is the minimal income that equalizes the pre-merger relative deprivation and $RD\left(w'_{i+1}, w^i\right)$. We note that this change in income cannot affect the relative deprivation of those having incomes $w'_i, ..., w'_1$ because, according to Lemma 2, $w'_{i+1} \leq w'_j$ for $j = 1, ..., i$. It is a trivial feature of the relative deprivation index that the relative deprivation of an individual earning v does not increase when incomes that are lower than v are raised, as long as the raised incomes do not surpass v.

The preceding reasoning leads us to conclude that for $i = 1, ..., n$, the w'_i income is the lowest possible level of income which guarantees, first, that the relative deprivation of an individual will be no higher than prior to the merger and, second, that this individual's relative deprivation will not be affected by the process of adjusting the incomes of individuals to his left in the w vector whose incomes are $w_n, ..., w_{i+1}$. This protocol delivers the optimality of Algorithm A_2.

Heuristically, in order to address problem Π_2 we first raise the incomes at the top of the constructed hierarchy of the levels of relative deprivation; we do so in order to equate the levels of relative deprivations of the top-income individuals with the pre-merger levels of relative deprivation. Then, because the comparisons that yield relative deprivation are with incomes on the right in the income hierarchy, the changes made at the top determine by how much incomes that are further down the hierarchy have to be raised as we move leftwards.

Example 3 Consider the merger of populations P_1 and P_2 with income vectors $x^1 = (1, 2)$ and $x^2 = (3, 4)$, respectively. The pre-merger levels of relative deprivation are $RD\left(1, x^1\right) = 1/2$, $RD\left(2, x^1\right) = 0$, $RD\left(3, x^2\right) = 1/2$, and $RD\left(4, x^2\right) = 0$. Therefore, in the merged population with income vector $x^1 \circ x^2 = (1, 2, 3, 4)$, we have that the vector w, ordered according to the descending pre-merger levels of relative deprivation (with the lower of two incomes associated with the same level of relative deprivation placed leftmost) is $w^0 = (w_4, w_3, w_2, w_1) = (1, 3, 2, 4)$. We pick first "for treatment" the individual with income $w_1 = 4$. Noting that his relative deprivation was not increased as a result of the merger, $w_1' = 4$ and thus, $w^1 = (1, 3, 2, 4)$. Moving leftwards, we next attend to the individual with income $w_2 = 2$. Because $RD\left(2, w^1\right) = 3/4 > 0 = RD\left(2, x^1\right)$, we need to raise income w_2 to the level $w_2' = 4$, because then $RD\left(4, w^1\right) = 0 = RD\left(2, x^1\right)$. Consequently, we obtain $w^2 = (1, 3, 4, 4)$. Proceeding further leftwards to $w_3 = 3$, we have that $RD\left(3, w^2\right) = 1/2 = RD\left(3, x^2\right)$, and so no increase in income is needed in this case. Thus, we obtain $w^3 = (1, 3, 4, 4)$. Because for the remaining individual with income $w_4 = 1$ we have that $RD\left(1, w^3\right) = 2 > 1/2 = RD\left(1, x^1\right)$, we need to increase his income to $w_4' = 3$ as then, $RD\left(3, w^3\right) = 1/2 = RD\left(1, x^1\right)$. Thus, the final income vector is $w^4 = (3, 3, 4, 4)$, which gives 4 as the total cost of implementing the policy.

Pulling together the results of *Example* 2 and *Example* 3, we have:

Example 2 Income vector $z^2 = (11/4, 11/4, 3, 4)$, cost of implementation 2.5;

Example 3 Income vector $w^4 = (3, 3, 4, 4)$, cost of implementation 4.

Not surprisingly, because the constraint on implementing policy Π_2 is stricter than the constraint on implementing policy Π_1, enacting policy Π_2 is costlier.

4 Discussion and Conclusions

Processes and policies that integrate economic entities also revise the social landscape of the people who populate the entities. We have considered the case in which the form that the revision takes is an expansion – be it the result of closer proximity to others, more intensive social interactions, or reduced barriers to the flow of information. We have argued that a consequence of the changing social milieu is the casting of a shadow on the anticipated economic gains.

An increase in aggregate relative deprivation is a dark side to the integration of regions and nations. It puts a strain on the individuals in the merged population, casting a shadow over the production and trade (scale and scope) benefits anticipated from integration. An increase in relative deprivation can itself cause an adverse physiological reaction such as psychosomatic stress, and could lead to social unrest and a collective response in the form of public protest. To aid a social planner who seeks cost-effectively to counter this negative effect, we analyzed policy measures in a setting in which incomes are not allowed to fall. In this setting, the policy

measure to be adopted depends on whether the policy objective is to bring the aggregate level of relative deprivation down to the sum of the pre-merger levels, or to ensure that no individual experiences more relative deprivation than prior to the merger. We formulated algorithms to guide the implementation of these policy measures, and in illustrative examples we calculated the associated cost that the social planner would need to bear.

The analysis performed in this paper did not take into account all the possible effects of a merger. As already mentioned, the integration of regions and nations is expected to increase efficiency. When the possibility of a merger is contemplated, an interesting question to address would be whether the anticipated boost in productivity will suffice to pay for the cost of the policies discussed above.

In related work (Stark 2014) we consider the viability of a self-contained, non-publicly-financed policy aimed at preserving the wellbeing of individuals at its pre-merger level. We find, though, that a policy that seems to be attractive may not be implementable. In that context, we assume that the wellbeing of an individual is a function of his absolute income and of his relative deprivation, with the partial first derivatives being, respectively, positive and negative. Correspondingly, we define the preferences of the individuals in population P with an ordered income vector x as

$$u_i = u\left(x_i, x\right) = \alpha_i x_i - \left(1-\alpha_i\right) RD\left(x_i, x\right) \tag{6}$$

where $0 < \alpha_i < 1$, $i = 1, 2, ..., n$.

The underlying idea of this policy response is to skim off income from those who reap a gain as a consequence of the merger, and to distribute that income to those who experience a loss as a consequence of the merger, such that following the merger no individual will be worse off in terms of the utility measure defined in Eq. (6). There are several difficulties with such a scheme, however.

First, a necessary condition is that there has to be at least one gainer. But as is quite obvious, there may not be any as, for example, when population with income vector $x^1 = (1, 2, 3, 4)$ joins population with income vector $x^2 = (5, 5)$.

Second, for the policy to be applicable, the policy maker would need to know the α_i's. If each individual has his own distinct preference structure, the required information is colossal. Two possibilities then come to mind: that all the individuals share the same distaste for relative deprivation, or that they do not. We have attended in detail to the former possibility: $\alpha_i = \alpha \ \forall i, \ i = 1, 2, ..., n$.

That all the individuals share the same distaste for relative deprivation eases drastically the information requirements, allowing us to work with a single α. But then, even in the simplest configuration of incomes, impossibility strikes; Stark (2010) presents an example of this impossibility for the simple case of the merger of a one-person population with a two-person population. In Stark (2014) we state and prove that what this simple case reveals generalizes to an impossibility that pertains to the merger of any two populations with a uniform α. We concluded that implementation of a self-contained "tax and transfer" scheme aimed at retaining individuals' wellbeing at their pre-merger levels is not viable because there is not enough of a gain to placate the losers while still keeping the gainers at least as

well off as prior to the merger. A governmental infusion of funds is needed, or the efficiency gains to be had from integration need raise incomes sufficiently to facilitate an effective tax and transfer scheme.

Our analysis is essentially of the "comparative statics" type, with the revision of the social landscape occurring at the time of the merger, and the expected increase in incomes in the wake of the merger yet to come. Introducing dynamics need not erode our main argument, however. The revision of the comparison group could be gradual and coincide with the processes of scale economies and scope economies taking hold. Still, as long as the latter processes do not result in sufficient convergence of incomes,[7] the former process could still damage the post-merger sense of wellbeing.

References

Beckfield, J. (2009). Remapping inequality in Europe: The net effect of regional integration on total income inequality in the European Union. *International Journal of Comparative Sociology, 50*(5–6), 486–509.

Fan, C. S., & Stark, O. (2011). A theory of migration as a response to occupational stigma. *International Economic Review, 52*(2), 549–571.

Henrekson, M., Torstensson, J., & Torstensson, R. (1997). Growth effects of European integration. *European Economic Review, 41*(8), 1537–1557.

Hey, J. D., & Lambert, P. J. (1980). Relative deprivation and the Gini coefficient: Comment. *Quarterly Journal of Economics, 95*(3), 567–573.

Rivera-Batiz, L. A., & Romer, P. M. (1991). Economic integration and endogenous growth. *Quarterly Journal of Economics, 106*(2), 531–555.

Runciman, W. G. (1966). *Relative deprivation and social justice: A study of attitudes to social inequality in twentieth century Britain*. Berkeley, CA: University of California Press.

Stark, O. (2010). Looking at the integration of nations through the lens of the merger of populations: Preliminary superadditivity and impossibility results. *Swiss Journal of Economics and Statistics, 146*(4), 661–675.

Stark, O. (2013). Stressful integration. *European Economic Review, 63*, 1–9.

Stark, O. (2014). A policy response to a downside of the integration of economies: An impossibility theorem. *Economics Letters, 124*(2), 180–184.

Stark, O., & Fan, C. S. (2011). Migration for degrading work as an escape from humiliation. *Journal of Economic Behavior and Organization, 77*(3), 241–247.

Stark, O., & Hyll, W. (2011). On the economic architecture of the workplace: Repercussions of social comparisons among heterogeneous workers. *Journal of Labor Economics, 29*(2), 349–375.

Stark, O., Hyll, W., & Wang, Y. (2012a). Endogenous selection of comparison groups, human capital formation, and tax policy. *Economica, 79*(313), 62–75.

Stark, O., Rendl, F., & Jakubek, M. (2012b). The merger of populations, the incidence of marriages, and aggregate unhappiness. *Journal of Evolutionary Economics, 22*(2), 331–344.

Yitzhaki, S. (1979). Relative deprivation and the Gini coefficient. *Quarterly Journal of Economics, 93*(2), 321–324.

[7] In the case of *Example* 2, sufficiency stands for convergence of $z = (1, 2, 3, 4)$ to $z^2 = (11/4, 11/4, 3, 4)$.

Where Have All the Flowers Gone? The Locational Choice of High-Skilled Immigrants Within the EU-15 Countries: Some Empirical Evidence on the Roy-Borjas Model

Peter Nannestad

Abstract Increasing immigration into western countries over the last 25 years has re-kindled scholarly and political debates over the economic consequences of such immigration for the receiving countries. So far, research has come up with diverging empirical results. There seems, however, to be agreement on one point: the economic impact of immigration is strongly dependent on who the immigrants turn out to be. Thus, the skill-based sorting of immigrants into various host countries can be expected to be highly relevant. Borjas has proposed a sorting model in which the skill composition of migration depends on the second moments of the income distribution in the sending and the receiving countries, provided that skills and income are correlated in both. Previous empirical tests of this model have failed to turn up consistent results. Using data from the OECD Migration Database on the educational levels of immigrants from EU-15 countries into other EU-15 countries the paper shows that the skill-based sorting of this immigration is consistent with the core prediction from the Roy-Borjas model. This result also obtains after control for network (or herd) effects, which are often marshaled as alternative explanations of the locational decisions of immigrants.

1 Introduction

Increasing immigration into western countries over the last 25 years has re-kindled scholarly and political debates over the economic impacts of immigration on the receiving countries. So far, research has come up with diverging empirical results, which are hard to synthesize (Nannestad 2007). There seems, however, to be

Helpful comments by George J. Borjas, Arye Hillman, Dennis Snower, and Oded Stark on earlier versions of this paper are gratefully acknowledged. Remaining errors are the sole responsibility of the author.

P. Nannestad (✉)
Department of Political Science, Aarhus University, Universitetsparken, 8000 Aarhus C, Denmark
e-mail: pnannestad@ps.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_15

agreement on one point: the economic impact of immigration is strongly dependent on who the immigrants turn out to be. Given the characteristics of the economies and the labor markets of western countries, the human capital endowment of immigrants can be expected to be of crucial importance to their economic impact on the host country. Thus, the skill-based sorting of immigrants into various host countries becomes a highly relevant question.

1.1 The Roy-Borjas Model: Theory

A basic model of immigrant self-selection according to skill levels has been suggested by Borjas (1987; 1994: 1687–1692), building on Roy's model of how workers sort themselves among available employment opportunities (Roy 1951). The model assumes that the migration decision is determined solely by earnings opportunities in the source country, earnings opportunities in the host country, and migration costs. Immigrants will be positively self-selected with respect to skill levels—and have above-average earnings in both the source and the host country—if the correlation between the return to skills in both countries is high and if the dispersion in the earnings distribution is higher in the host than in the source country. Immigrants will be negatively self-selected with respect to skill levels—and have below-average earnings in both the source and the host country—if the correlation between the return to skills in the two countries is high and if the dispersion in the earnings distribution is higher in the source than in the host country (Borjas 1994: 1689).[1]

Thus, differences between source and host country with respect to the spread of their earning distributions give rise to selection biases in the skill distribution of immigrants (relative to the distribution of skill levels in the source country). The skill distribution of immigrants can be biased upwards as well as downwards. The size of the bias will, ceteris paribus, depend on the size of the difference between source and host country in the spread of their earning distributions. If there is no difference between the spread of the income distributions in the source and the host country, migration might still occur (depending on whether the first moments of the income distributions differ), but there should be no selection bias. In that case the immigrants should be expected to be a random sample of the populations in their countries of origin with respect to the distribution of skill levels.[2]

[1] Borjas also considers the case of a low correlation between skills in the source and the host country (Borjas 1994: 1689, eq. 15).

[2] This reading of the model differs somewhat from Chiswick's interpretation "... that immigrants from countries with greater inequality constitute the least able members of the origin labor force" (Chiswick 1999: 184).

1.2 The Roy-Borjas Model: Empirics

With respect to immigration into the US, Borjas (1987; 1990; 1994; 1998; 1999a; 1999b) has reported empirical evidence suggesting a process of negative self-selection among immigrants from source countries with a greater spread in their earnings distribution than the one found in the US. One indicator considered by Borjas is the declining skill level in the post-1965 immigrant cohorts. The declining level of skills in the immigrant population is what one would expect from negative self-selection, if the national origin mix of immigrants changes in favor of less developed countries where the dispersion in the earnings distribution will tend to be (even) higher than in the US. Such a change in the national origin mix of immigrants into the US did indeed occur as a consequence of the 1965 Immigration Act.

Even closer to the core of the Roy-Borjas model is the finding that measures of income inequality in the source country are negatively related to the earnings of male immigrants in the US (Borjas 1990, ch. 7; 1994, 1690–1691). Cobb-Clark (1993) has reported similar patterns for immigrant women. Thus, it would seem that immigrants into the US indeed tend to be negatively self-selected in a way consistent with the model.

Other studies, however, report more mixed empirical evidence or findings that contradict the Roy-Borjas model in various respects. Hatton and Williamson (2002) find that per capita income and education levels in source countries have significant effects on migration. They also find, in line with Borjas (1990) and Cobb-Clark (1993), that immigration into the US increases with the degree of income inequality in source countries. On the other hand, they find that factors like cultural affinity (language) and pure geographical distance are also important for immigration into the US. So is the size of the immigrant population from a particular source country already residing in the US, which indicates network or herding effects.

A number of studies apply the Roy-Borjas model to migration from Mexico into the US. Chiquiar and Hanson (2002) report results that do not seem to be consistent with the Roy-Borjas model. Since the dispersion of the Mexican earnings distribution exceeds the dispersion of earnings in the US, Mexican immigrants to the US should tend to be self-selected from the lower part of the Mexican skills distribution. But in fact Mexican immigrants turn out to be on average better educated than the average Mexican. Later research, e.g. Kaestner and Malamud (2010) and Moraga (2011), using different data, have come up with results that are largely compatible with the predictions from the Roy-Borjas model.

Empirical evidence for negative self-selection of immigrants into Western European countries with respect to skill levels also seems to be somewhat mixed. On the one hand, non-EU immigrants into Germany, Denmark, the Netherlands, Belgium, and France have been found to be considerably less well educated than EU-citizens in these countries (Brücker et al. 2002: 74–77). This could be interpreted as indicating that non-EU immigrants into these countries are primarily self-selected from the lower end of the skills distribution in their home countries, as predicted by the Roy-Borjas model. But as the study of Mexican immigration into the US

has shown, immigrants may well be selected from the higher end of the skills distribution in the source country and still be substantially less educated than the natives in the destination country (Chiquiar and Hanson 2002). On the other hand, non-EU immigrants into the UK, Greece, Spain and Portugal turn out to be better educated than the average native (Brücker et al. 2002: 74–77). Thus, based on education levels, there is no clear evidence of pervasive negative self-selection of immigrants into Western European welfare states.

In partial analogy with Zavodny (1999), the comparatively low educational levels of non-EU immigrants into Germany, Denmark, the Netherlands, Belgium, and France could also be explained by a combination of the characteristics of the original immigrant population in these countries and later network effects rather than by an ongoing process of negative self-selection with respect to skill levels. All five countries recruited immigrants under guest worker schemes up to 1973. Guest workers were mainly recruited for manual and "3-d" (dirty, dangerous, and demeaning) work. As a consequence, immigrants with comparatively low skill levels were recruited. Thus, it seems likely that the chain migration following the initial establishment of immigrant communities under guest worker schemes would likewise primarily consist of individuals from the lower end of the skill distribution. Thus, the low skill levels among non-EU immigrants in the five countries can be due to the initial selection of low-skilled immigrants by the host countries, perpetuated later through family reunification, rather than to the negative self-selection of immigrants.

Pedersen et al. (2004) have examined the determinants of immigration from 129 source countries into 27 OECD countries over the period 1990–2000. The study includes as predictors, among others, cultural and linguistic distance between source and destination country, networks, and a proxy for the generosity of the welfare regimes in the destination countries. They do not find clear evidence of negative self-selection of immigrants. A partial explanation for this lack of evidence could be the restrictive migration policies of many OECD countries since 1973. On the other hand, Mayda (2005) finds that inequality in the source and host economies is related to the size of emigration rates into 14 OECD countries over the period 1985–1995 "as predicted by Borjas' selection model".[3] Grogger and Hanson (2011), studying migration into 15 high income nations in the OECD, find that the higher the skill-related difference in earnings in the host country, the stronger is the tendency for migrants from a source country to be better educated relative to non-migrants. Again, however, it is not possible to disentangle the effects of positive self-selection from the effects of selective immigration policies, e.g. point systems favouring well-educated would-be immigrants.

Abramitzky (2009) tests the Roy-Borjas model using data on migration into and out of Israeli kibbutzim. Since the income distribution in the Kibbutz used to be

[3]It should be noted, however, that Mayda (2005) is testing the impact of source and host country inequality on the *size* of immigration rates, whereas the Roy-Borjas model concerns the impact of source and host country income inequality on the *skill composition* of migration.

much more equal than in the rest of Israel, one should expect people leaving the kibbutz to be disproportionally drawn from the upper end of the skill distribution, while people joining the kibbutz should be disproportionally drawn from the lower end of this distribution. The data confirm both predictions. However, since migration into a kibbutz is not free—there is an extensive screening process, and in the end the members of the kibbutz have to vote on whether or not to accept an applicant—part of these results may de due to (negative) selectivity in the admission process.

1.3 The Roy-Borjas Model: Another Empirical Test

The aim of this chapter is to test the Roy-Borjas model of skill-based self-selection of immigrants using data on the skill composition of the migration between all of the EU-15 countries. Focusing on migration between EU-15 countries appears advantageous in two respects.

First, the data at hand allow one to concentrate on the "proper" dependent variable in the Roy-Borjas model directly. This variable is the proportion of high- (or low-) skilled among the migrants moving from country A to country B relative to the proportion of high (or low) among the native population in the sending country A. Thus, it becomes possible to measure to what extent the immigrants from country A into country B are disproportionally drawn from the upper (or lower) tail of the skill distribution in the sending country (A) and to relate this measure of disproportionality to differences between country A and country B in the spread of their income distributions. Most previous empirical tests of the Roy-Borjas model have not been able to stay that close to the model's core variable.

Second, migration between EU-15 countries is far less restricted than migration between most other countries. While the "free movement of labor" within EU-15 is not in fact entirely free,[4] national immigration regimes have been abolished as far as citizens of EU-15 are concerned. Thus, with respect to immigration policies migration of EU-15 citizens between EU-15 countries takes place in a setting that approaches the setting of the Roy-Borjas model where immigration restrictions are absent. The "noise", caused by various and varying restrictive immigration policies, in the data used in most previous empirical tests of the Roy-Borjas model can be eliminated, at least to a considerable extent.[5]

Focusing on the locational choice of high-skilled immigrants within the EU-15 countries, the main research question in the present analysis then becomes *if the proportion of high-skilled immigrants from country A into the various receiving*

[4]The main restriction is that while citizens of an EU-15 country have the right to move to another EU-15 country and stay there for up to 6 months looking for work, they can only exercise that right if they are able to provide for themselves (and their families) in the host country. Also they can be repatriated if they lose their job in the host country and become dependent on public support.

[5]As will be discussed below, this type of "noise" cannot be eliminated completely, however. Also, the data used here may introduce other kinds of "noise" in the analysis.

countries relative to the proportion of high-skilled natives in country A is increasing in the ratio of the spread in the income distribution in the various receiving countries relative to country A, when we control for migration cost in the manner suggested by Borjas.[6] If this is indeed found to be the case, the second question becomes if this finding is robust to different specifications of "migration costs" taking into account factors like cultural and linguistic distance as suggested by the findings of Hatton and Williamson (2002) as well. The third question becomes if a positive finding is also robust when it is controlled for network or herd effects on the locational choice of high-skilled immigrants within the EU-15 countries.

2 Data, Variables and Models

2.1 Data

The data on the skill levels of inter-EU-15 migrants, as well as of the natives in the EU-15 countries, are taken from the OECD Migration Database (OECD 2005). This database on immigrants and expatriates in OECD countries is the first internationally comparable data set with detailed information on the skill distribution of the foreign-born population for almost all member countries of the OECD, among them the EU-15 countries. The database contains a file on the foreign-born population aged 15+ by educational attainment, country of birth and nationality (nationalized or not). The file also contains information on the educational attainment of the native-born population aged 15+. The data in the OECD Migration Database are primarily acquired from national population censuses or population registers. In two cases (Germany and the Netherlands) the data come from labor force surveys, however. The time reference of the data varies somewhat between countries (between 1999 and 2002).

Thus, the OECD Migration Database is cross-sectional. It contains data on the stocks of foreign-born in the EU-15 countries and their skill levels, but not on immigration flows between these countries and their skill composition.[7] While immigration flows and stocks must be ultimately related, testing the Roy-Borjas model on stock data introduces noise in (at least) two ways, which may bias the results.

In the first place, part of the immigration flows that have created the present stock of immigrants in the EU-15 countries may have occurred prior to the introduction of the free movement of labor and the abolishment of national immigration controls as

[6]Borjas uses geographical distance as a proxy for migration costs, assuming migration costs to be proportional to geographical distance. He also assumes migration costs to be the same for all migrants.

[7]Even with several such cross-sectional data sets, each referring to different points in time, it would not be possible to construct data on the skill composition of immigration flows.

far as nationals of other EU-15 countries are concerned. As suggested by Pedersen et al. (2004) immigration policies—especially if they favor immigration based on family ties, as was generally the case with immigration policies in the EU-15 countries prior to the implementation of free movement of labor—may counteract or obscure a skill-based self-selection of immigrants. Thus, even if the Roy-Borjas model were a perfect description of the processes that shaped the skill composition of migration flows between the EU-15 countries *after* the implementation of the free movement of labor principle, this might be difficult to detect when we examine the skill composition of the stock of immigrants in the EU-15 countries. Hence, a test of the Roy-Borjas model based on data from the OECD Migration Database will most likely be biased against the model.

In the second place, differences in the spread of the earning distributions in the sending and the receiving countries at a given point in time should, strictly speaking, be related to the skill composition of the migration flows at the same point in time. This is not possible with data on stocks. Relating differences in the spread of the earning distributions in the sending and the receiving countries at a given point in time to the skill composition of the stock of immigrants at that point in time will introduce noise into the relationship unless these differences are (approximately) stable over time.[8] While the differences between the spreads of the earnings distributions in the countries of EU-15 appear relatively stable over a reasonable time period, changes may still introduce some noise, which once more will bias a test of the Roy-Borjas model based on data from the OECD Migration Database against the model.

As a consequence, a rejection of the Roy-Borjas model based on data from the OECD Migration Database will not be very informative, as it can (too) easily be ascribed to noise in this data. Conversely, if the Roy-Borjas model is confirmed, this should be considered reasonably strong evidence in the model's favor, since the deck was stacked against such a result from the outset.

2.2 Variables

The dependent variable in the following analyses is the ratio of the proportion of high-skilled among the immigrants aged 15+ from country (i) to country (j) to the proportion of high-skilled aged 15+ in the native population in country (i) $[Highrat_{ij}]$. This variable expresses the extent to which immigrants from country (i) to country (j) are disproportionally drawn from the upper or the lower tail of the skill distribution in the sending country (i). Values close to 1 indicate that the proportion of high-skilled among the immigrants is approximately equal to their proportion in the native population. Values >1 or <1 indicate that the immigrants

[8]There is no equivalent problem in using the distance between countries as a proxy for migration costs, as this variable is time invariant.

are drawn disproportionally from the upper or from the lower tail, respectively, of the skill distribution in the country of origin. The data are taken from OECD (2005).

The central independent variable is operationalized as the ratio of the Gini coefficient in the receiving country (j) to the Gini coefficient in the sending country (i) *[Ginirat$_{ij}$]*. From the Roy-Borjas model we should expect a positive relationship between this ratio and the dependent variable. The data are taken from World Bank (2006). In keeping with standard practice in the empirical literature on the Roy-Borjas model, *Ginirat$_{ij}$* is considered exogenous to *Highrat$_{ij}$*.

Borjas uses geographical distance between the sending and the receiving country as a proxy for migration costs. Therefore geographical distance—measured as the distance (in 1000 km) between the capitals in the sending and the receiving country[9]—will be used as additional independent variable *[Dist$_{ij}$]*.

It might be argued, however, that using geographical distance as a proxy for migration costs entails a fairly narrow conception of migration costs. Hatton and Williamson (2002) have shown that cultural and linguistic distances may play a role in the migration decision as well, which can be interpreted as a second dimension of migration costs. To cover this dimension three dummy variables are included. The first one indicates if the sending and the receiving country share a common border *[Cborder$_{ij}$]*,[10] the intuition being that migration within border regions, e.g. from Copenhagen in Denmark to the Swedish city of Malmoe just 20 miles across the border, may entail few migration costs. The second dummy variable indicates if the language (or one of the languages) spoken in the sending country is also spoken in the receiving country *[Clanguage$_{ij}$]*,[11] the idea being that the absence of language barriers reduces migration costs. The third dummy variable indicates if the host country is English-speaking *[Elanguage$_{j}$]*.[12] Since most high-skilled can be assumed to know at least some English (of a kind), migrating to an English-speaking country should entail lower migration costs.

Several studies have shown network or herd effects to be important for the migration decision and for the immigrants' choice of destination country (Epstein and Gang 2004). As a proxy for network or herd effects the number of immigrants

[9]Calculated using the distance calculator at http://www.mapcrow.info.

[10]Though Denmark and Sweden are separated by the Oeresund, they have been coded as having a common border.

[11]Source: http://www.infoplease.com/ipa/A0855611.html. When we classify pairs of countries with respect to whether or not they have a common language some arbitrariness can be hard to avoid. Obviously Austria and Germany or Ireland and the United Kingdom have a common language by all reasonable standards. Belgium has a common language with both France and the Netherlands, as both French and Dutch (of some kind) are spoken in Belgium. Denmark has been coded to have a common language with Sweden, even though Danish and Swedish differ in quite some respects. Nevertheless they are close enough for most Danes and Swedes to communicate with each other using their respective mother tongue. On the other hand Denmark is coded not to have a common language with Germany, even though there is a small German-speaking minority in Denmark and even though most Danes know at least some German.

[12]Source: see preceding note.

Table 1 Descriptive statistics

Variable	Mean	SD	Min	Max
$Highrat_{ij}$	1.96	1.00	0.27	5.83
$Ginirat_{ij}$	1.02	0.19	0.67	1.50
$Dist_{ij}$	1.34	0.71	0.173	3.36
$Cborder_{ij}$	0.17	0.37	0	1
$Clanguage_{ij}$	0.07	0.25	0	1
$Elanguage_{j}$	0.13	0.37	0	1
$Immig_{ij}$	32.18	74.20	0.02	567.70
$Gdprat_{ij}$	1.05	0.37	0.36	2.76
$Hnative_{j}$	17.30	5.20	7.90	23.45

(in 1000) from country (i) living in country (j) will be used *[*$Immig_{ij}$*]*. The data are taken from OECD (2005).

Two further variables, intended to tap into the potential effect of earnings possibilities of (high-skilled) immigrants from country (i) to country (j) on the skill-based sorting of immigrants, will be used. The first one is the ratio of GDP per capita in the destination country (j) to GDP per capita in the sending country (i) *[*$Gdprat_{ij}$*]*.[13] If GDP per capita is taken as a proxy for average wages, this ratio should be found *not* to systematically influence the skill composition of the migration from country (i) to country (j). According to the Roy-Borjas model, differences in average earnings between the sending and the receiving country should affect the *size,* but not the *skill composition* of the migration between the two countries. The second variable in this group is the proportion of high-skilled individuals in the native population in the receiving country (j) *[*$Hnative_j$*]*. The higher this proportion the greater the employment opportunities of high-skilled individuals are expected to be in this particular country. The data are taken from OECD (2005).

Table 1 shows descriptive statistics for these variables (including 0–1 variables).

2.3 Models

The central relationship between the ratio of the proportion of high-skilled among the immigrants aged 15+ from country (i) to country (j) to the proportion of high-skilled aged 15+ in the native population in country (i) *[*$Highrat_{ij}$*]* and the ratio of the Gini coefficient in the receiving country (j) to the Gini coefficient in the sending country (i) *[*$Ginirat_{ij}$*]* will be estimated over pairs of countries ($i = 1, \ldots, 15$; $j = 1, \ldots, 15$; $i \neq j$) using weighted OLS. This implies testing for an affine (or linear, if the constant is omitted) relationship between these two variables. Strictly speaking, the Roy-Borjas model does not claim this relationship to be affine or linear. Any monotonic (non-decreasing) relationship should suffice. Thus, the

[13] Computed from Heston et al. (2002).

operationalization of the model tested here may be more restrictive than its theoretical formulation, which may introduce still another conservative bias into the results.

Four different versions of the model will be estimated separately:

$$\text{(I)}\quad Highrat_{ij} = a + b_1 Ginirat_{ij} + b_2 Dist_{ij} + e_{ij}$$

$$\text{(II)}\quad Highrat_{ij} = a + b_1 Ginirat_{ij} + b_2 Dist_{ij} + b_3 Cborder_{ij} + b_4 Clanguage_{ij} + b_5 Elanguage_j + e_{ij}$$

$$\text{(III)}\quad Highrat_{ij} = a + b_1 Ginirat_{ij} + b_2 Dist_{ij} + b_3 Cborder_{ij} + b_4 Clanguage_{ij} + b_5 Elanguage_j + b_6 Immig_{ij} + e_{ij}$$

$$\text{(IV)}\quad Highrat_{ij} = a + b_1 Ginirat_{ij} + b_2 Dist_{ij} + b_3 Cborder_{ij} + b_4 Clanguage_{ij} + b_5 Elanguage_j + b_6 Immig_{ij} + b_7 Gdprat_{ij} + b_8 Hnative_j + e_{ij}$$

Country-of-origin fixed effects are added in all models.

Model (I) is a "raw" version of the Roy-Borjas model with only the difference between the spread of the earnings distribution in the receiving and the sending countries and geographical distance as proxy for migration costs as predictors of the skill composition of migration. Model (II) contains an extended operationalization of migration costs, controlling also for the potential effects of cultural and linguistic distances on the locational choice of high-skilled immigrants in the EU-15. Model (III) adds a proxy in order to control for network or herd effects on the locational choices of high-skilled immigrants within EU-15. If found, such effects would seem to contradict the logic of the Roy-Borjas model. Model (IV) introduces additional controls for the potential effect of differences in the first moments of the earnings distribution on the skill composition of migration within EU-15 and for the effect of differences across receiving countries in the employment opportunities of high-skilled immigrants. According to the Roy-Borjas model differences between two countries in the first moment of their earnings distributions should not affect the skill-based composition of the migration between these two countries. It would hence strengthen the model if no effect can be found.

The data would thus seem to lend empirical support to the Roy-Borjas model if

(A) $b_1 > 0$ in all four versions of the model.
(B) $b_6 \leq 0$ in models (III) and (IV).
(C) $b_7 = 0$ in model (IV).

3 Results

Table 2 shows the results from fitting each of the four models separately. As can be seen, the core variable in the Roy-Borjas model—the difference between the spread of the earnings distribution in the receiving and the sending countries—is positive and significant in all four models, although only marginally so in Model II. This most probably reflects an omitted variable bias. As predicted by the Roy-Borjas model, the proportion of high-skilled among immigrants from a given

Table 2 Determinants of the locational choice of high-skilled immigrants within EU-15 (weighted OLS, unstandardized coefficients, robust standard errors in parentheses)

Dependent: *Highrat*	Model I	Model II	Model III	Model IV
Ginirat	3.427***	1.652*	3.051***	3.257***
	(0.821)	(0.880)	(0.813)	(0.716)
Dist	0.010	−0.161	−0.300*	−0.217
	(0.124)	(0.186)	(0.170)	(0.189)
Cborder		−0.269	−0.136	−0.135
		(0.217)	(0.175)	(0.168)
Clanguage		−0.247	0.198	−0.221
		(0.179)	(0.143)	(0.141)
Elanguage		0.756**	0.447	0.341
		(0.358)	(0.283)	(0.258)
Immig			−0.003***	−0.003***
			(0.001)	(0.000)
Gdprat				0.254
				(0.529)
Hnative				0.023**
				(0.011)
Constant	−0.767	1.385	0.194	−0.709
N	201	201	201	201
Root MSE	0.606	0.557	0.460	0.454
R^2 (adj.)	0.43	0.56	0.68	0.69

Fixed effects (country-of-origin dummies) included but not shown
***, **, *indicate significance levels of 1 %, 5 %, and 10 %, respectively

EU-15 country, relative to their proportion among the natives in this country, tends indeed to be highest in host countries within the EU-15 with the greatest spread in their income distribution relative to the spread in the income distribution in the immigrants' country of origin.

As a check on the sensibility of the estimated coefficients to $Ginirat_{ij}$ one can use them together with the other coefficients in the models (including the constant and the fixed effects) to estimate the expected value of $Highrat_{ij}$ if $Ginirat_{ij}$ is set equal to one and the remaining variables are fixed at their means, cf. Table 1. Setting $Ginirat_{ij}$ equal to 1 means that the income equality is the same in the sending and the receiving countries. For that case the Roy-Borjas model predicts that the ratio of the proportion of high-skilled among the emigrants to the proportion of high-skilled in the sending countries should not be significantly different than 1. When we use the estimated coefficients from Model I, the average expected value of $Highrat_{ij}$ is found to be 1.28. While this value is slightly bigger than 1 it deviates from that value by less than one half of the standard deviation of the expected values of $Highrat_{ij}$ (0.56) and would hence appear to be sufficiently close to the value predicted from the Roy-Borjas model.

Geographical, cultural and linguistic distances between sending and receiving countries do not show a robust effect on the locational sorting of high-skilled

immigrants. One interpretation of this finding could be that geographical distances between EU-15 countries are modest and that also cultural and linguistic differences are moderate, at least from the perspective of the high-skilled.

The number of immigrants from country (i) living in country (j) has a significant effect on the skill composition of immigration from country (i) to country (j), which suggests that network or herd effects also affect the skill composition of inter-EU-15 migration. However, since the coefficient is < 0, network or herd effects tend to *lower* the proportion of high-skilled among immigrants (relative to their proportion among the natives in their countries of origin). Most probably this does not mean that networks deter or counteract high-skilled migration. Rather it may be taken to mean that networks increase the migration propensity of low- and medium-skilled more than the migration propensity of high-skilled individuals, thus lowering the relative proportions of the latter among the migrants.[14]

As expected from the Roy-Borjas model, differences in per capita wealth—a proxy for average earning levels—between sending and receiving countries in the EU-15 do not affect the proportion of high-skilled individuals among the immigrants (relative to their proportion among the natives in their country of origin). Thus, even large differences in per capita wealth between two countries have no effect on the extent to which the migrants are disproportionally drawn from either the upper or the lower tail of the skill distribution in the sending country.

The proportion of high-skilled individuals among the native population in the host country tends to increase the proportion of high-skilled individuals among immigrants (relative to their proportion among the natives in the sending country). This could indicate that employment opportunities for high-skilled individuals may have an effect, if a high proportion of high-skilled individuals in the native population in a country signals a knowledge-based economy, which should make a country more attractive as a host country for high-skilled immigrants from other countries in EU-15.

4 Discussion

The present analysis has turned up clearer evidence in favor of the Roy-Borjas model than several similar studies, e.g. Pedersen et al. (2004) or Mayda (2005). The most likely explanation appears to be that the study has focused on migration between EU-15 countries only, which is less restricted by national immigration regimes than the migration focused on in other studies. If this interpretation is correct it implies that the skill composition of migration should be expected to approach the skill composition predicted by the Roy-Borjas model as national immigration restrictions are liberalized or lifted.

[14]This interpretation is supported by the fact that the percentage of low-skilled individuals among immigrants from one EU-15 country into another is positively related to the number of immigrants from the sending country living in the receiving country (r = 0.37).

One possible objection to this generalization could be that the composition of migration between EU-15 countries is special, since much of it is likely to be temporary only, for instance students taking a semester abroad, business executives being posted at branches in other EU-countries for a limited period, etc. In order to get an indication of the extent to which the results in Table 2 are actually driven by temporary migration, Model IV is re-estimated splitting the immigrants into two groups: those who have become naturalized in their host country and those who have not. Obviously the former group can be considered permanent immigrants. Since data on citizenship status of immigrants are not available for all EU-15 countries, the number of observations is reduced to 140.

Table 3 indicates that the results in Table 2 on the skill-based sorting of immigrants into the various EU-15 countries are not driven by temporary migration only. The difference between the spread of the income distributions in the receiving and the sending countries has the same impact on the over- or under-representation of the high-skilled among those immigrants who have become naturalized in the receiving country—and supposedly are permanent immigrants—and those who have not. If anything, the effect seems actually stronger and more significant with immigrants who have become naturalized.

Table 3 Determinants of the locational choice of high-skilled immigrants within EU-15, for nationalized and non-nationalized immigrants separately (weighted OLS, unstandardized coefficients, robust standard errors in parentheses)

Dependent: *Highrat*	Nationalized	Not nationalized
Ginirat	2.623***	2.129**
	(0.730)	(1.047)
Dist	−0.098	−0.336
	(0.169)	(0.206)
Cborder	−0.137	−0.090
	(0.143)	(0.271)
Clanguage	−0.205**	−0.706***
	(0.101)	(0.173)
Elanguage	0.140	0.463
	(0.242)	(0.381)
Immig	−0.002***	−0.002***
	(0.000)	(0.000)
Gdprat	0.280	−0.214
	(0.339)	(0.465)
Hnative	0.036***	0.001
	(0.010)	(0.024)
Constant	−1.102	2.259
N	140	140
Root MSE	0.236	0.483
R^2 (adj.)	0.79	0.73

Fixed effects (country-of-origin dummies) included but not shown

***, **, *indicate significance levels of 1 %, 5 %, and 10 %, respectively

The main difference between immigrants who have become naturalized and those who have not seems to be that with the former group the educational attainment level of the receiving country has a significant effect on the extent to which the immigrants are disproportionally drawn from the upper tail of the skill distribution in their countries of origin; with immigrants who have not become naturalized, this variable shows no significant effect.

Another objection could be that the results might be driven by the particular set of values of the ratios of the Gini-coefficients in the sending and receiving countries and hence not be very robust. As mentioned, relating differences in the spread of the earning distributions in the sending and the receiving countries at a given point in time to the skill composition of a stock of immigrants accumulated over a much longer period may be problematic, unless these differences are (approximately) stable over time.

As a check on the robustness of the results in Table 2, Table 4 shows the results when Model IV is re-estimated with the variable *Ginirat* replaced with a dummy variable, *Ginidum*, where $Ginidum_{ij}$ takes the value 1 if $Ginirat_{ij} > 1$ and zero otherwise. Obviously this way of expressing the difference between source and host countries with respect to the relative spread of their earning distributions will

Table 4 Determinants of the locational choice of high-skilled immigrants within EU-15, with dummy variable for relative differences in the spread of the income distributions in the sending and receiving countries (weighted OLS, unstandardized coefficients, robust standard errors in parentheses)

Dependent: *Highrat*	Model IV
Ginidum	0.361**
	(0.163)
Dist	−0.16
	(0.197)
Cborder	−0.058
	(0.194)
Clanguage	−0.303*
	(0.162)
Elanguage	0.855***
	(0.252)
Immig	−0.002***
	(0.000)
Gdprat	−0.220
	(0.485)
Hnative	0.023**
	(0.013)
Constant	−1.102
N	201
Root MSE	0.483
R^2 (adj.)	0.48

Fixed effects (country-of-origin dummies) included but not shown

***, **, *indicate significance levels of 1 %, 5 %, and 10 %, respectively

normally be considerably more robust over time to changes in the Gini-coefficients involved than the ratio of these Gini-coefficients. According to the logic of the Roy-Borjas model, one should expect a positive coefficient to *Ginidum.* As can be seen in Table 4, the coefficient to *Ginidum* is indeed positive and significant. This demonstrates that the results in Table 1 are not critically dependent on the particular set of values for *Ginirat* used in the estimation.

Finally, Table 5 shows the results when Model IV is re-estimated leaving out of the analysis one country at the time. Regardless of which country is omitted, the coefficient to $Ginirat_{ij}$—arguably the central variable in the present context—remains significantly positive. Thus, the positive effect of the ratio of the income inequality in the sending to the income inequality in the receiving countries for the skill-based locational sorting of immigrants is not driven by a single country.

5 Conclusion

The data appear to support the following conclusions about the determinants of the skill composition of inter-EU-15 migration: Immigrants from one EU-15 country into another EU-15 country are most likely to be disproportionally drawn from the upper tail of the skill distribution in their country of origin,

- if the difference between the spread of the income distributions in the receiving and the sending country is big and positive;
- if the proportion of high-skilled natives in the receiving country is high;
- and if the number of migrants from the sending country living in the receiving country is small.

These findings, especially the first one on the importance of differences in the spread of the income distributions in the sending and the receiving countries for the skill-based locational sorting of immigrants within EU-15, are in agreement with what one would predict from the Roy-Borjas model.

Furthermore, the finding on the importance of differences in the spread of the income distributions in the sending and the receiving countries for the skill-based sorting of immigrants would also appear to have some obvious policy implications. For some time now the EU Commission has been pushing for a harmonization and liberalization of the rules for labor immigration into the countries of the EU. It has also been pushing for an extension of the principle of free movement of labor to third-country nationals residing in EU-countries. An "out of sample prediction" from the present study would suggest that the consequences of these policies in terms of the skill distribution of immigrants can be expected to be less advantageous for highly distributive welfare states like Denmark or Sweden.

Table 5 Determinants of the locational choice of high-skilled immigrants within EU-15, leaving out one country at the time (weighted OLS, unstandardized coefficients, robust standard errors in parentheses)

Omitted	AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
Ginirat	3.101***	2.978***	2.632***	3.611***	3.250***	3.132***	1.314**	2.947***	3.202***	2.866***	2.102***	3.644***	3.557***	3.946***	3.970***
	(0.782)	(0.793)	(0.964)	(0.814)	(0.734)	(0.840)	(0.604)	(0.777)	(0.697)	(0.772)	(0.775)	(0.701)	(0.778)	(0.669)	(1.015)
Dist	−0.202	−0.248	−0.240	−0.222	−0.365*	−0.221	0.003	−0.176	−0.275	−0.152	−0.179	−0.133	−0.264	−0.174	−0.299
	(0.194)	(0.207)	(0.169)	(0.190)	(0.209)	(0.207)	(0.138)	(0.154)	(0.175)	(0.200)	(0.192)	(0.216)	(0.207)	(0.192)	(0.205)
Cborder	−0.114	−0.229	−0.396*	−0.144	−0.053	−0.118	−0.062	0.107	−0.196	−0.034	−0.130	−0.111	−0.144	0.094	−0.139
	(0.172)	(0.214)	(0.232)	(0.169)	(0.158)	(0.172)	(0.193)	(0.154)	(0.157)	(0.183)	(0.204)	(0.173)	(0.181)	(0.162)	(0.170)
Clanguage	−0.343**	−0.090	−0.253	−0.241	−0.337**	−0.243	0.179	−0.150	−0.234	−0.084	−0.355***	−0.196	−0.248	−0.217*	−0.245
	(0.159)	(0.216)	(0.233)	(0.151)	(0.144)	(0.155)	(0.163)	(0.122)	(0.143)	(0.142)	(0.132)	(0.134)	(0.159)	(0.126)	(0.154)
Elanguage	0.410	0.304	0.565**	0.299	0.332	0.370	0.871***	0.138***	0.225	0.574	0.443	0.290	0.261	0.288	0.270
	(0.272)	(0.291)	(0.225)	(0.263)	(0.277)	(0.265)	(0.201)	(0.496)	(0.236)	(0.401)	(0.301)	(0.247)	(0.280)	(0.221)	(0.271)
Immig	−0.003***	−0.003***	−0.002***	−0.003***	−0.003***	−0.003***	−0.006***	−0.002***	−0.003***	−0.003***	−0.002***	−0.003***	−0.003***	−0.005***	−0.003***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gdprat	0.245	1.089	0.179	0.279	−0.016	0.212	−0.263	0.188	0.343	0.154	0.400	0.892	0.325	0.276	0.303
	(0.246)	(0.548)	(0.546)	(0.531)	(0.560)	(0.551)	(0.247)	(0.438)	(0.545)	(0.500)	(0.504)	(0.717)	(0.565)	(0.414)	(0.565)
Hnative	0.024*	0.033**	0.031**	0.022**	0.020	0.024**	0.014	0.026***	0.028**	0.022***	0.020	0.020*	0.025**	0.001	0.016
	(0.013)	(0.013)	(0.017)	(0.012)	(0.013)	(0.012)	(0.012)	(0.010)	(0.012)	(0.011)	(0.023)	(0.011)	(0.012)	(0.012)	(0.013)
Constant	−2.083	−1.519	−1.435	−2.262	−1.965	−2.095	0.198	−0.204	−2.482	−0.155	−0.125	−3.881	−2.231	−2.859	−3.036
N	174	173	174	174	173	174	173	173	174	174	173	175	181	174	174
Root MSE	0.461	0.464	0.442	0.455	0.450	0.462	0.400	0.417	0.416	0.463	0.422	0.443	0.466	0.404	0.448
R^2 (adj.)	0.67	0.71	0.76	0.70	0.71	0.69	0.76	0.64	0.72	0.70	0.72	0.71	0.69	0.76	0.71

Fixed effects (country-of-origin dummies) included but not shown

***, **, *indicate significance levels of 1 %, 5 %, and 10 %, respectively

References

Abramitzky, R. (2009). The effect of redistribution on migration: Evidence from the Israeli kibbutz. *Journal of Public Economics, 93*, 498–511.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review, 77*, 531–553.

Borjas, G. J. (1990). *Friends or strangers: The impact of immigrants on the U.S. Economy*. New York: Basic Books.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature, XXXII*, 1667–1717.

Borjas, G. J. (1998). *Immigration and welfare magnets*. NBER Working Paper No. 6813, National Bureau of Economic Research, Cambridge, MA.

Borjas, G. J. (1999a). *Heaven's door*. Princeton, NJ: Princeton University Press.

Borjas, G. J. (1999b). Immigration and welfare magnets. *Journal of Labor Economics, 17*, 607–637.

Brücker, H., Epstein, G. S., McCormick, B., Saint-Paul, G., Venturini, A., & Zimmermann, K. F. (2002). Managing migration in the European welfare state. In T. Boeri, G. Hanson, & B. McCormick (Eds.), *Immigration policy and the welfare system* (pp. 1–168). Oxford: Oxford University Press.

Chiquiar, D., & Hanson, G. H. (2002). *International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States*. NBER Working Paper No. 9242, NBER, Cambridge, MA.

Chiswick, B. R. (1999). Are immigrants favorably self-selected? *American Economic Review, 89*, 181–185.

Cobb-Clark, D. A. (1993). Immigrant selectivity and wages: The evidence for women. *American Economic Review, 83*, 986–993.

Epstein, G. S., & Gang, I. N. (2004). *The influence of others on migration plans*. IZA Discussion Paper No. 1244, IZA, Bonn.

Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics, 95*, 42–57.

Hatton, T., & Williamson, J. G. (2002). *What fundamentals drive world migration?* NBER Working Paper No. 9159, NBER, Cambridge, MA.

Heston, A., Summers, R., & Aten, B. (2002). *Penn world table version 6.1*. Center for International Comparisons at the University of Pennsylvania (CICUP).

Kaestner, R., & Malamud, O. (2010). *Self-selection and international migration: New evidence from Mexico*. NBER Working Papers No. 15765, NBER, Cambridge, MA.

Mayda, A. M. (2005). *International migration: A panel data analysis of economic and non-economic determinants*. IZA Discussion Paper No. 1590, IZA, Bonn.

Moraga, J. F. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics, 93*(1), 72–96.

Nannestad, P. (2007). Immigration and welfare states: A survey of 15 years of research. *European Journal of Political Economy, 23*(2), 512–532.

OECD (2005). *Database on migrants and expatriates*. http://www.oecd.org/document/51/0,2340,en_2649_37415_34063091_1_1_1_37415,00.html

Pedersen, P. J., Pytlikova, M., & Smith, N. (2004). *Selection or network effects? Migration flows into 27 OECD countries, 1990-2000*. IZA Discussion Paper No. 1104, IZA, Bonn

Roy, A.-D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers, New Series, 3*, 135–146.

World Bank (2006). *World Development Indicators*, CD-ROM. The World Bank, Washington, DC.

Zavodny, M. (1999). Determinants of recent immigrants' locational choices. *International Migration Review, 33*(4), 1014–1030.

Part V
Globalization and Policy Issues

Which Conflicts Can the European Neighbourhood Policy Help Resolve?

Thorvaldur Gylfason and Per Magnus Wijkman

Abstract Under what conditions can the European Neighbourhood Policy (ENP) achieve one of the objectives of the European Union (EU): to resolve conflicts in its neighbourhood? In the spirit of Montesquieu and Monnet, the basic hypothesis of the EU is that closer economic integration helps resolve political conflicts and vice versa, creating a virtuous circle of prosperity and detente. The EU has a strong self-interest in resolving conflicts in its neighbourhood since their consequences spill over into the EU itself through refugee flows and trans-border criminal activities. The paper identifies factors important for the neighbourhood countries to succeed: access to active facilitators of compromise, a well-functioning market economy, strong common institutions to realise potential intra-regional free trade, the prospect of accession to the EU and the absence of third parties to the conflicts. However, characteristic of the key conflicts in the European Neighbourhood is that few of the parties are prospective members of the EU and that a third party outside the Partnership is involved in major conflicts. To succeed in conflict resolution, the EU must, therefore, provide these parties significantly stronger arbitration, more technical and financial resources and stronger institutions anchoring the partner country to the EU. This applies to conflicts involving partners in the Eastern Partnership as well as in the Mediterranean Partnership.

The natural effect of trade is to bring about peace.
Montesquieu (1748)

The authors thank T. Y. Mathä, Banque Centrale Luxembourg, and an anonymous referee for helpful comments as well as S. Munyaneza, UNCTAD, for guidance concerning the use of WITS.

T. Gylfason (✉)
University of Iceland, Reykjavík, Iceland
e-mail: gylfason@hi.is

P.M. Wijkman (✉)
School of Commercial Law and Economics, University of Göteborg, Gothenburg, Sweden
e-mail: per.wijkman@telia.com

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_16

The European Union (EU) was founded on the premise that close economic integration, with common administrative institutions, could contribute decisively to stable and peaceful relations among democratically governed countries (Monnet 1976).[1] The successive enlargements of the EU extended these principles to an ever-larger number of countries that formerly were enemies or dictatorships. In parallel, the EU applied the principle of free trade and successively closer economic integration in various forms to a number of countries in the European Neighbourhood: seven Balkan countries,[2] ten Barcelona Process (BP) countries on the southern Mediterranean coast,[3] and six Eastern Partnership (EaP) countries in Eastern Europe and the Southern Caucasus.[4] Conflicts in the EU's vicinity can easily spill over into the Union itself. Disrupted energy supplies, inflows of refugees and migrants and trans-border criminal activities give the EU a strong self-interest to help resolve conflicts in its neighbourhood. Based on its own experience, the EU sees free trade and close economic integration with common institutions as a means to resolve these conflicts and to foster good-neighbourly relations.

1 Keys to Conflict Resolution in the Neighbourhood Policy

The Neighbourhood consists of very disparate countries, each wanting to be judged on its own, widely differing objectives and merits (Appendix 1). Among these countries some may be eligible for EU membership while others clearly are not; a few are relatively rich countries while many are poor; all are in different stages of transition from plan to market; some are budding democracies and others harsh dictatorships[5]; most are relatively small in terms of population and may wish to compensate for small size through trade[6]; and, lastly, some have excellent relations with neighbours while others are effectively at war with them. Each of these factors has a bearing on whether closer economic integration can help resolve conflicts.

Supposedly a coherent and consistent policy, the European Neighbourhood Policy (ENP) adopted in 2004 is, in fact, an improvised amalgamation of policy initiatives undertaken at different times to satisfy the interests of different EU Member States. It includes the Barcelona Process initiated during the Spanish

[1] Appendix 2 contains acronyms.

[2] Albania, Bosnia-Herzegovina, Croatia, Macedonia, Montenegro, Serbia and Kosovo formed the Central European Free Trade Area (CEFTA 2006). Moldova was included although formally part of the Eastern Partnership (EaP) as were Bulgaria and Romania, soon to accede to the EU.

[3] Algeria, Egypt, Israel, Jordan, Lebanon, Libya, Morocco, Palestinian Authority, Syria and Tunisia.

[4] Belarus, Moldova, Ukraine and Armenia, Azerbaijan and Georgia, respectively.

[5] While the Arab Spring has felled some of these dictators, the outcome does not currently suggest more democratic governance.

[6] Kowalczyk (2000) shows how a small country's optimal strategy may be to join several free-trade areas.

Presidency (Spring 1995) and transformed into the Union for the Mediterranean by the French Presidency (Autumn 2008) and the Eastern Partnership (EaP) proposed by Poland and Sweden and adopted by the European Council during the Czech Presidency (Spring 2009).[7] These improvisations co-habit in the framework of the ENP with its bilateral Action Plans and its flora of Cooperation Council and Committee meetings. When the Commission presented the EaP, it "emphasized the need for a differentiated approach respecting the character of the ENP as a single and coherent policy framework." (Com (2008) 823 final, p. 2). This formulation captures the contradictory nature of the ENP: One size does not fit all but even though the countries have widely different objectives and preconditions the suits are uniform in important respects. Effectiveness when dealing with so many countries requires sorting them into a limited number of groups sharing similar objectives, initial conditions and geopolitical situations and designing common procedures and appropriate incentives for each group. If the situation of each country is *sui generis* then the ENP degenerates into a baker's dozen of bilateral relationships. The ENP needs to reduce the number of differentiated approaches enough to maintain a single and coherent policy framework that is cost effective.

We identify five factors as important if closer economic integration is to contribute to conflict resolution in the Neighbourhood: the existence of a trustworthy facilitator of compromise (Sect. 1.1); the institutional incentives provided by the EU, e.g., the prospect of accession to the EU or of a common institutional framework, as well as financial incentives (Sect. 1.2); whether each party to a regional conflict participates in the ENP or is a third party to it (Sect. 1.3); the potential leverage of regional foreign trade (Sect. 1.4); and last but not least, how far the parties to the conflict have progressed in transition to a well-functioning market economy (Sect. 1.5). This paper groups the ENP countries in these respects to identify which conflicts have high or low likelihood of successful resolution through closer economic integration with each other and to suggest how the likelihood of success can be increased.

1.1 Active Facilitators of Compromise

A conflict which has not ended through unconditional surrender by one party may require that a facilitator assist the parties to settle their differences. In cases where conflicts have involved much death and destruction, outside mediators are usually essential. Bitterness can be so widespread and so deep that neither party is willing either to take an initiative or to respond to one. The EU is in a unique position to act as a facilitator of compromise in its neighbourhood. In major cases it may require

[7]This paper refers to the Union for the Mediterranean as the Barcelona Process (BP) or Mediterranean Partnership. It refers to the partner countries of the Eastern Partnership (EaP) as the Eastern partners.

support from other facilitators as well, but the facilitators must be well coordinated. In the case of the Western Balkans both the US and the European Commission (EC) played key roles while Member States of the EU (e.g., UK, Germany, Sweden) and EFTA (Norway, Switzerland) contributed actively. Facilitators of compromise help the parties to take the long view necessary to resolve their problems.

1.2 Sufficiently Strong Incentives

The prospect of accession to the EU is a key incentive. If both parties to a conflict wish to accede to the EU, they must first resolve their differences. Examples include settlement of disputes between Slovenia and Croatia as well as between Hungary and Romania prior to accession. An unresolved conflict will effectively prevent accession. When the prospect of EU accession is lacking, other incentives are necessary to encourage the parties to solve a regional conflict. The Neighbourhood Economic Community (NEC), proposed by the EU as an institutional structure to manage the ENP, could provide such an incentive but remains as yet undefined. Without it, significant financial incentives will be necessary.

The neighbourhood countries fall into the following groups as concerns their interest in and eligibility for accession to the EU.

1. *Acknowledged candidates for full membership*. The EU has concluded accession negotiations with Croatia[8] and is conducting them with Turkey. The time-table for Turkey is uncertain. The start of effective negotiations with Macedonia has been delayed due to a dispute with Greece over the country's name. In October 2011 the Commission proposed to open negotiations with Montenegro, which had applied for membership in 2008, and negotiations started in June 2012. The EU conferred candidate status on Serbia in March 2012, but it was not until June 2013 that the Serbian Government had made sufficient progress in normalising relations with Kosovo for the EU to declare that negotiations with Serbia could start in September 2013.
2. *Acknowledged potential candidates*. The Thessaloniki Declaration of 2003 states that the "future of the countries of the Western Balkans lies in the European Union." Albania submitted its application for membership in 2009 and has achieved candidate status. Bosnia-Herzegovina must first fulfil a number of conditions, as must Kosovo.
3. *Unacknowledged potential candidates*. The six countries of the Eastern Partnership could be potential candidates if the EU deems them to be European States in accordance with Article 49 of the Treaty of the EU (TEU).[9] This is likely to be

[8] Croatia acceded to the European Union 1 July 2013.

[9] Article 49 of the TEU states *inter alia* that "Any European State which respects the values referred to in Article 2 and is committed to promoting them may apply to become a member of the Union." The official website of the European Union lists all EaP countries as "European countries".

Table 1 Two dimensions of free trade agreements affecting incentives

	Few countries	Many countries
Deep agreement	Difficult, requires strong incentives	Doubly difficult, requires extra strong incentives
Shallow agreement	Easy, weak incentives sufficient	Difficult, requires strong incentives

the case for the three countries of Eastern Europe (Belarus, Moldova, Ukraine) which are west of the Urals, considered by geographers to be the Eastern limit of Europe. The three countries in the Southern Caucasus (Armenia, Azerbaijan, Georgia) lie east of the Bosporus, another traditional geographic limit of Europe, and may require further consideration.[10] The States in the Southern Caucasus border Turkey, an acknowledged candidate. Thus, Turkey's possible accession may improve the eligibility of Georgia and Armenia in accordance with Article 49. Georgia, Moldova and Ukraine have expressed interest in membership. Several Member States (Sweden, Finland, Estonia, Poland, the Czech Republic and Hungary) strongly support the Eastern Partnership.

4. *Non-candidates.* None of the Southern and Eastern Mediterranean coastal States qualify for accession in accordance with Article 49 of the TEU. Morocco applied for membership in 1987 but was rejected by the EU as not being a European State. This set an important precedent. Several EU Member States (France, Spain, Italy) have strong ties to and interests in several of the Southern Mediterranean coastal States and support the Barcelona Process.

The incentives provided to conclude free trade agreements (FTAs) must be proportionate to their costs. A shallow and limited FTA is less costly to negotiate and implement than a deep and comprehensive FTA (DCFTA).[11] A bilateral agreement with a single country is less costly to negotiate than is a plurilateral agreement with many countries. Table 1 illustrates schematically how the depth of the agreement and the number of countries involved combine to affect the costs of negotiating and implementing a trade agreement.

The more ambitious the FTA, the stronger are the incentives needed for a neighbourhood country to wish to participate. Each of these alternatives entails different costs and benefits for different countries.

[10] Regardless of geography, several of these States are currently disqualified because they do not respect the values referred to in Article 2 of the TEU.

[11] Shallow agreements remove primarily tariff barriers and quotas. Deep agreements remove other non-tariff barriers (NTBs) and provide for some trade-related regulatory convergence such as sanitary and phytosanitary standards and intellectual property rights. The scope can be limited to industrial goods or be extended to include agricultural products and services. The free trade agreement, shallow or deep, can be bilateral or plurilateral.

Table 2 Key factors indicating likelihood of success in resolving regional conflicts in ENP. The horizontal line segment in second column is dotted since the operational definition of Europe is unclear

	Conflicts confined to region	Conflicts involving third countries
Potential candidate countries	Western Balkans (e.g., Kosovo-Serbia, Bosnia-Herzegovina)	Ukraine-Moldova-Transnistria (Russia) Georgia-South Ossetia-Abkhasia (Russia)
Not potential candidate countries	Western Sahara (Morocco-Algeria) Middle East (Israel, Palestine and neighbours)	Armenia-Azerbaijan-Nagorno Karabakh (Russia, Turkey)

1.3 No Third Party to Conflict

A third key factor in resolving conflicts is whether the conflict is limited to parties that are members of the Neighbourhood or involves a third party from outside the Neighbourhood. When two ENP countries are in conflict with each other, trade-offs within the framework of the ENP can offer them incentives to resolve the conflict. If a third country (outside the ENP) is involved, a separate deal will need to be struck with that country. Russia is not a member of the ENP.[12] It is, however, involved as a third party directly or indirectly in several of the conflicts in the European Neighbourhood.[13] This reduces the ability of the European Neighbourhood partners to resolve these conflicts in that framework.

Combining the factors laid out in Sects. 1.1 and 1.2 indicates the likelihood of successful conflict resolution (other factors influence this likelihood as well). The various combinations of these two factors are illustrated in Table 2. The likelihood of successful conflict resolution is greatest if all parties to the conflict are members of the region *and* if they all share an ambition to accede to the EU. The EU can then make resolution of regional conflicts a precondition for trade agreements with the EU itself and insist on regional economic integration as a means to resolve these conflicts. Accession is the reward for conflict resolution. Thus, conflicts that fall in the North West quadrant in Table 2 are likely to be the easiest to resolve, other things being equal. The ex-Yugoslavia countries, after bitter conflicts, all wanted to accede to the EU. The prospect of EU accession finally convinced them to settle their differences.

Conflicts that fall in the South East quadrant are likely to be the most difficult and costly to resolve, since the parties are not potential candidates for accession and a third party is involved in the conflict. The countries in conflict thus have

[12]Russia requested not to be included in the ENP preferring instead to have its own Strategic Partnership with the EU.

[13]Russia is involved in the conflict with Moldova concerning Transnistria, with Georgia concerning South Ossetia and Abkhasia and in the conflict between Armenia and Azerbaijan concerning Nagorno Karabakh. Gogolashvili (2011).

weak incentives to resolve their differences. For instance, the conflict over Nagorno-Karabakh will be difficult to resolve if Armenia and Azerbaijan are not both eligible for accession and Russia remains involved in the conflict. The prospects to solve the Transnistrian conflict in the North East quadrant are poor if neither Moldova nor Ukraine is eligible for accession and Russia remains a third party to the conflict. Likewise, the conflict between Georgia and Russia concerning Abkhasia and South Ossetia appears likely to remain frozen as long as Russia remains a third party to the conflict.

1.4 Strong Leverage of Foreign Trade

Whether the potential for regional free trade is large or small affects how much leverage regional free trade agreements can have on conflict resolution. The larger the potential for increased mutual trade, the more likely that normalization of political relations and trade liberalization can increase trade flows and standards of living.

The prospect of large potential gains from trade can help overcome reluctance to resolve conflicts. The actual share in GDP of a country's exports to countries in a conflict situation is a poor indicator of trade leverage. It is the potential for increased barrier-free trade that it is important to know. Kaminski and de la Roche (2002) used a gravity model of trade to estimate potential intra-regional trade for the Balkans in 2000 and found that trade liberalization could triple regional trade. Few comparable estimates of potential trade for ENP countries are available. In 2005 actual intra-regional trade flows of both the Euro-Mediterranean partners and the Eastern partners were extremely low, about 5 % (Sekarev 2011). Based on this observation, these countries may have believed that regional free trade would have little leverage and focused instead on exports to the EU which accounted for about half of their total exports.

1.5 Well-Functioning Market Economies

Implementing a deep and comprehensive free trade agreement with the EU or with each other requires that a Partner country has a well-run market economy. Each EaP country is a former republic of the USSR and thus as yet lacks strong national institutions and a well-functioning market economy. Many of the countries on the Southern coast of the Mediterranean also have weak market economies, though for other reasons. Presumably, each country wants to modernize its economy. However, if doing so imposes a severe economic burden on a poor country, it may opt for shallow rather than deep agreements.

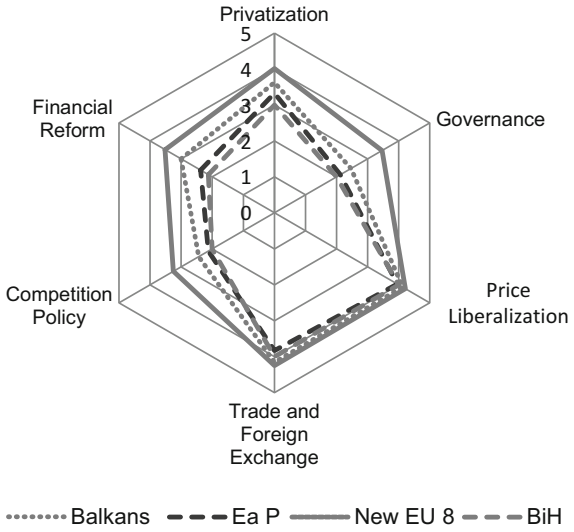

Fig. 1 Indicators of transition to a market economy for three country groups and Bosnia-Herzegovina, 2009. EBRD has published these qualitative assessments by experts annually for 20 years. The EBRD score 1 represents "little or no change from a rigid centrally planned economy" and 4.25 represents "the standards of an industrialized market economy" (*source*: EBRD Transition Report 2010, Table 1.1)

According to the European Bank for Reconstruction and Development (EBRD), the Balkan countries on average lag significantly behind the eight Central and Eastern European countries, that became EU members 2004, in competition policy and governance and enterprise restructuring (Fig. 1). These are essential prerequisites for a DCFTA. The EaP countries on average lag even further behind.

1.5.1 Price Liberalization

These differences between the county groups suggest that the Balkan countries (except Bosnia-Herzegovina) were significantly better prepared than the EaP countries to negotiate a DCFTA, either with the EU or with each other. These differences are, of course, even greater when we consider individual countries rather than group averages. Participation in a DCFTA will require a Neighbourhood country to invest significant resources to build institutions and train personnel in select parts of the *acquis communautaire*.[14]

The next section considers how these indicators—prospects for accession, existence of a third party to a regional conflict, the leverage of foreign trade and the

[14]*Acquis communautaire* refers to European law, including all treaties, regulations and directives adopted by the EU.

functioning of market economy—affect the likelihood that a neighbourhood country will succeed to negotiate a DCFTA.

2 Cases with Favourable Prospects for Conflict Resolution

Success in regional conflict solving has been greatest in the Balkans. Even so, progress there has been slow and incomplete and some countries still have far to go. Nevertheless, most countries have progressed step by slow step (Sect. 2.1). This progress has been due to the following factors:

(1) Active mediators and facilitators of compromise ('godfathering' by the EU).
(2) Credible prospects of accession to the EU and absence of third parties to conflict.
(3) Sufficient progress in transition to a well-functioning market economy.
(4) Large potential for intra-regional trade.

Bosnia-Herzegovina is the exception that proves the rule. Some reasons for its failure to make progress are presented in Sect. 2.2.

2.1 Lessons of Success: The Balkans

The facilitator of compromise in this case was the Stability Pact for South Eastern Europe.[15] Its purpose was to normalize relations between the newly independent States of former Yugoslavia after the bloody civil wars of the 1990s. In 2001 the Stability Pact proposed that these once closely integrated States plus others in the Balkans (Albania, Bulgaria and Romania) negotiate a regional free trade agreement.[16] This was initially rejected by the ex-Yugoslav States because of strong mutual resentments following Yugoslavia's break-up. Each country's prime objective was a trade agreement with the EU followed by EU accession, and not free trade with other Balkan countries. As shown in Figs. 2 and 3, the EU took about 60 % of the Balkan countries exports and imports in 2005, whereas less than 10 % was with other Balkan countries. Instead, they opted for a network, not necessarily complete, of bilateral free trade agreements with selected partners. Intra-regional trade was limited because of the disruptions of the war. However, regional trade had significant unrealized potential, estimated by the World Bank to be about three times

[15]This organization was initiated by the EU and the US in 1999 after the end of the Balkan wars. A number of other states were active in this group (UK, Germany, Sweden, Switzerland, Norway) as well as adjacent Hungary, Slovenia and Turkey. Their presence provided support and encouragement and dampened animosities.

[16]Gylfason and Wijkman (2011a).

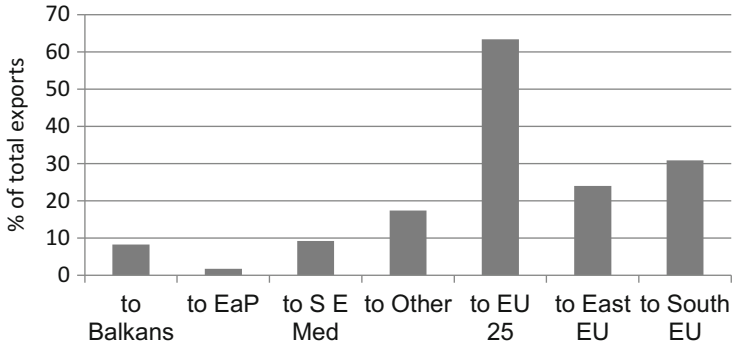

Fig. 2 Distribution of Balkan countries' exports to neighbours 2005 (*source*: Authors' calculations using The World Bank's WITS program)

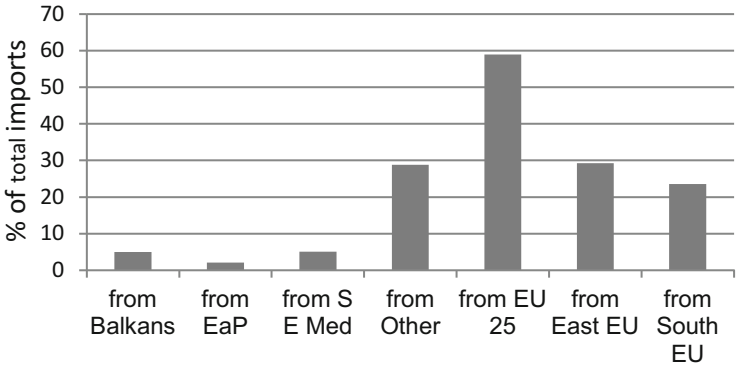

Fig. 3 Distribution of Balkan countries' imports from neighbours 2005 (*source*: Authors' calculations using The World Bank's WITS program)

larger than actual trade in 2000.[17] The Stability Pact proposal aimed to realise this potential.

The number of bilateral agreements in the network grew, to avoid preferential discrimination, and the Balkan States finally agreed to convert the network into a regional free trade agreement. This agreement was to include all Balkan countries and also to liberalise more than just industrial products. The countries signed it in December 2006 and, significantly, called it CEFTA 2006.[18] Why did it take 5 years

[17]Kaminski and de la Rocha (2002). Trade between the core countries Bosnia-Herzegovina, Croatia and Serbia had been badly disrupted by the wars in the 1990s.

[18]On 1 January 2007 Bulgaria and Romania left CEFTA and entered the EU. Their somewhat reluctant presence in CEFTA was important to ensure the ex-Yugoslav States that regional free trade was not a Yugoslavia redivivus. CEFTA's sole remaining member would have been Croatia if it had not opened CEFTA for the Western Balkan countries and revised the treaty. Croatia, which

to achieve this? In addition to the low level of actual intra-regional trade following the wars, the Balkan countries initially feared that regional free trade was a substitute for EU membership. It took long for the EC representative to convince them that regional free trade was, in fact, a prerequisite for accession.[19] A key event was the Thessaloniki Declaration by the European Council in June 2003 stating that the future of the western Balkans lay in the EU. By repeating these words often enough during the Stability Pact meetings the European Commission finally convinced also the sceptics that accession was a credible prospect. This made all the difference. As if to drive home the point, the Commission invited the Parties to conduct the final rounds of negotiations on a single regional free trade agreement in the Charlemagne building in Brussels mid-2006. Once each Balkan State was convinced that it had a credible prospect of accession, it accepted to negotiate a regional free trade agreement, as a precondition for a Stabilization and Association Agreement (SAA) with the EU, which in turn was a precondition for EU membership.

The regional free trade agreement negotiated within the framework of the Stability Pact for Southeastern Europe and godfathered by the EU thus put the Balkan countries on track for accession. In February 2008 the Stability Pact was dissolved and replaced by the Regional Cooperation Council, 'owned' by the States of the region and headquartered in Sarajevo. The establishment of this common institution was an indication of the significant progress made in regional conflict resolution and an important instrument for continued progress.

The accession track contained the following stations: Regional free trade, deep and comprehensive free trade agreements of each country with the EU (Stabilisation and Association Agreements, SAAs) and requests for and negotiations on accession to the EU. Each country progressed from one station to the next on its own merits. Thus, in some cases progress was rapid, in others it was slow, but in almost all cases it was steady (Table 3). The fact that there was a common objective (EU accession) and a structured road map on how to get there proved essential for progress. The accession track allowed the Balkan countries to start with the easy things and finish with the more difficult (Table 1): relatively shallow agreements with select partners, a plurilateral deep agreement with each other, and finally a bilateral SAA with the EU with the prospect of future accession.

The Western Balkan countries passed the various stations on the track to accession in an order closely related to the progress each had made toward a functioning market economy (as measured by EBRD *Transition Report* indicators). The ability to negotiate a deep and comprehensive free trade agreement depends not only on how far the country has progressed in transition to a market economy, but also on its government's preparedness to resolve any regional conflicts and the

was already negotiating accession with the EU, found this especially appropriate: CEFTA had been a transit station for Central and Eastern European countries on their way to the EU.

[19]The EU was not prepared to admit as member a country with unresolved conflicts with its neighbour. Such conflicts would become an external conflict of the EU, if one country acceded, or an internal conflict, if both countries became members.

Table 3 Key dates for Balkan States en route to EU membership

Country	Signing of SAA	SAA enters Into force	Application for EU membership	Commission issues positive opinion	Council confers candidate status	EU Accession negotiations start	EU Member
Albania	2006-06-12	2009-04-01	2009-04-24	2012-10-10	2014-06-24		
Bosnia- Herzegovina	2008-06-16	2008-07-01					
Bulgaria			1995-12-16			2000-02-15	2007-01-01
Croatia	2001-10-29	2005-02-01	2003-02-21-	2004-04-01	2004-06-01	2005-10-03	2013-07-01
Macedonia	2001-04-09	2004-04-01	2004-03-22	2005-11-09	2005-12-16	2012-03-09	
Montenegro	2007-10-15	2010-05-01	2008-12-15	2010-11-09	2010-12-17	2012-06-29	
Romania			1995-06-22			2000-02-15	2007-01-01
Serbia	2008-04-29	2013-09-01	2009-12-22	2011-10-12	2012-03-01	2014-01-21	

Source: Data assembled from www.EuropeanUnion.org

degree of national support it enjoys. Therefore, it is hardly surprising that of the Balkan countries Croatia made the most rapid progress and Bosnia-Herzegovina the slowest progress toward EU membership. Bosnia-Herzegovina has the lowest transition indicators and the most dysfunctional political system of all the Balkan countries.

2.2 Lessons of Failure: Bosnia and Herzegovina

The conflicts that once caused the violent disintegration of Yugoslavia remained unresolved *within* the Federation of Bosnia-Herzegovina. These internal dissensions are the main reason for its lack of progress since the signing of the Dayton Accords in 1995.[20]

Bosnia-Herzegovina is a dysfunctional Federation where the politicians of various ethnic groups are unable to agree on the key measures necessary for a trade agreement with the EU: to accede to the WTO, to revise the Constitution imposed by the Dayton Accords and to replace the Office of the High Representative (OHR) by a democratically appointed office.[21] In addition, Bosnia-Herzegovina has failed to improve the functioning of its market economy. The Dayton Accords were imposed by outsiders on the Serbian aggressors, making Dayton a cease-fire rather than a peace treaty. War-time political leaders remained in power both in Bosnia-Herzegovina and in its immediate neighbours. National animosities in Croatia and Serbia became ethnic enmities within Bosnia-Herzegovina, freezing the conflict there.[22]

The resulting deadlock of Bosnia-Herzegovina's political system crippled its economic recovery and growth after the war. Per capita GDP fell sharply during the war 1992–95 and remains significantly below its neighbours' (Fig. 4). Bosnia-Herzegovina is heavily dependent on remittances from abroad. Few refugees have returned and foreign direct investment is low. The wage claims of a large public sector erode the competitiveness of the tradable sector. Other Balkan countries continue to show modest growth, especially Serbia (with Montenegro) where the Kosovo war had long-lasting economic consequences. Only Croatia, which opened EU membership negotiations in 2005, recovered quickly and continues to grow rapidly.

To avoid a break-up of Bosnia-Herzegovina, its constituent ethnic groups must find a compromise that provides a *modus vivendi*. If the entity breaks up, possible

[20]Gylfason and Wijkman (2011b).

[21]The international community created the OHR and empowered it to veto laws. This reflected its distrust of Bosnia-Herzegovina's ability to govern itself in an initial phase. This prolonged foreign tutelage has led to dependency and irresponsibility on the part of the country's politicians.

[22]A petty detail was parliamentarians' inability to agree on new identity cards prior to summer 2013. This had tragic consequences, causing the death of several infants who were unable to cross borders for medical care.

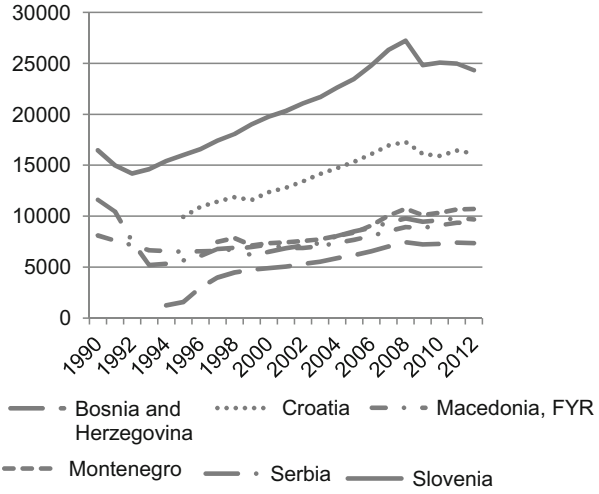

Fig. 4 Balkan countries: GDP per capita 1990–2012 (PPP, constant 2005 international $) (*source*: World Bank, *World Development Indicators* 2013)

successor states are unlikely to qualify for membership in the EU, especially if the breakup is bloody. But can Bosnia-Herzegovina transform itself into a viable state whose citizens share a vision of a common future? Perhaps. If the entities cease to believe they have another option, they might accept the necessity to live together. Reconciliation will require active facilitators of compromise, including the EU as well as the US, and provision of financial assistance for economic reconstruction and restitution of property to returning refugees.

3 Cases with Unfavourable Prospects for Conflict Resolution

The Barcelona Process has made little progress to resolve conflicts in the region. This is because

(1) Accession to the EU is not an option, and no close substitute has been presented.
(2) Active and concerted facilitation in the BP has not been forthcoming to resolve the 'local conflicts'.
(3) The amount of intra-regional trade is small and its potential is uncertain.
(4) Most countries have poorly functioning market economies and governments incapable of implementing DCFTAs.

In addition, having been governed by nepotistic, corrupt and dictatorial regimes for decades, many States in the region have a strongly entrenched legacy of inept governance. The fundamental problems being primarily political, political solutions are necessary. The Arab Spring provided a rare window of opportunity to support

far-reaching political reforms in some of these countries. But this window has now closed in most countries as a result of deeply imbedded political obstacles to reform. Democratic reform will require a more assertive and innovative EU policy and hard decisions by partner countries to effect necessary political reforms and conflict resolution (Sect. 3.1).

The Eastern Partnership is now launched after a difficult start. However, the prospects for successful conflict resolution appear dim due to the low potential regional trade and to the existence of a third party to the conflicts. In this case, the fundamental problems impeding progress are economic and require improving the functioning of market economies (Sect. 3.2).

3.1 The Barcelona Process: Addressing Political Obstacles to Success

The Barcelona Process set up as a target that each Euro-Mediterranean partner should conclude a bilateral free trade agreement (FTA) with the EU as well as with each other before the end of 2010. This would create a vast free trade area around the Mediterranean.

This goal has been achieved to a degree but with an important exception and with a low level of ambition. The EU concluded bilateral FTAs with all countries but Ghadafi's Libya. However, most of these agreements are 'shallow'. Those with Egypt, Israel, Jordan, Morocco and Tunisia are deeper although none are as deep as the trade agreements that the EU has with the countries of the Western Balkans. Free trade has been achieved between all Arab countries on the Mediterranean coast through the agreement on the Pan-Arab Free Trade Area (PAFTA), established at the initiative of the Economic and Social Committee of the Arab League in 1997.[23] PAFTA claims to have realized free trade in industrial and agricultural goods already in 2005. This, too, is a shallow agreement. In addition, country coverage is incomplete, since it does not include Israel.

Four countries have been more ambitious. Egypt, Jordan, Morocco and Tunisia established free trade between them in 2005, 5 years ahead of the Barcelona schedule, through the Agadir Agreement.[24] This Agreement is deeper than PAFTA. It provides *inter alia* for adoption of the Euro-Med protocol on preferential rules of origin, thereby allowing for diagonal cumulation of origin for participating countries in their trade with the EU. The volume of trade among the Agadir countries is small, about 2 % of the total and half that of trade among the other PAFTA countries. However, it has increased significantly as a result of the entry into force of the deep

[23]It thus also includes eight Arab League countries that are *not* members of the Barcelona Process.

[24]This Agreement entered into force in 2004 and was negotiated after the signing of the Agadir Declaration in 2002.

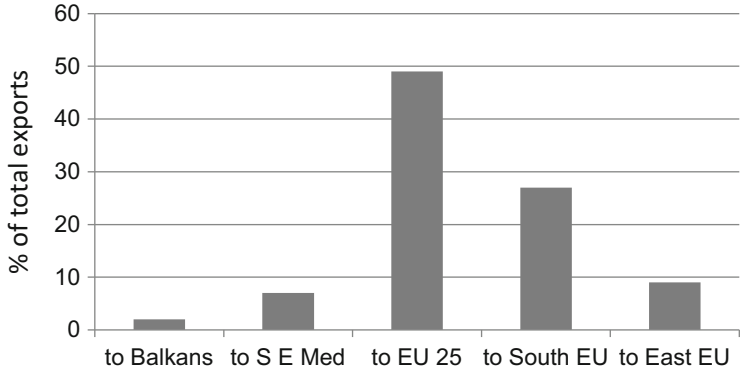

Fig. 5 Distribution of South and East Mediterranean countries' exports to neighbours 2005 (*source*: Authors' calculations using The World Bank's WITS program)

Agadir Agreement in 2005. No such increase can be noted for the other PAFTA countries, whose free trade agreement has remained shallow.

Viewed as an exercise in conflict resolution, the Barcelona Process has not produced results. No progress has been made to resolve the conflict over Western Sahara (between Morocco and Polisario/Algeria) or over the former League of Nation's mandate Palestine (between Israel and its neighbours). Indeed, the BP does not appear to have even attempted to resolve these conflicts.

3.1.1 Limited Regional Trade Leverage?

What explains this shallow and incomplete pattern of regional trade liberalisation? Is potential intra-regional trade really small? Independently of their factor endowments, the Southern and Eastern Mediterranean countries should be able to benefit from economies of scale and scope like any other small country. Actual intra-regional trade among these countries is strikingly low (Fig. 5)—less than 8 % of their total trade.[25] Of this small amount the States in the Middle East take about 50 % more than the States in North Africa, which in turn take marginally more than the four Agadir States (Egypt, Jordan, Morocco, Tunisia). Perhaps policy makers believe that little leverage is to be gained from attempts to expand intra-regional trade. But before such conclusions can be drawn, studies of potential trade are needed.

[25]Trade Commissioner Peter Mandelson described the situation bluntly in his address to the First Agadir Investment Forum, in Brussels 8 April 2008. "... creating a regional market place along the southern shore of the Mediterranean ... has turned out to be ... a political challenge of the first order. You don't need me to tell you how far away you are from ... [your] ... goals. ... regional integration in the Southern Mediterranean is lower than anywhere else in the world. ... You do not trade with your neighbours. It makes you almost unique in the world."

If intra-trade between the countries expands as a result of the Agadir Agreement, the other PAFTA countries may be encouraged to deepen the PAFTA agreement (and their bilateral agreements with the EU), including adopting Euro-Med preferential rules of origin. For the present it seems that few countries are prepared to raise their level of ambition for intra-regional trade by moving from shallow to deep FTAs.

There is one case where unrealized potential trade is likely to be large: between Israel and its neighbours, as indicated by Gylfason et al. (2012). The failure to realize this economic potential is solely political. No Arab country has signed an FTA with Israel, the country with the largest and fastest growing GDP in the region. Two Arab countries signed peace treaties with Israel: Egypt in 1979 and Jordan in 1994. They also have trade agreements with Israel, but not FTAs as the BP calls for. Such trade agreements between Israel and its bordering countries could generate significant regional growth. The failure of the Barcelona Process to encourage conclusion of FTAs between Israel and neighbouring countries has contributed to continuation of this conflict.

When it started in 1995, the Barcelona Process was fuelled by optimism generated by the Madrid Conference of 1991, the Oslo Accords of 1993 between Israel and the PLO and the Israel-Jordan Peace Treaty of 1994. However, when Israel and the Palestinian Authority failed to reach a final settlement at the Camp David Summit in 2000, hostilities between Israel and Palestinians increased. It would be too easy to blame the failure of the BP on the renewed hostilities: the task of the BP is to assist the protagonists to resolve these conflicts not to give in to them. The failure to hold Union of the Mediterranean Summits or Foreign Minister meetings after the Gaza crisis 2008–9 was not an appropriate response.

In order to achieve its target of a Mediterranean free trade area, the BP should encourage the following steps. Egypt and Jordan each negotiate and sign deep and comprehensive FTAs with Israel. These three countries have already signed peace and trade agreements with each other. The bilateral free trade agreements could be ‘carbon copies’ of the deep Agadir Agreement, making it easy at a later stage to include also Morocco and Tunisia, which have informal relations with Israel.

The potential for increased trade between Egypt, Israel and Jordan is likely to be significant, characterised as these countries are by different factor endowments, geographic proximity and large economic size.[26] The Palestinian territories would benefit from improved access to neighbouring markets. The prospect for achieving this is no less realistic today than the Common Market was in Europe after World War II.

[26]See Gylfason et al. (2012).

3.1.2 Can the EU Provide Incentives for Conflict Resolution?

Can the EU encourage the signatories of the Agadir Agreement to include Israel and others in it? The EU's *New Response to a Changing Neighbourhood (COM (201)202* offers the following general incentives:

1. Improved access to the EU market.
2. Institutional anchoring in a community of functioning market economies.
3. Financial and political support for democratisation in the Southern and Eastern Mediterranean.

The EU is the major market for the countries along the Southern and Eastern coast of the Mediterranean, taking about 50 % of their exports, of which the Southern EU States account for slightly more than half. Improved access means opening the EU market to agricultural products, the main export from the Southern Mediterranean coastal countries, energy services (not least solar energy), and other services (in particular labour). The EU has taken a first important step. It opened DCFTA negotiations with Morocco in 2013 and is considering a DCFTA with Tunisia. These could serve as a model for agreements with Egypt and Jordan. The Agadir four have relatively advanced market economies and close ties to key countries in the EU. Israel is also closely integrated with the EU. An arrangement whereby the EU negotiates DCFTAs with each of the Agadir four and with Israel and requires these to have DCFTAs with each other could produce significant economic and political gains for these countries.

The EU foresees that an institutional framework called the Neighbourhood Economic Community (NEC) encompassing the EU and participating neighbours will administer such a deep and comprehensive free trade area. The EU does not define either the scope or the content of this common institution except by way of a vague reference to the institutional structure of the European Economic Area (EEA).[27] Appropriately designed, the NEC could be a powerful institutional framework for reform and modernisation, much as the Organisation for European Economic Cooperation (OEEC) was in post-war Europe. It is time for the EU to give functional precision to this vague NEC. Those countries that have DCFTAs with both the EU and with each other would thus have an institutional anchor to manage close economic integration and to lock in its significant economic benefits.

An institutional superstructure is by itself insufficient to resolve the deep-rooted crises in the Middle East. The EU's new approach to the changing neighbourhood in 2011 provides a number of important initiatives (Sect. 4). It will require addressing also the resentments of past wars. The Palestinian refugees, now registered at about 4.8 million (including their descendants), need decent living conditions and job opportunities. The immediate parties to the conflict are unable or unwilling to bear these costs themselves. Outside donors will need to contribute financially to create the political preconditions for a settlement.

[27] The EEA consists of the EU, Iceland, Liechtenstein and Norway.

3.2 The Eastern Partnership Countries: Addressing Economic Obstacles to Success

The Eastern Partnership has taken off in spite of initial difficulties. Ukraine and Georgia entered talks on DCFTA with high expectations only to find that these negotiations were more complex and took longer than they had expected. Just getting started took time.[28]

The EU started negotiations with Ukraine on a DCFTA in February 2008 in the wave of enthusiasm that followed the Orange Revolution in 2004/5. Progress was slow. Negotiations were concluded in 2011. The Ukraine administration lacked technical competence, institutions and experience for running market economies. In addition, uncertainty grew concerning the commitment of Ukraine's political leadership to basic democratic values, to non-corrupt governance and the country's geopolitical role.

Georgia requested a DCFTA following the Georgian-Russian war in August 2008. The difficulties encountered in the negotiations with Ukraine led the Commission to send a fact-finding mission to assess Georgia's preparedness. In its report in 2009 providing Key Recommendations, the EC identified a number of issues where the Commission suggested that Georgia had to make progress before negotiations could start. After 2 years of 'pre-negotiations', the EU announced in December 2011 that negotiations on a DCFTA with Georgia would start in February 2012. Surprisingly, they were concluded already in July 2013. One reason for the acceleration of the pace of pre-negotiations after a slow start was that the EU initiated pre-negotiations also with Moldova and Armenia. This converted what had been EU's bilateral tug of war with Georgia into a competitive race to the fore between these three countries.

The EC sent a fact finding mission to Moldova at the end of May 2010 to ascertain its capacity to negotiate and implement a DCFTA and it issued its Key Recommendations on 26 October 2010. Already in December 2011 the EU announced that it was prepared to launch negotiations on a DCFTA. These started in February 2012 and were completed in July 2013. The rapid pace of negotiations with Moldova can be explained partly by its familiarity with the EU *acquis* after negotiating the CEFTA 2006 with the Balkan countries. However, this cannot explain why it took Armenia even less time to negotiate a DCFTA.

The Association Agreements with Armenia, Georgia and Moldova were to be initialed and the Agreement with Ukraine signed at the EaP Summit in Vilnius in November 2013.[29] However, Russia exerted strong pressure on these former republics of the USSR to join instead its Eurasian Customs Union. Armenia chose

[28]The disappointments that initially characterized the EaP are described in Messerlin et al. (2011) in the case of Georgia.

[29]Azerbaijan and Belarus are not yet WTO members, a precondition for a DCFTA, so for this reason as well as for political reasons, an assessment of these countries' preparedness for a DCFTA is far off.

to do this.[30] The Ukraine President chose not to sign the Association Agreement with the EU after strong pressure from Russia.[31] His decision led to popular demonstrations in Kiev (Euromaidan) surpassing those of the Orange Revolution, to a new government in Kiev, to Russia's annexation of Crimea and support of separatists in Eastern Ukraine. Georgia, Moldova and Ukraine finally signed Association Agreements on 27 June 2014.

Judging from experience to date "negotiations to start negotiations" on a DCFTA with the EU can take 2–3 years, whereas the negotiations on a DCFTA itself can take at least 2 years. The total of 4–5 years is not a short period, but given these countries' poor initial conditions it is an impressive achievement. No EaP State has as yet opened negotiations with another EaP State on a DCFTA and there is no requirement for this in the EaP and hence no deadline. Realisation of a free trade area encompassing the EU and key EaP States will take time due to the failure to resolve conflicts in this region. Progress is likely to be slow also for the following reasons:

(1) Insufficient transition from plan to market in the EaP countries.
(2) Low levels of intra-regional trade in the EaP.
(3) Mixed economic and political objectives of some participants.

We discuss these in turn.

3.2.1 Poorly Functioning Market Economies

Most EaP countries have made only modest transition to a market economy since independence. Administrations in some republics of the former USSR lack functioning national institutions and retain residual reflexes of central planning, including significant corruption. With slow progress toward a market economy, most of the EaP countries had not recovered their real per capita income 15 years after independence (Fig. 6).

The EaP countries must build the institutions necessary to run a market economy and to train staff in relevant parts of the *acquis communautaire*. Therefore, implementing the DCFTA with the EU will take time. This is not a quick fix!

[30]Reports indicate that Armenia's decision not to sign the Association Agreement with the EU was influenced by Russia's threat to remove its security guarantee concerning Nagorno Karabakh.

[31]Russia threatened to retaliate by cancelling purchases of goods and services, investments, increasing border controls and cancelling oil and gas deliveries. In addition, the Ukraine government failed to meet the EU's conditions concerning application of "selective justice" (e.g. to release Yulia Tymoschenko). See Gylfason et al. (2014) and Gylfason and Wijkman (2014).

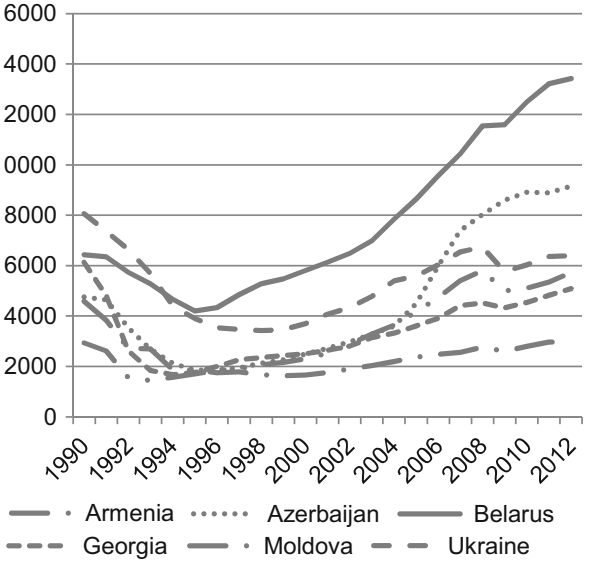

Fig. 6 EaP countries: GDP per capita 1990–2012 (PPP, constant 2005 international $) (*source*: World Bank, Development Indicators 2013)

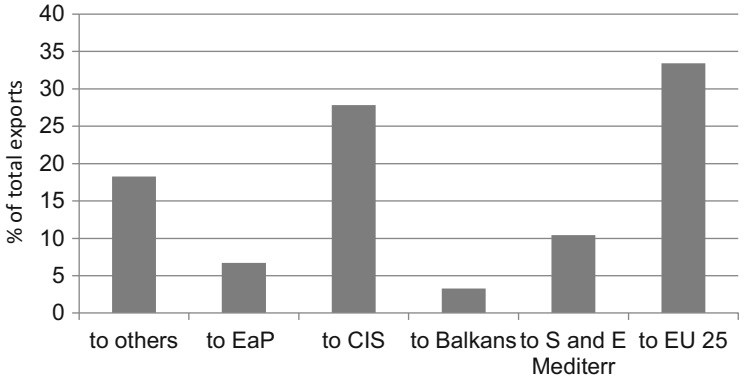

Fig. 7 Distribution of EaP exports to neighbours in 2009 (*source*: Authors' calculations using The World Bank's WITS program)

3.2.2 Limited Regional Trade Leverage?

Intra-regional trade in 2009 is small for the EaP six—around 7 % of exports (Fig. 7). It is thus less important than intra-regional trade was in the Balkans after the disruptions of the Yugoslav civil wars (about 10 %).

Most of the EaP six have regional FTAs with each other, albeit very shallow ones, through participation in Russia's Commonwealth of Independent States

(CIS). It is, therefore, unlikely that the priority of any of the EaP six will be to conclude a bilateral DCFTA with others. Prime candidates are neighbouring countries. The EaP countries have dramatically reoriented their trade from East to West in the past 10 years. Other CIS countries, including Russia, take almost 30 % of EaP countries exports while the EU now takes over 30 %. Nevertheless, the EU still takes a significantly smaller share of the EaP partners' trade than it takes of the Balkan's exports or of the Southern and Eastern coastal States of the Mediterranean. Political—rather than economic—considerations have played a major role in making a DCFTA with the EU attractive.

3.2.3 Mixed Objectives: Mixed Minds

In Georgia and Ukraine political considerations dominated economic ones when considering the prospect of a DCFTA with the EU. Georgia, originally satisfied with a shallow agreement, opted for a DCFTA only after the August 2008 war with Russia. Ukraine aimed at a political reorientation to the West following the Orange revolution in 2004/05. In both countries, the DCFTA was seen as a means of improving national security in the face of an external threat with internal dimensions. NATO membership was the priority objective and in its absence a DCFTA was a second-best option. Thus, a mismatch arose between what these EaP governments really wanted at the time, namely national security when faced with an external threat, and what the EU really could deliver, namely closer economic integration. The DCFTA was an economic slow-fix applied in a situation where geopolitical threats to vital national interests required a security quick-fix. The tedious nitty-gritty of trade policy and the absence of security policy alternatives to divert the threat from a third party outside the EaP, resulted in a slow and hesitant start that caused disappointment and frustration. This difficult economic first step has now been taken and, if successful, will make additional steps easier to take.

3.3 A Summary View of Progress

Both the Balkan countries and the Mediterranean partners started with the easier negotiations, those for a shallow agreement. By contrast, the EaP countries started with the most difficult: a deep and comprehensive bilateral FTA with the EU. The trajectories of the EaP countries, the Balkan countries and the BP countries are illustrated by the arrows in Fig. 8. This is a dynamic version of Table 1, illustrating how the complexity of free trade agreements increases as they grow deeper and include more countries. These factors explain the varying progress of the different country groups.

The Balkan countries started with shallow, bilateral agreements with each other. They proceeded to negotiate a single regional deep free trade agreement only when it was clear to them that this was a precondition for Stabilisation and Association

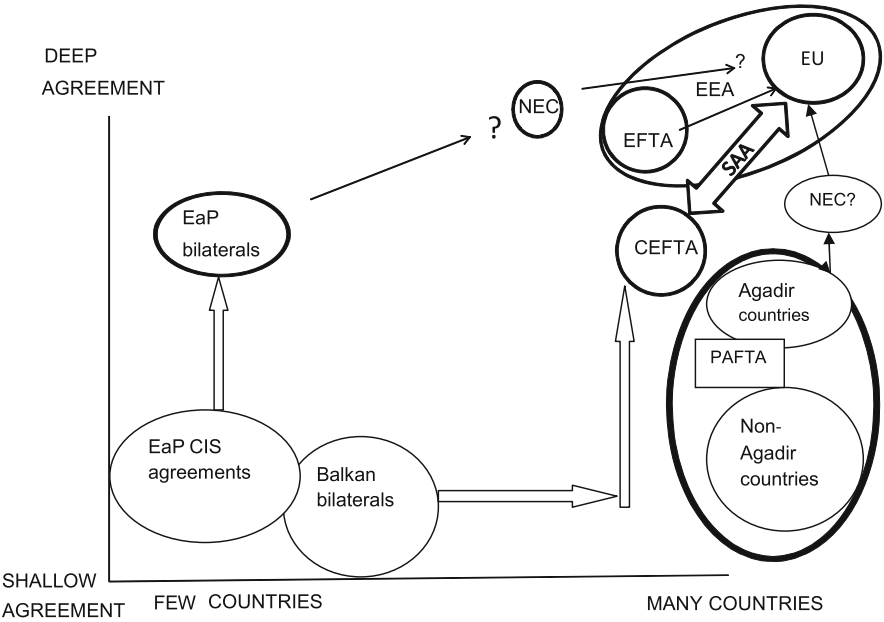

Fig. 8 Schematic presentation of complexity of various DCFTAs in terms of depth of agreements and number of participating countries. The figure illustrates how complexity grows from a shallow free trade agreement between two Balkan countries in the lower left corner to the deeply integrated 28-country EU in the upper right corner. The various groupings of countries are linked with each other through more or less complex trade agreements

Agreements (SAAs) with the EU. Normalisation of relations between former belligerents was necessary for EU accession. Each Balkan country thus started with the easy parts before proceeding step by step to the more difficult task of conflict resolution.

The Euro-Mediterranean partners also started with the easy parts and avoided what was economically and politically difficult. They concluded shallow free trade agreements with each other and with the EU. Most realised the target date (2010) for achieving free trade but none entered free trade agreements with Israel. Only the Agadir four entered deeper agreements with each other. The Barcelona Process provided no sanctions for failure to reach its target of a vast free trade area including countries on both sides of the Mediterranean.

The EU required that the EaP partners approximate much of the EU trade-related *acquis*. The EU does not specify a target date for the EaP as it did for the BP. Nor does it make regional free trade among the EaP countries a condition for a bilateral agreement with the EU as it did for the Balkan countries. Nevertheless, the first EaP countries encountered significant difficulties in doing this. By opening the prospect of negotiations on a DCFTA also for Moldova and Armenia, the EU introduced a competitive element that allowed a State to show whether it was both willing and able to negotiate a DCFTA in a reasonable period of time. Unlike the Balkans, no

prospect of accession was held forth and, consequently, no accession track provided. This removed an important incentive. The national security concerns, a priority of some of the EaP countries, were addressed only indirectly through an unspecified institutional framework, the Neighbourhood Economic Community (NEC). Conflict resolution had initially a low profile given that Russia, not a member of the EaP, had a significant role as a third party in most of the conflicts.

A low threshold of complexity contributes to the initial success of the Euro-Mediterranean partners, while a high threshold explains the initial difficulties of the Eastern partners. The slow start of the EaP countries is finally paying off in substantive and substantial progress for several countries. The Euro-Mediterranean partners still face the need to make substantive progress in conflict resolution.

4 Implications for the Partner Countries

The High Representative and the Commission in May 2011 outlined a new approach for the ENP in response to the Arab Spring (COM(2011)201).[32] Importantly, the new approach provides the Euro-Mediterranean partners with the option of negotiating DCFTAs. Among other proposals are setting up a Partnership for Democracy and Shared Prosperity in the Southern Mediterranean, diversifying programmes to address countries' specific needs, expanding support from government to civil society organizations through a new Civil Society Facility, providing also the Mediterranean countries with a Comprehensive Institution Building Programme, and increasing assistance in the ENP programme of €1,242 million up to 2013. The European Bank for Reconstruction and Development (EBRD) extended its area of operations to include the Southern and Eastern Mediterranean in 2012. To avoid that the Euro-Mediterranean partners experience similar frustrations as the first EaP partners in the initial negotiations for a DCFTA it is necessary to ensure that they and the EU have a common understanding of what is negotiable.

4.1 Understanding What Is Negotiable and What Is Not

The Commission has defined the elements that a DCFTA is to contain. In addition to eliminating tariffs and quotas as in shallow FTAs, DCFTAs also approximate *inter alia* technical regulations, sanitary and phytosanitary standards, intellectual property rights, competition policy, state aid and liberalise certain services and capital movements. To avoid trade diversion and extra administrative costs for both government and business the DCFTAs must be uniform and not vary from country to country.

[32] *A New Response to a Changing Neighbourhood,* Brussels25/05/2011, COM(2011) 303.

Thus, little of the content of a DCFTA can be individually negotiated. However, it is possible to negotiate how and when to implement its various parts. Both the *sequence* in which the various elements are adopted, the *time period* over which they are introduced (transition periods), the degree of *symmetry* in the 'concessions' exchanged between the parties and the *technical and financial assistance* provided to the partner country can vary from country to country (Wijkman 2011). These are appropriate subjects of negotiations. Thus, a government should use its negotiating credibility to reach agreement on sequencing of issues, transition periods, asymmetric concessions and financial and technical assistance rather than squander it on trying to modify elements of the *acquis* that define the DCFTA. Kowalczyk and Davis (1998) provide evidence that suggests that an important element of trade negotiations concerns the terms of phasing in preferential trade agreements. They found *inter alia* that the length of transition periods for eliminating tariffs on particular goods between Mexico and the USA were a variable in the NAFTA negotiations.[33]

4.2 Establishing Credibility and Mutual Trust in the Negotiating Process

The Euro-Mediterranean partners can expect to face several years of 'pre-negotiations' during which the European Commission assesses whether they are prepared to negotiate a DCFTA. How long it will take to qualify for negotiations will depend *inter alia* on the negotiating credibility of the partner country. Can it make a credible commitment to approximate the relevant *acquis*, present a credible time-table for such approximation and a credible plan to implement the *acquis* by an agreed-upon date? If so, this will hasten the start of negotiations. If not, it may have to 'deliver' substance before the negotiations can start: afterall, seeing is believing. Since the negotiations on a DCFTA can take 4 years or more, the partner country should use the time to approximate necessary legislation, train personnel and build institutions. Additional time to adapt is available during transition periods after the entry into force of the DCFTA. Negotiating credibility enables a country to accelerate the negotiations.

4.3 Furthering Transition from Plan to Market

A DCFTA is a powerful instrument for modernising economies but a costly one. It requires a long-term plan for the partner country to take each of the many steps involved to legislate and implement significant parts of the *acquis*. A realistic view

[33]Mexico tended to accept US terms of phasing out tariff rates, advocated by special interests, so as not to jeopardize its preferential access to the large US market.

of how long it will take to build the institutions, approximate the legislation and to train the staff needed to manage market economies allows the partner to sequence events effectively, negotiate appropriate transition periods and secure funding of the costs of implementing the *acquis*. Thus, the partner should view negotiating a DCFTA as an effective mechanism to make the transition from plan to market.

5 Implications for the European Union

The more ambitious policy vis-à-vis the Euro-Mediterranean partners following the review by the High Representative of the Union for Foreign Affairs and Security Policy and the European Commission of the ENP increases the demands on the EU. Negotiating DCFTAs with "willing and able partners" in the Mediterranean basin will require the Commission to devote significantly greater resources to the ENP. In general, the EU has raised its expectations on the Euro-Mediterranean partners, increased the incentives it provides them ("more for more") and demanded greater efforts by the institutions of the ENP (Comprehensive Institution Building Programme, Partnership for Democracy and Shared Prosperity, Civil Society Facility). This intensified work will require the EU to devote special attention to the following tasks.

5.1 Strengthen the Resources of the Trade Directorate in the European Neighbourhood Policy

As the lead directorate for a DCFTA, Directorate General Trade must have sufficient 'command and control' over the time of experts involved in the negotiations. Otherwise experts in specific areas may give insufficient priority to DCFTA issues or 'stiffen' conditions previously presented in a loose form. It is also necessary for Directorate General Trade to have clear negotiating mandates on key issues. When Member States disagree among themselves, the Council may fail to give Directorate General Trade sufficiently clear instructions, making it difficult for the negotiators to give quick and clear responses on key issues to negotiating partners. Furthermore, the Commission now has to report continually on negotiations to the European Parliament as well as to Member States in the Trade Policy Committee (TPC). With 'two masters', it risks having its position on various topics 'revised'. In short, Directorate General Trade needs a sufficiently 'complete' negotiating mandate.

The revitalisation of the BP adds more elements than a DCFTA and will increase the strain on the Commission's resources. This strain could be eased by making more effective use of the secretariat set up by the Union for the Mediterranean.[34]

[34]It has two co-chairs (one from the EU and one from the Southern Coastal States) to symbolise joint ownership. It is manned by experts seconded by member governments. Neither of these factors is conducive to achieving rapid results.

5.2 Increase Financial and Technical Assistance

Most of the ENP countries can ill afford the expense of approximating and implementing parts of the *acquis* necessary for a DCFTA. To postpone this 'crash course' until it is affordable would be penny-wise but pound-foolish. An early start of the Comprehensive Institution Building Programme is necessary given that institution-building and personnel training are investments that take a long time to yield results. While the ENP fared well in the tough EU budget for 2014–2020, the increased level of ambition calls for additional financial resources. The Commission needs to mobilise and coordinate funding with interested Member States' own bilateral assistance programmes and with financial institutions. Friends of the particular partner countries are prepared to provide such long-term bilateral assistance. The EC and interested Member States need to coordinate their respective efforts.

5.3 Common Institutions to Facilitate Conflict Resolution

A key element of the European experience was the importance of common institutions to prevent and resolve conflicts. The ENP is vague concerning the character of the institutions that will administer the deep and comprehensive free trade agreements between the partner countries and the EU and between the partners themselves. This ambiguity concerns which partners are eligible for accession to the EU and also the structure of the proposed Neighbourhood Economic Community which will apply to those countries which are not eligible for accession. The first ambiguity was intentional. The original purpose of the European Neighbourhood Policy in 2004 was to stave off membership applications by offering "anything *but* membership" (italics here). After admitting eight Central and Eastern European States as members in 2004, the EU was not ready to think of new members.

Ten years later, ambiguity concerning the limits of Europe is no longer constructive. It is time for the EU to state explicitly whether or not interested EaP countries are eligible for accession under Article 49, provided they meet the conditionality established by the Copenhagen criteria.[35] Providing an accession track, such as the EU provided for the countries of the Western Balkans, does not mean that accession will take place any time soon, if at all. It will, however, encourage conflict resolution in the region.

Establishment of a common institutional structure for the EU and the relevant Euro-Mediterranean partners can help prevent and resolve conflicts among them.

[35]In 1993, in Copenhagen, the European Council determined that a prospective member state must have stable institutions guaranteeing democracy, the rule of law, human rights, respect for and protection of minorities; a functioning market economy and the capacity to cope with competition and market forces; administrative and institutional capacity to effectively implement the *acquis* and ability to take on the obligations of membership.

Establishment in 2008 of the Regional Cooperation Council as a forum for cooperation between the countries of the Western Balkans is an important role model. The EU has indicated that the proposed Neighbourhood Economic Community (NEC) could be similar to the European Economic Area (EEA) that joins the EU and three EFTA countries in the internal market. The EU now needs to give it greater precision. A strong institutional structure encompassing the EU and the partner countries that have DCFTAs with it and with each other could help to lock in the resolution of regional conflicts.

It is clear that a small country can achieve significant economic benefits from a deep and comprehensive free trade agreement with the huge neighbouring EU market. The rapid progress made by Armenia, Georgia and Moldova in concluding their agreements testifies to this understanding. The EU needs to make it clear to the Euro-Mediterranean partners what common institutions will administer DCFTAs and contribute importantly to resolving and preventing conflicts.

6 Conclusion

This chapter has identified factors conducive to conflict resolution and indicated those countries where the ENP can most likely resolve conflicts.

Several countries in the Balkans have taken important first steps toward greater integration with each other and with the EU. The following factors have contributed to this: The prospect of EU accession has been a powerful incentive to normalize political and economic relations. The Stability Pact for South Eastern Europe, with the European Commission supported by several States, acted as a consistent, persistent and active facilitator of compromise. The potential for intra-regional trade was large and thus economic integration had considerable leverage on jobs and incomes. The Balkan countries had better scores on key indicators for a functioning market economy than the EaP countries (though consistently lower than the central European countries that became EU members in 2004). Serbia long remained unable to accept the sovereignty of Kosovo but is now normalizing relations. With the exception of Bosnia-Herzegovina, prospects are encouraging in the Balkans, but continued support from the European Union remains essential.

The prospects for conflict resolution among the Eastern partners are poor in the short run. These conflicts involve Russia, a third country, making solutions difficult in the context of the EaP alone. The prospect that integration involving Moldova and Ukraine alone could contribute to resolving the Transnistria problem is dim. Two EaP countries do not qualify for a DCFTA (Azerbaijan, Belarus) while Armenia and Ukraine have encountered serious obstacles, making regional integration unlikely in the near future. This means that the EU must be highly selective, focusing its attention and resources on implementing the DCFTAs with Georgia and Moldova while maintaining Armenia and Ukraine in a long-term perspective. Speeding up transition to market economies will encourage economic growth and trade, including eventually trade between the partners themselves. Given their low per

capita incomes and small size, and poorly functioning market economies, these countries will require significant financial assistance from the EU.

Finally, no progress has been made by the Euro-Mediterranean partners in resolving the longstanding conflicts in the Western Sahara and over Palestine. They have made no serious attempt to solve the conflict between Israel and its neighbours, damaging the credibility of the BP. This conflict's past history weighs so heavily that an active facilitator of compromise is necessary. The EU has not assumed this role because of disagreement among its member states. Furthermore, the conflict's international dimension is sufficiently large to require major powers, such as the EU and the USA, to act in concert. On the positive side, there is significant potential for increased trade between Israel, Egypt and Jordan and these countries are on the whole well-functioning market economies with peace agreements with each other. However, significant trade, political and financial incentives appear to be required for these countries to sign free trade agreements with each other. The EU could offer a DCFTA with tariff and quota free access for agricultural products, require any two Euro-Mediterranean partner with DCFTAs with the EU to have one with each other, provide a strong institutional structure (Neighbourhood Economic Community) to manage the network of DCFTAs, and, finally, offer financial resources to fund modernization of the economic and political structures of the BP States and ease the plight of the Palestinian refugees. This is a tall order, but cheaper than the alternative.

The EU and the countries in its Neighbourhood face a truly historic opportunity. After about a century as part of the USSR, some of the EaP partners are now democratic states aiming to establish market economies. The wave of popular demonstrations in North Africa and the Middle East may result in democratic governments there, although the many bloody conflicts are a bad omen. Much is at stake. The population of the Arab countries is now greater than the population of the communist countries of East and Central Europe was in 1989. A new world of opportunities is opening for those that can make the difficult transition to democracy and market economy. A DCFTA is an effective instrument to accelerate the transition. Sooner or later the partners must undertake this transition. It is better that it take place sooner rather than later. The introduction of democratic governance in these countries will be supported by the Partnership for Democracy and Shared Prosperity set up by the ENP. The EU must adopt a long-term perspective and provide substantially greater support if it is to respond effectively to the changes in its neighbourhood. To meet these challenges will require political leaders to display statesmanship not seen since the days of Jean Monnet and George Marshall.

Appendix 1: Selected Indicators for the ENP and Western Balkan Countries

	Country	GDP/cap	Size	Governance	Religion	Independ.	Ethnic groups
Balkans	Albania						
	Bosnia and Herzegowina	8000	4.6	Federal democratic Rep.	Islam 48 %, Serb orthodoxy 35 %,	1992	Bosniaks 48 %, Serbs 37 %, Croats 14 %
	Croatia	18000	4.5	Parlamenary Rep.	Roman Catholicism 88 %, Atheism 5 %	1991	Croats 90 %, Serb 4.5 %
	Kosovo	6600	1.8	Parlamenary Rep.	Majority muslim	2008	Albanians 88 %, Serbs 7 %
	Macedonia	9700	2.1	Parlamenary Rep.	Christian Orthodoxy 65 %, Muslims 33 %	1992	Macedonian 65 %, Albanians 24 %
	Montenegro						
Eastern Partnership Countries	Belarus	13900	9.6	Presidential Rep.	close to 100 Christianity	1991	Belarusian 81 %, Russian 11 %
	Ukraine	6700	45.9	Unitary semi-presid. Rep.	close to 100 Christianity	1991	Ukrainian 78 %, Russian 17 %
	Moldova	3100	3.6	Parlamentary Rep.	close to 100 Christianity	1991	Moldovan 70 %, Ukrainian 11 %, Russian 9 %
	Armenia	5100	3.3	Presidential Rep.	predominantly Christians	1991	Armenian 98 %
	Azerbaijan	10300	9.0	Presidential Rep.	Muslim 99 %	1991	Azerbaijani 91 %
	Georgia	5100	4.6	Unitary semi-presid. Rep.	Orthodox Christianity 94 %	1991	Georgian 84 %, Azeri 6.5 %, Armenains 5.7 %

North-African Countries	Algeria	6900	35.4	Semi-presid. Rep.	Muslim 99 %	1962	Algerian Arabs
	Marocco	4800	32.2	Constitutional Monarchy	Muslim 99 %	1956	Berber 78 %, Arabs 20 %
	Tunisia	9000	10.4	Presidential Rep.	Muslim 98 %	1956	Arab-(berber) 98 %
	Libya	13800	6.4	Dictatorship	Muslim 97 %	1951	Arab and Arab-Berber (predom.)
	Egypt	6400	80.0	Military junta	Muslim 90 %	1922	Egyptians 99 %
	Israel	29500	7.7	Parlamentary Democracy	Jewish 80 %, Muslim 16 %	1948	Jewish 75 %, Arab 21 %
	Jordan	6000	6.4	Constitutional Monarchy	Muslims 92 %	1946	Jordanians, Palestinians, Iraqis
	Syria	5000	22.5	Secular single party state	Muslim 87 %	1946	blend of aramaic speaking Syriac groups

Source: Wikipedia data kindly assembled by T. Y. Mattä, Banque Centrale Luxembourg

Appendix 2: Acronyms

APC	African-Pacific-Caribbean
BiH	Bosnia and Herzegovina
BP	Barcelona Process
CEFTA 2006	Central European Free Trade Area after 2006 (Albania, Bosnia-Herzegovina, Kosovo, Moldova, Montenegro, Macedonia, Serbia). Bulgaria, Croatia and Rumania left CEFTA 2006 upon joining the EU.
CIS	Commonwealth of Independent States established in 1993 contains today nine members. The loose Free Trade Area it established in October 2011 includes Armenia, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, and Ukraine. Georgia withdrew in August 2008 from the Commonwealth following the war with Russia, Ukraine participates but does not consider itself a member.
DCFTA	Deep and Comprehensive Free Trade Agreement
DG	Directorate General
EaP	Eastern Partnership (Armenia, Azerbaijan, Belarus, Georgia, Moldova, Ukraine)
EBRD	European Bank for Reconstruction and Development
EC	European Commission
EEA	European Economic Area (EU, Iceland, Norway, Liechtenstein)
ENP	European Neighbourhood Policy (Eastern Partnership countries and Barcelona Process countries)
EU	European Union (Austria, Belgium, Bulgaria, Croatia, Czeckia, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Malta, Poland, Portugal, Rumania, Slovakia, Slovenia, Spain, Sweden, UK)
FTA	Free Trade Agreement
NEC	Neighbourhood Economic Community
NTB	Non-Tariff Barriers
OHR	Office of the High Representative
OEEC	Organization for European Economic Cooperation, precursor to the OECD
OECD	Organization for Economic Cooperation and Development
PAFTA	Pan-Arab Free Trade Area (Algeria, Egypt, Israel, Jordan, Lebanon, Libya, Morocco, Palestinian Authority, Syria and Tunisia which are members of the Barcelona process plus Bahrain, Djibouti, Iraq, Kuwait, Mauritania, Oman, Palestine, Saudi Arabia, Somalia, Sudan, UAE, Qatar, Yemen)
SAA	Stabilization and Association Agreement
TEU	Treaty of the EU

TPC	Trade Policy Committee
WITS	World Integrated Trade Solution
WTO	World Trade Organization

References

COM (2011) 202, European Commission & High representative of the European Union for foreign affairs and security policy, Brussels 25/05/2011, *A new response to a changing neighbourhood; A review of European neighbourhood policy*.

Gogolashvili, K. (2011). The conflicts in the South Caucasus and their impact on the eastern partnership. *Eastern Partnership Review* No. 9.

Gylfason, T., Martínez-Zarzoso, I., & Wijkman, P.M. (2012, July). *How free trade can help convert the 'Arab Spring' into permanent peace and democracy.* CESifo Working Paper No. 3882.

Gylfason, T., Martínez-Zarzoso, I., & Wijkman, P. M. (2014, June 14). A way out of the Ukrainian quagmire. *VoxEU*.

Gylfason, T., & Wijkman, P. M. (2011a, January 1). *Economic integration as a Balkan peace project. VoxEU*.

Gylfason, T., & Wijkman, P. M. (2011b, January 3). Dayton's unfinished business. *VoxEU*.

Gylfason, T., & Wijkman, P. M. (2014, January 25). Meeting Russia's challenge to EU's eastern partnership. *VoxEU*.

Kaminski, B., & de la Rocha, M. (2002), Policy induced integration in Balkans: Policy options and their assessment. In *Trade policies and institutions in the countries of South Eastern Europe in the EU association and stabilization process.* World Bank.

Kowalczyk, C. (2000). Welfare and integration. *International Economic Review, 41*(2), 483–494.

Kowalczyk, C., & Davis, D. (1998). Tariff phase-outs: Theory and evidence from GATT and NAFTA. In J. Frankel (Ed.), *The regionalisation of the world economy*. Chicago: University of Chicago Press.

Messerlin, P., Emerson, M., Jandieri, G., & Le Vernoy, A. (2011), *An appraisal of the EU's trade policy towards its eastern neighbours; The case of Georgia.* Group d'Economie Mondiale, Science Po, Paris, and Centre for European Policy Studies, Brussels.

Monnet, J. (1976), *Mémoires*, Librairie Arthème Fayard.

Montesquieu. (1748), C.-L. de S., *On the spirit of the laws* (*De l'esprit des lois*), Book XX.

Sekarev, A. (2011).Medium-term trade flows between EU27 and EaP6: What trends are visible? *Eastern Partnership Review* No. 10.

Wijkman, P. M. (2011), Fostering deep and comprehensive free trade agreements for the eastern partners. *Eastern Partnership Review* No. 8.

The European Union's Role in the Development of Global Environmental Law

Ellen Margrethe Basse

Abstract This article is focusing on the role that the European Union (EU) is playing as a global actor when it is using its 'internal' regulatory power as well as its parallel market power 'externally' to strengthen and to fill out gaps in the global environmental law. The EU is at the forefront of international efforts to reduce greenhouse gas (GHG) emission as one of the most important global challenges, and the mentioned regulatory role of the EU is illustrated in this article by three examples related to GHG emissions from transport: The EU's use of the 'cap and trade' principle on all aircrafts that arrive or depart from its territory; the EU's proposal on the use of monitoring, reporting and verifications (MRV) obligations on all ships that arrive or depart from its territory; and the EU's use of mandatory sustainability criteria on all biofuels used by road traffic. These examples involve transnational regulation of private sector actors. The EU ambitions are to ensure a model for global norms in situations, where the international regimes are ineffective. The EU legislator has the European Court of Justice's word for its right to permit a commercial activity to be carried out by foreign enterprises only on its territory provided that the operators comply with the EU norms; and it has the Court's word for its right to act with exclusive external competences to ensure an export of the norms also in situations where it is the Member States of the EU—and not the EU—that are the accepted parties of the international organisation that the EU wants to push.

Abbreviations

AGP	The Agreement on Government Procurement
Annex I parties	Developed countries
Annex II parties	Developed countries in transition from a plan economy to a market economy
BAP	The Bali Action Plan

E.M. Basse (✉)
Department of Law, Aarhus University, Aarhus BSS,
Bartholins Allé 16, 8000 Aarhus C, Denmark
e-mail: ellenmb@law.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_17

CAP	Common Agricultural Policy
CBDR/RC	Common but differentiated responsibilities and respective capabilities
CDM	Clean development mechanism
Commission	The European Commission
COP	Conference of the Parties
Council	The European Council
Court	The European Court of Justice
EEDI	Energy Efficiency Design Index
ETS	Emission trading scheme
ETS Directive	The EU's Directive on Emission Trading (2003/87/EC as amended)
EU	The European Union
EU ETS	The EU's Emission Trading Scheme
GATT	The General Agreement on Tariffs and Trade
GHG	Greenhouse gas
ICAO	The International Civil Aviation Organisation
ILUC	Direct and indirect land-use changes
IMO	The International Maritime Organization
IPCC	The Inter-governmental Panel on Climate Change
ISPS Code	The International Ship and Port Facility Security Code
IUCN	The International Union for the Conservation of Nature
JI	Joint implementation
LDC	Least developed countries
MARPOL 73/78	The Convention for the Prevention of Pollution from Ships
MBM	Market-based measures
MEPC	The Marine Environmental Protection Committee of the IMO
MFN	Most favoured nations
MRV	Monitoring reporting and verifications
NAMAs	Nationally appropriate mitigation actions
NMFT	No more favourable treatment
Non-Annex I parties	Developing countries
RED	The Directive on Promoting Renewable Energy (Directive 2009/28/EC)
SEEMP	The Ship Energy Efficiency Management Plan
SIDC	Small islands developing countries
SOLAS Convention	The International Convention for the Safety of the Life at Sea
TBT Agreement	The Agreement on Technical Barriers to Trade
TEU	The Treaty on the European Union
TFEU	The Treaty on the Functioning of the European Union
Treaty	The Treaty of Lisbon
UN	The United Nations
UNEP	The United Nations' Environmental Programme

UNFCCC	The United Nations Framework Convention on Climate Change
US	The United States of America
WCED	The World Commission on Environment and Development
WTO	The World Trade Organization
Aarhus Convention	The UNECE Convention on Access to Information Public Participation in Decision Making and Access to Justice in Environmental Matters

1 Introduction

Climate changes represent one of the most serious threats to the global society and the absence of effective traditional solutions to mitigate has proved to be among the central challenges of globalization. The characteristics of climate changes create unique policy challenges, as nearly the full range of human activities is associated with GHG emissions. Trade liberalization and the dramatic growth of industrial production, transport and trading activities have facilitated the growth of global markets, the integration of national economies, the problems of ecological and economic interconnectedness, and the challenges of climate changes. Globalisation is changing the nature of institutions and their role and consequences for economic or societal performance. Globalization is also positively influencing the public perceptions, popular norms, and governmental views of how environmental issues—including climate changes—should be properly addressed by the design of law. Although international environmental law historically has been focused mainly on inter-state relations and the rights and duties of sovereign states, new approaches and new actors are accepted as a consequence of globalisation.

Since the early 1990s, the European Economic Community (now the EU) has established itself as an international leader in environmental governance pushing especially for stringent global commitments on GHG mitigation under the United Nation's Environmental Programme (UNEP) and outside this regime. The institutional set-up is a challenge for the EU, as the IMO Convention and the Chicago Convention establishing two agencies do not accept the EU as a party. As a 'regional economic integrated organisation' the EU is exercising a strong power of influence especially in the cooperation on global commitment, e.g. under the United Nations (UN)—directly as a party to multilateral agreements, and indirectly through its Member States in the regimes, where the EU has not been accepted as a party. This article is explaining the role of the EU in the emerging 'global environmental law' with focus on the EU's use of 'internal' regulatory power and its global market positions to fill out gaps in the multilateral environmental agreements on measures related to the challenges of climate change.

The traditional, legal theory on state-based legal order and the fragmentation of law is shortly presented in Sect. 2. The phenomenon of fragmentation of law is an important consequence of globalisation and an important reason for the

development of new legal theories. The general issue related to environmental law as a relatively new legal discipline and the ideas on the emerging 'global environmental law' is presented in Sect. 3. In Sect. 4 it is followed by a presentation of the legal competences of the EU, where it is illustrated how the European Court of Justice takes part in the designing of a new legal order. The EU environmental law is increasingly interrelated, reflecting: its role in multilateral agreements; its use of its internal regulatory power to push the international negotiations on new measures; its vertical harmonization of international and national law; and its ambition on a more holistic approach used in the legal design is described. The involvement by the general society in the gap filling and vertical harmonization *inside* the international, the EU and the national legal regimes is, as illustrated *infra* in Sect. 4.4., ensured by the implementation of a multilateral agreement on access to environmental information, a broader participation of the 'public concerned' (especially NGOs) in decision-making processes, and by access to justice for the 'public concerned'. A general presentation of the multilateral agreements in the ordinary climate change regime under the UN, where the EU is accepted as an ordinary party, is the content of Sect. 5. Section 6 is supplementing this presentation with a description of the negotiations on measures covering GHG emissions from international aviation and shipping, respectively, in two special agencies under the UN—the International Civil Aviation Organisation (ICAO) and the International Maritime Organization (IMO)—that do not accept the EU as a member. The mentioned difference between the approaches to the EU as a party versus not being accepted as a party in international negotiations is one of several examples on the fragmentation of law. Three examples of the EU's internal regulatory initiatives related to the transport sector with extraterritorial effects are described in Sects. 7–9. The trade aspects of such regulatory initiatives are discussed briefly in Sect. 10, and the concluding remarks are stated in Sect. 11.

2 Traditional State-Based Law: Or?

For roughly the last 200 years, law has been considered a product of the democratic national political process with the powers, rights and obligations laid down in a constitution (constitutionalism). The established theoretical and doctrinal premises of legal thinking are still bound to the concept of sovereignty of national states and provides mostly for a national-state perspective. The 'Westphalian system' with its clear separation between national and international levels of law and only relative thin forms of coordination in the latter is still part of the most accepted legal theory. The formal 'international law' that is based on the mentioned theory consists of multilateral environmental agreements (conventions and protocols) and declarations formed as a result of voluntary agreements between states as well as universal usage of practices that are considered by states to be customs that are legally binding upon them (customary international law). The Vienna Convention on the Law of the Treaties, agreed by the national states in 1969, deals with the interpretation of agreed

multilateral agreements. The authoritative catalogue of sources of international law is found in article 38 of the Statute of the International Court of Justice established by the Charter of the UN. In most national legal systems (for example Germany, the Netherlands, Japan, Russia, Denmark, Norway and Sweden), international law is considered a body of law distinct from national law. Such legal systems that require the intervention of the national democratically elected legislator before an international multilateral agreement, which can be transformed and applied into national law, are based on the dualism-approach. Other legal systems are based on monism (for example France, the USA, Indonesia and South Africa). They have constitutional provisions that allow for a specific class of multilateral agreements to be applied self-executing, which means that they do not require, as a transformative step into national law, any intervention of the legislator. They view international and national laws as interconnected as a single system (Rose 2011 at p. 6).

The legal order of the EU is different from this traditional approach to law. The EU, which is not viewed as an international regime (confederation of sovereign states), is neither a federation, which shares its sovereignty with its Member States, nor a nation state. The EU has a constitutional basis (now the Treaty of Lisbon) regulating powers, rights and obligations of the EU institutions, Member States, and citizens. The concept of shared sovereignty as well as the exclusive external competences for the EU institutions explained *infra* in Sect. 4 are contrary to the historical views of Westphalian sovereignty. The EU legal order differs from traditional international law due to its body of autonomous law-making institutions and especially due to the role of the European Court of Justice, which is not an international court in the traditional sense and has not limited itself to applying the law in a traditional legal dogmatic sense (Sommer 2013, at p. 121). The legal authority of the European Court of Justice is like that of a supreme court enforcing federal constitution (Edgeworth 2002, at p. 179) it was stated by the European Court of Justice (In Case 2662 *Van Gend en Loos*) that the EU constitutes a new legal order in international law, for whose benefit the States have limited their sovereign rights, albeit within limited fields (Sommer 2013, at p. 119). By this judgement—and several latter judgements—the Court has developed a special legal order that has consequences for the competences of the Member States and for the traditional legal theory.

There is also another difference between the traditional legal state-based approach and the approach in the legal order of the EU: The traditional legal separation between public and private law is not a part of EU law (Semmelmann 2012, at pp. 195–196). The Treaty is using the terminology 'public' versus 'private', but does not contain an express the distinction between public versus private institutions or the distinction between private versus public law (Nielsen 2012, at pp. 105–106; and Semmelmann 2012, at pp. 200–206). It is not the public power of the authorities versus the private actors that is the focal point in the EU legal order—rather it is the access to a market without the states disturbing the market forces more than necessary.

The consequences that European integration has on the position of international requirements in the Member States, after the rules have entered the national legal

order via the EU legal order, see *infra* in Sect. 4, illustrate the emergence of a complicated triangular relationship between international law, EU law and national law.

Several political science and legal scholars—including the International Law Commission under the—UN's General Assembly (Koskenniemi 2006)—have started studies on the fragmentation of law, the challenges for national as well as international law, and the question of 'self-contained' regimes as a consequence of the globalization. It is described how state-based legislation is becoming less prominent, and rulemaking by multilateral agreements between states is in retreat as a consequence of globalisation (Barendrecht et al. 2012). As the divide between national and international law becomes still more blurred, the theory of pluralism in law is developed both on analytical and on normative grounds. In the theory on the emerging of global environmental law it is argued that a supplementary involvement of new actors—including market actors—is filling out the gaps in the traditional law, see *infra* Sect. 3. As described below, the EU is acting as such an actor by exporting its norms into multilateral agreements (Sommer 2013, p. 168, 176, 198, 215, 223 and 266) and pushing the international community to accept its internal norms as market conditions.

3 Environmental Law as a New Discipline

Environmental law is the area of law that seeks to manage human impacts on the environment with respect to the use of resources, water and ecosystem services, reduction of pollution and reduction of emission of GHG caused by human activities. Whereas mainstream legal disciplines (e.g. constitutional law, contract law, commercial law, tort law and administrative law) deal only with human-human interactions (e.g. government-citizen interactions), environmental law deals with the interactions among humans and the other several million species on the planet, in complex interconnected systems, trans-boundary interactions and over long-term horizons. In other words, environmental law is concerned with the inextricable connection between humanity and the common exterior world—the global environment as a whole and its component parts, dealing with the complex and uncertain risks, with shared externalities and diverse sources, and with hard trade-offs between competing values. The word 'law' used in environmental law refers broadly to those norms and principles that lay out behavioural prescriptions issued at international, regional or national levels by an authoritative source in a written form, whether or not mandatory or backed by a dispute settlement or an enforcement system.

Environmental themes have recurred through different generations of regulatory designs. It started in 1972 at the Stockholm Declaration with a presentation of ideas for the simplification of rules for national states in order to implement general principles of human rights and obligations towards nature, the environment and future generations. In its report of the World Commission on Environment and

Development, titled 'Our Common Future', the Brundtland commission is recognising the consideration of ecological/resource dimensions alongside economic and social dimensions, stating that the interdependency of the new challenges and issues today contrast sharply with the nature of the institutions that exist:

> These institutions tend to be independent, fragmented, and working to relatively narrow mandates with closed decision processes. Those responsible for managing resources and protecting the environment are institutionally separated from those responsible for managing the economy. The real world of interlocked economic and ecological systems will not change: the policies and institutions must. (WCED 1987 at p. 310)

The 'concept of sustainability' addresses the need to balance and coordinate divergent collective interests, while protecting the regenerative capacity of nature in a concerted effort to synthesize and integrate environment and development issues. Environmental law of today is—based on the rationales of sustainable development—focusing on the merging of environmental and economic interests and consequences in the decision-making systems of legislative actors at different regulatory levels as well as the interaction between legal actors and private actors. Achieving sustainability is different from other economic criteria as the recognition of the present generation's responsibility to future generations by this concept has the implication that economic growth must respect the finite capacity of the natural environment both in terms of climate changes, exhaustible natural resources and the limited capacity of ecological functions (Voigt 2006, at p. 67). The concept also means a shift away from the traditional thinking based on state sovereignty in international law towards cooperative and inclusive governance structures founded upon a sense of responsibility for humanity's common future (Voigt 2006, at p. 21). Based on this concept, the protection of the global atmosphere, temperature, biodiversity, hazardous products and waste etc. is a common responsibility for all states (Birnie et al. 2009, at p. 128 ff.). The role of modern environmental law is both an issue relating to ensuring the competitive position of the private sector actors in global trade and an issue relating to environmental regulation and reduction of the negative impacts of global trading, transport etc. The three examples described in Sects. 7–9 show the many interests involved in environmental regulation of transport.

It has been argued that a 'global environmental law' is being developed in parallel with the international and the national environmental law (Yang and Percival 2009, at pp. 616f., 623, 626 and 635f.). As the traditional multilateral agreements are probably not the most effective solution, a supplementary global linkage between markets can be a necessary means to ensure environmental self-regulation by the actors responsible for the negative consequences of trade and production (Yang and Percival 2009, at pp. 642 and 645). Robert V. Percival describes the development leading to the emergence of 'global environmental law' in this way:

> Globalization is affecting law and legal systems throughout the world in profound new ways. With the growth of global concern for the environment, nations are transplanting environmental law and policy innovations even from countries with very different legal and cultural traditions. Private actors and nongovernmental organizations are driving the development of new legal and nonlegal strategies to protect the environment. These

> developments are blurring lines that traditionally separated conceptions of domestic and international law and public and private law. This is leading to the emergence of what I have called "global environmental law". (Percival 2010, at p. 37).

As mentioned above, the intention with this article is to explain the role of the EU in the emerging global environmental law. It will be done by explaining the competences of the EU, and the judgements of the European Court of Justice in Sects. 4 and 6, the acting of the EU inside and outside multilateral negotiations in Sects. 5 and 6, and the three examples in Sects. 7–9 on the EU's internal regulatory initiatives with external consequences in areas, where access to the European market is of importance for enterprises in non-EU Member States.

4 The Legal Competences of the EU

The constitutional basis for the EU is the Treaty of Lisbon that includes the Treaty on the European Union (TEU), the Treaty on the Functioning of the European Union (TFEU) and several protocols. Article 3(3) TEU states that the Union shall establish an internal market, and it shall work for the sustainable development of Europe based on balanced growth. The intent behind the Union's policy is to facilitate sustainable development by use of means respecting trade as well as environmental policy. The principle of integration is very important as the guiding principle for all the EU's policy based on the concept of sustainability. Article 11 TFEU states that

> Environmental protection requirements must be integrated into the definition and implementation of the Union policies, in particular with a view to promote sustainable development.

The competences of the EU have two dimensions: the relation between the EU and its Member States ('internal dimension'); and EU's external relations ('external dimension').

4.1 EU's Internal Dimension between the EU and its Member States

The principle of conferral is a fundamental principle of EU law stated in Articles 4(1) and 5(1)–(2) TEU and Article 4(1) TFEU. Under this principle, the EU is a union of Member States, and those Member States voluntarily confer all their competences on it. Within the areas of environmental, climate and energy issues, in Article 4(2) the TEUF contemplates 'shared competences' between the Member States and the EU institutions. A shared competence implies that both the EU and its Member States may legislate and adopt legally binding rules. The EU institutions must pass two tests, when they are regulating in such policy areas. *Firstly*, they have to demonstrate that the objectives of the initiatives cannot be sufficiently

achieved at the national level, and *secondly* they should demonstrate that the actual initiative by reason of its scale or its effects can be better—e.g. more effective, more democratic, more consistent with the internal market, more consistent with international obligations, more environmentally friendly etc.—achieved at EU level than at national level. On the other hand, the Member States shall only exercise their competence to the extent that the EU has not exercised, or has decided to exercise, its competence. The EU has exclusive competence in all areas as to the conclusion of an international or regional agreement when its conclusion is provided for in an environmental legislative act of the EU or is necessary to enable the EU to exercise its internal competence, or in so far as its conclusion may affect common rules or alter their scope, see Article 3 TEU. When the EU has passed secondary environmental law, which is the case in relation to the use of the regulatory initiatives described *infra* in Sects. 7, 8 and 9, all subject matters of such secondary law fall within the scope of the exclusive external EU power—also in situations where the international agencies do not accept the EU as a member. As a consequence of the mentioned power, the international agreements concluded and other external initiatives taken by the EU in such areas are an integral part of the EU legal order. The Member States have no right to enter into international or other commitments, which might affect or alter the scope of 'internal' EU rules and the external negotiations by the EU.

The environmental rule in the treaties—Article 192 TEUF—is the legal basis for most environmental legislation. It is the legal basis used by the EU in the legislation on the 'cap and trade' principle described *infra* in Sect. 7, and in the legislation on the monitoring, reporting and verification (MRV) obligations described *infra* in Sect. 8. By Article 193 TEUF the Member States are expressly permitted to take more stringent national environmental measures in the areas regulated by this legal basis, as long as they respect the requirements of the treaties.

Distortion caused by competition between undertakings of the various Member States can be reduced by means of a strict uniformity (or total harmonization) in the secondary EU legislation by use of the treaties' rule on the internal market, currently Article 114 TEUF, as the legal basis (Jans et al. 2012, at p. 59ff.). The EU has used this legal basis in the regulation of biofuels described *infra* in Sect. 9. It means that Member States may not exclude or include products on other sustainability grounds than the measuring compliance with the sustainability criteria.

4.2 EU's External Dimension and the Member States Duty of Loyalty

The Treaty of Lisbon is formulating the principles and objectives of the EU's external actions, and it confirms the legal personality of the Union, which is in effect an international personality, cf. Article 47 TEU. It is stated in Article 3(5) TEU that one of the aims of the EU is to promote its values globally supporting the role of it

as a normative global power. Among the general provisions on its external actions is the obligation to respect the principles, which have inspired its own creation, including the principle of sustainability, see Article 21 TEU.

It is the European Commission (Commission) that is negotiating, and the European Council (Council) that makes the final decision on international commitments, unless secondary EU legislation accepts that the Commission has the competence to make the final decision—or the international regime does not accept the EU as a party. Once the Council has concluded a multilateral agreement, its legal effects are twofold: the EU as well as its Member States are bound, as the ratification by the EU has binding effects under Articles 216(6) and 288 TEUF. Both the Commission (representing the EU) and the individual Member States (represented by their governments) are represented in the international negotiations in the regimes accepting them as parties, even though they largely act jointly and are recognized as one unitary actor. Article 191(4) TEUF provides, that

> Within their respective spheres of competences, the Union and the Member States shall cooperate with third countries and with the competent international organisations. The arrangements for the Union cooperation may be the subject of agreements between the Union and the third parties concerned.
>
> The previous subparagraph shall be without prejudice to member states' competence to negotiate in international bodies and to conclude international agreements.

Both the UN's Climate Convention (UNFCCC) and its Kyoto Protocol have the EU and all EU Member States as parties. But not all the international regimes under the UN accept the EU as a party. According to the principle on external competence developed by the European Court of Justice in Case C-22/70 *Commission v. Council (the AETR-case)*, the EU acquires external competence, also in situations where it adopts internal EU rules on the same subject matter that is discussed on transport issues in an international agency that do not accept the EU as a party, see *infra* in Sect. 6.

4.3 Vertical Harmonisation of International, EU and National Law

The multilateral agreements ratified by the EU—as well as the agreements ratified by its Member States under IMO and ICAO, see *infra* in Sect. 6—are transformed by the EU into European and national law either by strategies, directives or regulations. Whilst individual Member States are responsible for the implementation of EU directives, the EU regulations are directly applicable and therefore form part of national legislation, cf. Article 288 TEUF. Transposition of directives is to be made within the context of the principle of loyalty, cf. Article 4(3) TEU, and the national regulatory traditions. Basic principles and doctrines, such as the supremacy of EU law, direct effect of some rules laid down in directives, and the state responsibility for failure to implement EU law have been elaborated by the European Court of

Justice and are now important elements in the description of the legal basis for the Member States' national environmental law. The Member States must ensure an effective implementation at national level and, if necessary, establish effective, dissuasive and proportionate sanctions in order to implement and enforce the law.

In its judgments the Court of Justice has held that under certain circumstances a provision in a multinational agreement concluded by the EU may be directly applicable in the Member States, see for example Case C-213/03 *Syndicat professionnel coordination des pêcheurs de l'étang de Berre and de la region v. Èlectricité de France* and Case C-239/03 *European Communities v. France*. Where a lack of direct effect gives room for discretionary interpretation, the duty of loyal interpretation based on Article 4(3) TEU has to be taken into account. Based on this, the Member States have a duty of consistent interpretation of international obligations ratified by the EU as stated in the Case 14/83 *van Colson*. In order to prevent conflicts, the primacy of the EU's international obligations over secondary EU legislation must be interpreted as far as possible, in a manner that is consistent with those agreements (Sommer 2013, at p. 127). Thus, the provision of a multilateral agreement can become part of national legislation without the national transformation of it into legislation—also in Member States with a legal system based on the dualistic approach, cf. *supra* Sect. 2.

Article 17 TEU states that the Commission shall oversee the application of international and EU law under the control of the European Court of Justice. In cases of suspected non-implementation, the Commission initiates Articles 258 and 260 TEUF infringement proceedings with financial sanctions.

4.4 Involvement by the General Society Inside the Legal Regimes

The UNECE Convention on Access to Information, Public Participation in Decision Making and Access to Justice in Environmental Matters (Aarhus Convention) from 1998—that has the EU and all EU Member States as parties—is the most important regional regime ensuring a vertical harmonisation of international, EU and national law through procedural rights. The Convention marks a definitive step forward in the field of environmental governance—and the emerging of a global environmental law—as it seeks to protect the right of every person of present and future generations to live in an environment adequate to his or her health and wellbeing. To do so, it sets minimum standards empowering members of the 'public concerned' to call public authorities to account by providing tools for the following three pillars: (a) access to environmental information; (b) participation in decision-making; and (c) access to justice in matters relating to the environment. The term 'public concerned' means: the public affected or likely to be affected, or having an interest in, the environmental decision-making, including the nongovernmental organisations. The rules on access to justice cover e.g. situations where there are

significant problems with enforcement of the international, EU and/or national environmental legislation.

The EU has implemented these procedural rules by two directives and one regulation: the Directive on Public Access to Information (2003/4/EC); the Directive on Public Participation (2003/05/EC); and the Regulation of the Application of the Provisions of the Aarhus Convention to the EU's institutions and bodies (1367/2006/EC). The actual preparation of a proposal on a new Directive on Access to Justice will ensure that also the third pillar of the Convention is transformed into EU law. The enforcement of international environmental law as well as the enforcement of EU environmental law at the national level is ensured by litigations at the Compliance Committee under the Convention as well as litigations at the European Court of Justice.

In cases where the international provisions do not contain any clear and precise obligations capable of directly ensuring the legal rights of the 'public concerned', the national courts in the Member States must interpret their national law in accordance with the objectives of the international provision to the fullest extent possible. The European Court of Justice stated such an obligation by its judgement in the Case C-240/09 concerning access to justice under the Aarhus Conventions Article 9(3). In this way the EU takes part in the filling up of gaps in environmental law by involving also private actors in the vertical harmonisation of environmental law. As is apparent from well-established European Court of Justice case law, the detailed procedural rules governing actions for safeguarding an individual's rights under EU law must not be less favourable than those governing similar national actions (the principle of equivalence) and must not make it impossible or excessive difficult to exercise rights conferred by EU law in practice (the principle of effectiveness).

5 UNFCCC/Kyoto Protocol and the Role of the EU

Internationally, 195 parties—including the United States (US), China and the EU—joined the UN's Framework Convention on Climate Change (UNFCCC) in 1992, and some of them have joined the Kyoto Protocol under the Convention in 1997. It is only the protocol that is imposing binding obligations on its parties. Some left the protocol on 1st January 2013.

The UNFCCC contains general environmental law principles and declarations without providing specific and precise obligations for the signatory parties. The ultimate objective is to achieve stabilization of GHG concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system. It has set up what is known as the COP (Conferences of the Parties) as its supreme authority meeting. In formal terms the negotiations take place in the COP that meet once a year in December, and each meeting is given a number. In December 2010, the COP recognized that global warming must not exceed 2 °C.

5.1 The Difference Between the Parties to the UNFCCC

The parties to the UNFCCC are currently divided into two main categories: Annex I parties (consisting of the 43 industrialized countries and countries in transition from a plan economy to a market economy); and others (consisting of 152 developing countries). There are 24 parties to the UNFCCC listed in Annex II (industrialized countries)—all of them are Annex I parties.

The Non-Annex I parties are also differentiated for the purpose of economic help into least developed countries (LDC) and small islands developing countries (SIDC) on the one hand; and the other Non-Annex I parties (China and other emerging economies) on the other hand.

5.2 Common but Differentiated Responsibility and Respective Capabilities

It is only the Annex I parties that have obligations regarding GHG emission reduction. They decide themselves the manner and the speed upon which they implement their obligations as long as they achieve the target set in Annex B to the Kyoto Protocol. The Inter-Governmental Panel on Climate Change (IPCC) assumes in its 4th Assessment Report that the Annex I parties must reduce the emission by 25–40 % no later than 2020 and by 80 % no later than 2050 (95 % in relation to the 1990 level) to keep the average temperature from rising more than 2 °C.

Working on compromises between the differences, the leading global principles of 'common but differentiated responsibilities and respective capabilities' (CBDR/RC), 'equity', and 'common concern' are used in the UNFCCC/Kyoto Protocol to ensure compromises between the developed and the developing countries. UNFCCC Article 3(1) states:

> [t]he Parties should protect the climate system for the benefit of present and future generations of humankind, on the basis of equity and in accordance with their common but differentiated responsibilities and respective capabilities. Accordingly, the developing country Parties should take the lead in combating climate change and the adverse effects thereof.

The concept of 'common concern' recognizes the atmosphere as a common resource—a global unity—whose protection is of common interest to all states, regardless of state sovereignty (Voigt 2006, at p. 101; and Birnie et al. 2009, at p. 129). The concept CBDR/RC is establishing the obligation to work in a spirit of global partnership. Under this principle the Annex II parties have a special responsibility in two respects: an implicit legal responsibility to others because of past and current damaging acts, and an equitable responsibility to use their technical and financial capacity to alleviate the world's environmental and development problems. The EU and its Member States are all Annex I parties; the 15 old Member States are Annex II partners. The LDC and SIDC deserve special consideration in

how that responsibility applies to them and special assistance in fulfilling it (Basse and Gaines 2010).

A relatively small number of countries (25 countries) produce a large majority (83 %) of global GHG emissions. China and the US are among the leading emitters of GHGs, together they are responsible for about 42 % of global emission. China is actually the biggest emitter of GHG in the world. As a non-Annex I party, China has no binding emission limits, and current forecasts foresee that its emissions will grow rapidly in line with its economy. Even though China is still considered a developing country in international environmental law, and consequently has a favourable position under the UNFCCC/Kyoto Protocol, it has grown rapidly in recent years and has become an important economic payer in the global world. China has pointed out that the Annex I parties have used a far larger proportion of the common atmosphere than they are entitled to—so they should surrender some space to the non-Annex I parties.

A reference to historical responsibility is found in the preamble and Article 4(1) of the UNFCCC, and in Article 10 of the Kyoto Protocol. Several analyses have, however, confirmed that the most developed countries do not undertake such a responsibility today. The US is an Annex I party, but not a party to the Kyoto Protocol, and is therefore without binding obligations to reduce the emission as the world's second-largest emitter of GHG. The US—as well as Canada, Japan, New Zealand and Russia—will not accept that major and better-off economies, such as China, should be permitted unlimited economic growth without accepting any GHG reduction commitments, and they want to redefine the word 'common' as well as the word 'developed' in the CBDR/RC principle. They are questioning the relevance of the dichotomy among countries as developed (Annex I parties) versus developing (Non-Annex I parties) that is embedded in the UNFCCC.

The EU has joined the LDC and the SIDS in an argumentation for the need for a strong and quick initiative in the interest especially of the LDC and the SIDS.

5.3 'Cap and Trade' and Measures, Reported and Verified GHG Emissions

The commitments on legally binding mitigation obligations for the Annex I parties made in the Kyoto Protocol were set-up at the 'Marrakech Accords' at the COP7 held in 2001 and provide for the flexible market-based measures (MBM) covering: the global emission-trading scheme (ETS); the joint implementation (JI) by use of agreements on emission reduction projects between Annex I parties; and the clean development mechanism (CDM) as agreements on emission reduction projects between an Annex I party and a Non-Annex I party. Under the JI and the CDM agreements the difference between the (high) emission of GHGs that would have resulted from the old technology (coal-fired plant, for instance) and the lower emissions resulting from producing the same amount of goods and services

(especially energy) using the new plant (water or wind power, for instance) is calculated and converted into carbon dioxide (CO_2) credits. The CO_2 credits form part of the global ETS market in which emission rights (credits) are transferred between parties.

The strength of a global 'cap and trade' system is that it places an absolute limit on the total amount of emissions permitted to occur. A 'cap and trade' mechanism, which is based on the idea of issuing credits on the basis of a given sector's reduction of GHG emissions—for example for the international aviation and shipping industries—is among the discussed possibilities under the UNFCCC/Kyoto Protocol. The main object of the sector-based 'cap and trade' scheme is to initially cover sectors that have a large reduction potential; sectors that are most exposed to carbon leakage; and sectors that significantly affect production costs of sectors exposed to carbon leakage. Carbon leakage arises when a carbon price causes national business to relocate to countries not pricing carbon or to increase imports of goods and services from countries not pricing carbon, resulting in no net reduction in global CO_2 emissions.

The 2007 Bali Action Plan (BAP) adopted at the COP13 called for the GHG emissions from all parties to be measured, reported and verified (MRV) to move further in the mitigation of climate changes. At that time the MRV elements were already elements in place under the UNFCCC/Kyoto Protocol. The Copenhagen Accord—agreed on at the COP15 (Decision 2/CP.15) in 2009—supplemented the requirements on MRV by stating that MRV of Annex I parties' targets and finance will be 'rigorous, robust and transparent'; and for Non-Annex I parties the nationally appropriate mitigation actions (NAMAs) will undergo 'international consultations and analysis' under guidelines respecting national sovereignty. It is a condition that the design of the MRV system must respect the CBDR/RC principle, as well as variation in the types of parties' actions. In the Cancun Agreements at the COP17 in South Africa, the LDCs and SIDCs received additional flexibility on reporting.

5.4 The EU and its Member States as Parties to the UNFCCC/Kyoto Protocol

As mentioned *supra* the Convention is accepting the EU as a party acting together with its Member States, cf. Article 1(6) of the Convention. Article 4 of the Kyoto Protocol—which became known as the 'EU Bubble'—allowed the EU and its Member States to meet their targets jointly through a differentiated commitment among the Member States. Part of the means used by the EU as an Annex I party in the comprehensive package of measures is the abovementioned MBM solutions established by the Kyoto Protocol as a flexible mechanism.

The EU is one of the most active parties to the UNFCCC and the Kyoto Protocol. It also proclaimed a leadership role under the Convention in December 2007 in Bali (Indonesia), COP13, where the COP resulted in the adoption of the Bali Action Plan (BAP). The plan had a deadline for concluding negotiations on the post 2012 regime at COP15 (December 2009). However, it did not result in binding mitigation obligations. Following the limited outcome of COP15, the European Parliament in its resolution of 10 February 2010 called for climate diplomacy under the UNFCCC involving the more progressive developing and emerging countries in the promotion of new mitigation obligations by individual activities. The first commitments of the international community under the Kyoto Protocol are expected to lead to a 3.5–4 °C temperature rise by 2050, and not the IPPC target on average global temperature increase below the 2 °C above pre-industrial level. The EU's overall aim is notably to solve this so-called 'gigatonnes gap'.

The EU is still the most proactive party to the UNFCCC, but not strong enough to push the development of a post 2012 regime. At COP17 in Durban, the EU agreed to a second commitment period for the Kyoto Protocol together with a few other Parties. At COP18 in Doha in December 2012 the second commitment period was agreed upon by the EU, Norway, Switzerland, Australia, Monaco and Liechtenstein, allowing the Protocol to move forward for the period 2013–2020. These parties are revisiting and increasing their commitments by 2014 in line with the 25–40 % emissions reductions called for by IPCC in its 4th Assessment Report.

5.5 EU Internal Initiatives and Instruments

Climate changes have become a major agenda item regularly discussed by the heads of states and governments in the EU (Oberthür and Kelly 2008, at p. 35). The European Environmental Agency is also following the GHG emission trends and projections in the member states (EEA Report, no. 4/2011).

The EU's Climate and Energy Package from 2008/2009 is the most important 'internal' initiative stating that the EU shall achieve at least a 20 % reduction of the GHG emissions below the 1990 levels by 2020. The EU's Emission Trading Directive (2003/87/EC as amended—hereinafter referred to as the ETS Directive) establishes a 'cap-and trade' scheme (EU ETS), making it mandatory for all Member States to reduce CO_2 emissions. In January 2005 it commenced operations as the largest multi-country, multi-sector GHG trading system in the world. It identifies certain types of industrial entities within the EU that are required to pay by means of CO_2 allowances for their emissions. CO_2 allowances are exchangeable, and may be bought and sold in all EU Member States under the 'EU Bubble'. Since the amendment of the ETS Directive in 2005 (the Linking Directive 2004/101/EC), the JI and CDM credits have also been accepted as part of the scheme. The prices for CO_2 allowance and JI and CDM credits are determined by supply and demand. The aviation industry has been included in the EU ETS since 2010, see *infra* Sect. 7.1.

The current version of the EU ETS (Directive 2009/29/EC) is established as one of the elements of the 'Climate and Energy Package'. The Directive is establishing the legal obligations under the third commitment period under the EU ETS, covering the period 1 January 2013–31 December 2020. The auctioning is now the basic principle of allocating allowances. This means that EST businesses have to bay an increasing proportion of allowances through auction. In 2013, 40 % of all allowances will be auctioned. The current EU ETS has introduced special mechanisms to mitigate the risk of carbon leakage—situations where enterprises shift production or investment outside the EU because they are unable to pass on with their business in the EU as a consequence of cost increases. Some ETS sectors are therefore provided with free allowances in order to assist their transition towards a low carbon economy. The 'EU Bubble' will reduce the number of available allowances by 1.74 % each year, delivering an overall reduction of 21 % below 2005 verified emissions by 2020. For this commitment period and onwards, the MRV under the EU ETS will be required to comply with two Commission Regulations, one specific to monitoring and reporting, and the other to verification and accreditation.

The EU has also set binding emission targets for non-ETS sectors (e.g. transport, agriculture, buildings and waste) by the first 'Effort Sharing Decision' (Council Decision 2002/358/EC) covering the emission in the 15 old Member States in the period 2005–2020. The Effort Sharing Decision for the period 2013–2020 (Decision 406/2009/EC) is covering the entire EU with its 28 Member States also affecting the shipping, see *infra* in Sect. 8.1. Based on this decision the GHG emissions from non-ETS sectors must be reduced by an average of 10 % compared to the 2005 level by 2020.

As the EU is not a party to the IMO Convention (see *infra* in Sect. 6.2.), the Council and the Parliament recalled an earlier commitment to take an internal action in 'Climate and Energy Package' adopted by their Decision 406/2009/EC:

> ... in the event that no international agreement which includes international maritime emissions in its reduction targets through the International Maritime Organisation has been approved by Member States or no such agreement through the UNFCCC has been approved by the Commission by 31 December 2011, the Commission should make a proposal to include international maritime emissions in the Community reduction commitment, with the aim of the proposed act entering into force by 2013. Such a proposal should minimise any negative impact on the Community's competitiveness while taking into account the potential environmental benefits.

The Effort Sharing Decision, combined with the reduction obligations for the EU ETS sector emission and other elements of the 'Climate and Energy Package', is designed to ensure that the EU's goal of a 20 % reduction of GHG emissions across the entire economy is met before 2020. By use of this package of instruments it is possible for the EU to fulfil its obligations under the agreed second commitment Kyoto period, see *supra* under Sect. 5.4.

6 Regimes on GHG from Flights and Ships

The emission reduction obligations under the Kyoto Protocol are only covering emissions from land-based activities. Article 2(2) of the Protocol declares that the Annex I parties *shall pursue* limitation or reduction of emissions of GHGs from aviation and marine bunker fuels, working through the ICAO and the IMO.

The international climate regime—UNFCCC/Kyoto—has consequently interacted with the two non-environmental agencies IMO and ICAO as the prime international agencies responsible for international shipping and aviation, respectively. Climate changes do not belong to these two agencies' main concerns, but they have both started slowly to address GHG emissions. Progress in the IMO and ICAO has, however, stalled partly due to the conflict between the South and the North on the relevance of the CBDR/RC principle.

6.1 Discussions on the CBDR/RC Principle Versus the NMFT Principle

The principle of '*common but differentiated responsibilities*' (CBDR) set out in Article 3 UNFCCC leads on to the theme of the differentiated responsibilities of developed countries (Annex I parties) on the one hand and the developing countries (non-Annex I parties) on the other hand, see *supra* Sect. 5.2.—a differentiation between North and South that is not actually relevant under the agreements signed under the IMO and the ICAO.

The guiding principle under the IMO regime is the 'no more favourable treatment' (NMFT) and the leading principle under the ICAO is the principle of 'non-discrimination'. Both these principles are relevant in the international trading system, see *infra* Sect. 10.

The difference between these guiding principles is important—especially in relation to ships, as most ships (representing 85 % of the GHG from international shipping) are registered under the flag of developing countries, but owned by companies in developed countries. It is a unique feature of shipping that fleets and ships cannot be neatly linked to individual countries or even to categories of countries under the UNFCCC—Annex I versus Non-Annex I parties. The use of the CBDR/RC principle is consequently problematic. The registration of the ships in the developing countries as flag countries will exclude the counting of their GHG emission from mitigation obligations, if the obligations are covered by a regime guided by the CBDR/RC. The registration will not exclude the ships from mitigation obligations under a regime guided by the NMFR principle.

Several developing countries have strongly reiterated that any measure for the international aviation and/or maritime transport sector must respect the CBDR/RC principle. In one of the latest meetings under the UNFCCC on this topic they argued that before any MBM can be considered, as a priority in the solutions on

future measures used in multilateral agreements, these agreements must be ensured via specific provisions for transfer of technology, financial resources and capacity building for Non-Annex I parties ('Bonn News Update', 15 May 2012).

6.2 The IMO Convention and the Chicago Convention

According to Article 92 of the Chicago Convention the International Civil Aviation Organisation (ICAO)—founded in 1944—is a regime only for national states. The institutions under the ICAO comprises an Assembly composed of representatives from all contracting states, a Council of limited membership with various subordinated bodies (including the Air Navigation Commission) and a Secretariat. The Council is empowered to make decisions on standards and recommended practices, which are binding on its parties as incorporated as annexes to the Convention on the contracting parties that include the EU's Member States. The EU—represented by the Commission—has had an observer status since 1974. Since 1989, following a decision of the ICAO Council, the Commission has been included in the list of organisations that may be invited by ICAO to participate as observers in the ICAO Assembly, as well as in committees, technical panels, and study groups. On 31 March 2011, the European Council adopted a decision (2011/530/EU) on the signing, on behalf of the Union, and a provisional application of a Memorandum of Cooperation between the EU and the ICAO providing a framework for enhanced cooperation as regards security audits and inspections.

The Convention on the Maritime Organization—the IMO Convention from 1948—that sets out the establishment of IMO is also limiting the circle of accepted parties to national states. It is therefore not possible for the EU (represented by the Commission) to take part in the negotiations under the IMO. In 1974 the Commission and the IMO Secretary-General signed an agreement on cooperation by which the Commission was accepted to participate in meetings with the status as an observer.

In 2002 the EU Commission published a recommendation to the Council to take steps to ensure an amendment of both the IMO and the Chicago conventions (SEC/2002/0381 final) in order to authorise the Commission to open and conduct negotiations within the agencies as the representative for the EU as an ordinary member. In 2005, the Council drafted a procedural framework for the adoption of common EU positions for IMO-related issues and rules governing their expression in the IMO. The procedure is published as the working document, SEC (2005) 449. The EU institutions as well as the Member States in the IMO bargaining situations follow it.

The Commission frequently issues several reasoned opinions to Member States before the meetings in the ICAO and the IMO. It does not hesitate to force the Member States to accept that it is not acceptable, if they themselves act as members to the ICAO and the IMO. The Commission has success in doing so, see *infra* in Sect. 6.3.

6.3 The European Court of Justice is Filling Out the Gaps

The consequence of the specialization under the UN described *supra* is actually that there is a regulatory area covering land-based emission of GHG that is clearly separated from the normative system that regulates global transport. In the opinion of the EU, the multi-faced character of global change makes it necessary to have multi-functional strategies based on a holistic approach covering all emitting sectors, and it is therefore necessary for the EU to use as well its internal regulatory power and its market position.

According to the principle developed by the European Court of Justice in Case C-22/70 *Commission v. Council (the AETR-case)*, the EU acquires external competence when it adopts internal EU rules on the same subject matter that is discussed on transport issues in an international agency. In line with this, the Court stated in the Case C-45/7 *Commission v. Greece* that Member States are not allowed to submit their own proposals—in this case a proposal on monitoring the compliance of ships and port facilities with the requirements of the International Convention for the Safety of the Life at Sea (SOLAS Convention), and the International Ship and Port Facility Security Code (ISPS Code)—to the IMO Marine Safety Committee. First, the Court reiterated that under the EU's treaty (now the Treaty of Lisbon) the setting of a common policy in the sphere of transport was specifically mentioned as one of the objectives of the EU. Second, the Member States must both take all appropriate measures to ensure fulfilment of the obligations arising out of the EU's treaty (now the Treaty of Lisbon) or resulting from action taken by the EU institutions and also abstain from any measure, which might jeopardise the attainment of the objectives of the EU's treaty. If those two provisions are read in conjunction, it follows that to the extent to which the EU rules are promulgated for the attainment of the objectives of the treaty, the Member States cannot, outside the framework of the EU institutions, assume obligations, which might affect those EU rules or alter their scope. The Court is explicitly stating that:

> The fact, that the Community is not a member of an international organization in no way authorises a member state, acting individually in the context of its participation in an international organisation, to assume obligations likely to affect Community rules promulgated for the attainment of the objectives of the Treaty.
>
> Moreover, the fact that the Community is not a member of an organisation does not prevent its external competence from being in fact exercised, in particular through the member states acting jointly in the Community's interest.

Consequently, the Court stated that by submitting the contested proposal to the IMO, Greece has failed to fulfil its 'duty of loyalty' as a Member State under the treaty.

In respect to the activities in the ICAO, the Court has also stated certain exclusive competence to the EU institutions in the field of international air transportation, and consequently the Chicago Convention, see Case C-476/98 *Commission v. Germany*.

By its judgements the Court has greatly expanded its external competence towards Member States in the field of international transport (Nengye and Frank 2012, at p. 288).

7 Initiatives on GHG Emission from Flights

According to the IPCC 4th Assessment Report, aviation represents approximately 2.5 % of the global GHG emissions. As mentioned above, the Kyoto Protocol requires the Contracting Parties to find ways to reduce GHG emissions from aviation by working through the ICAO. In 2007, the ICAO Assembly recognized the critical importance of providing leadership to international civil aviation on limiting the GHG emission from the aviation sector. The ICAO has established a climate change group, which has recommended the improvement of the average fuel efficiency. However, given annual growth in aviation traffic, this measure would not reduce the growth of GHG emissions (Meltzer 2012).

The EU has, as described *supra* in Sect. 6.2., only an observer status in the ICAO. It is working through its Member States on the development of market-based measures, to address the climate change impacts of aviation. It is the ambition of the EU that the EU ETS scheme shall serve as a model for the use of emission trading based on the principle of 'cap and trade' on flights worldwide.

The activities under the ICAO regime have, however, until now been without success, see *infra* Sect. 7.3.

7.1 The EU's Internal Regulatory Initiate on Flights with Extraterritorial Effects

From an economic and environmental point of view, the EU has—by the amendment of the ETS Directive in 2008 (Directive 2008/101/EC) included CO_2 emissions from all flights by the EU as well as non-EU carriers to and from an airport in the territory of the EU Member States in the EU ETS. All aircraft operators must have enough allowances to surrender allowances calculated in the light of the whole of their international flight. Consequently, the EU rules have extraterritorial effects. If an operator emits CO_2 in excess as to the level of emissions covered by its CO_2 quotas, including any extra CO_2 quotas purchased and extra JI or CDM credits obtained, the operator must pay an excess emission penalty for each tonne of CO_2 emitted without surrendering the allowances. The EU Member States have to ensure that any operator who does not surrender sufficient allowances by 30 April each year to cover the emissions from its activities during the preceding year is held liable. In addition to paying the excess emission penalty, the operator must surrender an amount of allowances equal to those excess emissions to the relevant administrating

Member State. If the enforcement measures used are not successful, that Member State may request the Commission to decide on the imposition of an operating ban at EU level on the aircraft operator concerned.

In the preamble of the 2008 amending ETS Directive it is stressed that the EU and its Member States should continue to seek an agreement on global measures to reduce GHG emissions from aviation. If a third country adopts measures, which give an environmental effect at least equivalent to that of the EU ETS, to reduce the climate impact of flights to the EU, the Commissions should consider the options available in order to provide for optimal interaction between the EU ETS and the country's measures, after consulting that country. The use of the amended EU ETS on foreign operators and the use of EU's market power are consequently only supplementary means used by the EU to ensure the effectiveness of the global climate regime.

7.2 Reaction to the Extraterritorial Effects and the Judgement of the European Court of Justice

A number of American and Canadian airlines and airline associations contested the measures transposing the 2008 amendment of the ETS Directive before the High Court of England and Wales. They were arguing that the EU was overreaching its legal competences by charging the entire passage of a flight, and not just the EU element of it. The High Court referred the case to the European Court of Justice on the question of interpretation and validity of the compatibility of the amended ETS Directive with international law.

In its judgment of 21 December 2011—in Case C-366/10 *The Air Transport Association of America and others v. Secretary of State for Energy and Climate Change*—the European Court of Justice concluded that the 2008 amendment of the ETS Directive is not in conflict with international law. It stated that the EU is not a party to the Chicago Convention that was signed by its Member States before 1958, and therefore EU was not bound by it, see Article 351 TFEU. The fact that one or more acts of EU law may have the object or effect of incorporating into EU law certain provisions that are set out in a multilateral agreement, which the EU has not itself approved, is not sufficient for the agreement to be incumbent upon the Court to review the legality of EU's legislation in the light of that agreement. The Court could therefore not examine the validity of the amendment to the ETS Directive (2008/101/EC) in the light of the Chicago Convention. The legality of the 2008-amendment to the EU ETS is therefore based on EU law. The Court points out that EU policy on the environment aims at a high level of protection, see Article 191 TFEU. Thus, the EU legislature may in principle chose to permit a commercial activity—in this instance air transport—to be carried out in its territory only on condition that the operators comply with the criteria that have been established by the EU. It specified that the rules of the amended

Directive were related only to aircrafts that arrive or depart from the territory of the EU Member States—and more specifically on an aerodrome situated on their territory. In such situations the aircrafts are subjects to the unlimited jurisdiction—and national environmental law—of the relevant Member States. Furthermore, the fact that certain matters contributing to the pollution of air, sea or land territory of the Member States originate in an event which occurs partly outside that territory is not such as to call into question, in the light of the principles of customary international law capable of being relied upon, the full applicability of EU law in that territory.

On 21–22 February 2012, a number of non-European countries met in Moscow to coordinate their opposition to the aviation's inclusion in the EU ETS. The meeting resulted in a joint declaration and a series of suggested retaliatory measures to be used against European industry. On 31 July 2012, the U.S. Senate Commerce Committee passed a bill that would allow the Secretary of Transportation to ban airlines from complying with the EU ETS. Despite the decision taken by the EU to defer the international aspects of the EU ETS, see *infra* Sect. 7.3., the Congress completed its approval of the bill in November 2012.

The 2008 amendment of the EU ETS is said also by the EU's own agencies to have a negative impact on the potential trade agreements—including the potentially lucrative EU-India Free Trade Agreement (EU-India FTA), which was started in June 2007 and which is in the final stages of negotiation (DG for External Relations, 2012, at 17).

However, more than 99 % of all major global airlines have complied with the first step of the EU ETS. Only eight Chinese airlines and two Indian ones did not comply by the 31 March 2012 deadline (press release from the EU Commission on 15 May 2012). On 20 September 2012, the EU and China signed a financial agreement promoting the use of the ETS concept in China based on the EU experiences. Such a partnership can have consequences for the just mentioned non-compliance by Chinese airlines.

7.3 The Expected Outcome of the ICAO Session in the Autumn of 2013

In November 2012, the ICAO Council announced that it would take steps towards the adoption, at the 38th session of the ICAO Assembly, which will be held from 24 September to 4 October 2013, of a global framework for emission reduction, which facilitates the application of market-based measures to emissions from international aviation, and on the development of a global agreement based on the use of such measures. On 24 April 2013, the European Parliament and the Council approved the proposal of the Commission—Decision 337/2013/EU—to exempt flights operating into and out of Europe from enforcement of the ETS Directive in 2010, 2011, and 2012. It is stated in the decision:

> This derogation is provided by the Union to facilitate an agreement at the 38th session of the ICAO Assembly on a realistic timetable for the development of a global MBM beyond the 38th session of the ICAO Assembly and on a framework for facilitating the comprehensive application of national and regional MBM to international aviation, pending the application of the global MBM.

The decision is only addressed to operators from non-EU Member States. The general EU ETS rules continue to apply without any exemptions to all flights within and between the Member States and states in the European Economic Area—also in 2010, 2011 and 2012.

The latest news—published in the ENDS Europe on Wednesday 24 July 2013—on the negotiation in the ICAO is that the US favours mandating the ICAO Council to hold further discussions on tackling airlines' emissions, rather than making a deal in September 2013 as expected by the EU. If the discussion is postponed at the session in the autumn of 2013, the next ICAO Assembly in 2016 should consider the results of that discussion.

The consequence of such an outcome of the 38th session of the ICAO Assembly in the autumn of 2013 is that all non-EU flights to and from an airport in the territory of the EU Member States will be included in the EU ETS again from 2013, see *supra* the citation of some of the text in the Decision 337/2013/EU. The US ban to American airlines from taking part in the EU ETS at any time will also come into force, and the weaknesses in the effectiveness of the state-based multilateral negotiations will be clear.

8 Initiatives on GHG Emission from Ships

Emissions from the global shipping industry amount to around 1 billion tonnes a year, accounting for 3 % of the world's total GHG emissions. Without mitigation actions, these emissions are expected to more than double by 2050. Between 1990 and 2007 CO_2 emissions from this industry increased by 87 %. Over the same period, world GDP increased by around 65 % and world seaborne trade increased by over 100 % by volume (RICARDO-AEA 2013, p. ii). The maritime industries have endeavoured to optimize ships' fuel consumption, e.g. through the development of more efficient engines and propulsion systems, optimizing hull designs and larger ships, and have thereby achieved reduction in fuel consumption and in GHG emissions on a capacity basis, but such voluntary initiatives will not be enough to reduce the emission of GHG from international shipping.

The international regulatory context for shipping emissions is set by the Marine Environmental Protection Committee (MEPC) of the IMO. The IMO has adopted the Resolution A.963(23), which urges the MEPC to identify and develop the mechanism or mechanisms needed to achieve the limitation or reduction of GHG emissions from international shipping. On that basis, technical measures for new ships (Energy Efficiency Design Index—EEDI) and operational reduction measures (Ship Energy Efficiency Management Plan—SEEMP) for all ships have been agreed

in July 2011. The rules are included in Chapter "Dynamic Capabilities and the Multinational Enterprise" in Annex VI (that is related to air pollution) to the Convention for the Prevention of Pollution from Ships (MARPOL 73/78)—entitled 'Regulations on energy efficiency for ships'. The EEDI requirements are developed for the largest segments of the world's merchant fleet and are covering 87 % of emissions from new ships—the requirements came into force in 2013. However, due to the long economic life of merchant ships, it will take about 20 years to reach this coverage without additional incentives. Industry forecasts indicate that CO_2 emissions from shipping will continue to rise—even if a 58–75 % increase in energy efficiency is achieved (ICS MBI Analyse Report 2009, at p. 3). Another weakness is that these amendments to the Annex VI were not adopted by consensus—it has been approved despite strong objections expressed by several developing countries ('Bonn News Update', 15 May 2012).

8.1 Discussions on the Use of the Principle of 'Cap and Trade'

It has been recognised by the parties to the IMO that technical and operational measures would not be, in the long term, sufficient to meet the overall reduction obligations needed. In November 2011 it was therefore announced in a note from the IMO Ad hoc Working Group on Long-term Cooperative Action under the UNFCCC that IMO has concluded that market-based measures are needed for the effective regulation of GHG emissions from international shipping.

A topic of debate within the IMO is how the wording of Article 2(2) of the Kyoto Protocol should be interpreted, and if the CBDR/RC should apply to a GHG regime for international shipping rather than the IMO's basic principle of NMFT. A number of developing countries maintained the view that any measure to reduce GHG emissions to be adopted by the IMO should only be applicable to Annex I parties to the UNFCCC and its Kyoto Protocol, while other parties (developed countries) have expressed the opinion that the IMO regulatory framework on GHG emissions should be applicable to all ships, irrespective of the flags they fly. This discussion is mentioned in the Second IMO GHG Study 2009, at p. 21, where it is concluded under point 2.45–2.46:

> 2.45 As demonstrated earlier, the ownership and management chain surrounding ship operations can involve many players, located in various countries. In addition, the registration of a ship can move between jurisdictions several times over its lifetime. It is worth noticing that about three quarters of the world tonnage, by deadweight, of all merchant vessels engaged in international trade is registered in developing countries (not in Annex I of the Kyoto Protocol), hence making it a large portion of the world fleet; it would be ineffective for any regulatory regime to act only on the remaining portion, namely one quarter of the world fleet.
>
> 2.46 Given IMO's global mandate, given by the IMO Convention itself as well as from UNCLOS, there is no precedence in any of the more than fifty IMO treaty instruments currently in existence where measures are applied selectively to ships according to their flag. On the other hand, there are several international environmental agreements which

have a different approach, such as the Montreal Protocol (on substances that deplete the ozone layer), yet, when IMO has dealt with the same issues, the principle of different approach has not been taken on board.

The MEPC 63th session held in 2012 included a discussion on a 'cap and trade' emission trading scheme, and a carbon levy or tax system as market-based measures, which could be linked to the package of special technical and operational reduction measures agreed in the IMO. The MEPC failed, however, in delivering terms of reference for a study assessing the impacts of proposed market-based measures. Progress in this area has stalled mainly due to the conflict between the NMFT principle under the IMO and respecting the CBDR/RC under the UNFCCC/Kyoto Protocol (RICARDO-AEA 2013, p. ii).

8.2 The EU's Initiative with Extraterritorial Legal Effects: A Proposal

As the deadline set by the European Council and Parliament in their Decision 406/2009/EC mentioned *supra* in Sect. 5.5. has been passed without sufficient international action, the Commission, on 28 June 2013, has presented two communications: one on 'Integrated maritime transport emissions in the EU's greenhouse gas reduction policies'—COM(2013) 479 final—and one on a 'proposal for a Regulation on the monitoring, reporting and verification of carbon dioxide emissions from maritime transport'—COM(2013) 480 final. The main aim of these initiatives is to introduce a mandatory MRV system for CO_2 emissions from ships as a first step of a staged approach to reduce the global emission from ships. As regards the geographical scope for the proposed internal EU rules, the following routes will in principle be covered in a non-discriminatory manner for all ships regardless of their flag: (1) intra-EU voyage; (2) voyages from the last non-EU port to the first EU port of call (incoming voyages); and (3) voyages from an EU port to the next non-EU port of call (outgoing voyages). Non-compliance with the provisions of the Regulation should result in the application of sanctions.

It is the ambition of the EU that the proposed MRV system will function as an example for the implementation of a global MVR with the aim to speed up the international discussion, and that the MRV system could be converted into a global system. There is actually not a strong reaction from non-EU Member States on the external effects of the MVR obligations.

9 Sustainability Criteria for European Road Transport

Renewable energy—including biofuels—as the alternative to conventional energy sources is one of the most important means used to reduce the emission of GHGs. This is also the case in the transport sector, where biofuels can be a good renewable source.

The biofuels used in the EU are, however, very often imported from developing countries, where the environmental legislation on deforestation etc. can be weak compared to EU standards. The global concerns on the effects of feedstock production to be used in the production of biofuels has increased in recent years as international demand for biofuels has increased with impacts on the environment, security, food prices etc. There are concerns regarding the production's pressure on eco-sensitive areas, the effect on soil fertility, water availability and quality, and pesticide use as well as emission of GHG from the soil in countries where a large-scale expansion of feedstock production is taking place (UNEP 2011, at pp. 233f.).

In order to try to guarantee a global environmental benefit from the use of biofuels, the EU has developed totally harmonized and binding quality criteria—titled 'sustainability criteria'—which focus on the whole life cycle of biofuels, including the production process and transport of biofuels (and the biomass used in the production of biofuels) that take place outside the territory of the EU. The Directive on Renewable Energy (Directive 2009/28/EC—hereinafter referred to as the RED)—is promoting the reduction of GHG emission from road traffic by such mandatory sustainability criteria, see *infra* Sect. 9.1.

9.1 Mandatory Sustainability Criteria used on Biofuels

By the use of Article 114 TEUF as the legal basis for the Articles 17–19 of the RED the criteria are established as totally harmonized legal conditions covering all EU Member States. Article 17(8) of the RED states that the Member States must offer access to their market and support schemes to biofuels that meet the sustainability criteria. They cannot use any additional criteria. The criteria are also included in the amended Fuel Quality Directive (amended by Directive 2009/30/EC).

There are three elements in the criteria—the mitigation of GHG emissions, the land-use aspects and the agricultural-environmental aspects. They aim to avoid or minimise GHG emissions and biodiversity loss. Concerning the mitigation aspects of the criteria it is mandatory for the Member States to ensure that the GHG emission savings in the period 2009–16 are at least 35 %, including the calculation of the full carbon effects of land conversion of growing biomass. The GHG savings with effect from 1 January 2017 shall be at least 50 %, and with effect from 1 January 2018 at least 60 %. If these targets are not met, the Member States cannot use the biofuels to fulfil their RES obligations. The criteria have to be met by the Member States for the counting of the renewable energy targets under the RED, and for compliance with the renewable energy obligations under the Fuel Quality Directive. The criteria are also relevant for investment and/or operating aid in accordance with the EU Guidelines 2008/C82/01 on State Aid for the Environmental Protection (Basse 2013).

Based on the land use-related criteria, raw materials (biomass) for biofuels and bioliquids cannot be produced on land with high carbon stock or on wetlands, forests and areas with other wooded land of native species. High biodiversity land

is defined to include: (1) wooded land where there is no evidence of human activity before 2008 and ecological processes have not been disturbed; (2) highly biodiverse grassland; and (3) internationally or nationally designated conservation areas. The criteria ensure the protection of rare, threatened and endangered ecosystems or species recognised by international agreements or included in lists drawn up by intergovernmental organisations or by the International Union for the Conservation of Nature (IUCN).

Concerning the agricultural-environmental aspects, the vertical as well as the horizontal coordination through the sustainability criteria is reflected in the RED incorporating the cross-compliance requirements that are mandatory in the allocation of financial support for agriculture under the Common Agricultural Policy (CAP), concerning the scope of these aspects see *infra* Sect. 10. The CAP has five objectives: (1) to increase agricultural productivity; (2) to ensure a fair standard of living for the agricultural community; (3) to stabilize markets; (4) to ensure the availability of supplies; and (5) to ensure that supplies reach consumers at reasonable prices; see Article 39 TFEU.

9.2 Extraterritorial Legal Effects

Most of the sustainability criteria have a global impact on biomass productions, markets and land use (Burrell 2010), as EU is the largest biofuel user in the world.

In respect to biofuels produced from raw materials cultivated outside the territory of the EU it has to be assessed whether the criteria are respected and whether the exporting country has ratified and effectively implemented several international labour and environmental treaties. The RED states that the Commission should monitor the impact of increased demand for biofuels on the social sustainability in the EU and outside the EU, and the impact on the availability of food for affordability, especially for people in developing countries. Other development issues must also be included in the implementation of the RED's criteria.

The biofuel producers and importers are responsible for showing that these environmental and social criteria have been fulfilled. Therefore, imported raw materials cultivated outside the territory of the EU have to provide information on several issues, including information about sustainable water and soil management plans and the use of dangerous agrochemicals. The information should be offered also on the producer's right to use the land and on procedures to consult local populations and interest groups on plans and activities that may affect their rights, property, resources, or livelihoods.

The reasoning behind these extraterritorial effects is the need to avoid unjustified competitive advantages for the producers of biofuels in third countries over their European competitors and to ensure the respect for the goals of the EU also in the extraterritorial relations.

9.3 GHG Emissions from Indirect Land Use: A Proposal

Cultivating biomass feedstock as the basis for production of biofuels can have direct and indirect land-use changes (ILUC), as production of biomass used in the production of biofuels is displacing current agricultural or forest production to other areas—e.g. grasslands or forested land, which causes direct land use effects at the new location. The production of biomass to be used for the production of biofuels can take place in areas, where ownership of land is more readily recognized, and where few restrictions on the farmers' activities exist—e.g. production of feedstock for biofuels—that can be carried out, and the goods and services provided by nature (e.g. forests) are often not valued property. Uncertain ownership rights, complex regulations and weal law enforcement and corruption can all lead to conversion of sensitive areas (e.g. by deforestation) to other uses such as agriculture. While standing forests store carbon, acting as carbon 'sinks', deforestation sends carbon back into the atmosphere, thus rendering forests major carbon sources. The displacement by ILUC could move previous agricultural production to areas outside the traditional producing countries. There are also concerns regarding the biofuel production's pressure on eco-sensitive areas, the effect on soil fertility, water availability and quality, and pesticide use as well as emission of GHG from the soil in countries, where a large-scale expansion of feedstock production is taking place.

Given that the negative environmental impact of the production of biomass to be used in biofuels have come to light, and that the low-carbon benefits of some biofuels have been questioned, the need for more sustainable regulation has been brought into the global political agenda—and the amendment of the current legislation is on the EU agenda. On 17 October 2012, the EU Commission published a proposal on an amendment of the RED and the Fuel Quality Directive aiming to address the ILUC of biofuels (COM(2012) 595 final). The proposal includes a cap of 5 % on the contribution of biofuels produced from food crops to the transport-fuel target of 10 % in the RED.

The proposal points out that after 2020, biofuels should only be subsidies, if they lead to substantial GHG savings that take ILUC into account, and provided that these fuels do not come from food crops. The aim of the proposal is to limit global land conversion for the purpose of biofuel production, and raise the climate benefits of biofuels used in the EU. In contrast to new biomass materials, the provision of waste materials and residuals as second generation biofuels does not require significant energy input. Therefore, the lowest GHG emissions are achieved using biofuels and biodiesel from vegetable oil from nutrition or methane from manure. For the purpose of stimulating the use of more sustainable types of waste materials for biofuel production, the contribution of such biofuels is counting by a factor 2 or 4.

10 Relations to Global Trade Rules

The three EU initiatives mentioned *supra* in Sects. 6–9 involving measures adopted as part of the EU's internal climate legislation to restrict GHG emission may affect international competitiveness. They may have negative consequences in non-EU Member States as well as in EU Member States.

The negative consequences of various climate policies and measures (spill-over effects) are taken into account in the current rules of the UNFCCC/Kyoto Protocol. Article 3(5) of the UNFCCC states that

> Measures taken to combat climate change, including unilateral ones, should not constitute a means of arbitrary or unjustified discrimination or a disguised restriction on international trade.

Article 2(3) of the Kyoto Protocol incorporates this approach stating that Annex I parties should strive to implement policies and measures in such a way as to minimize adverse effects, including on international trade. Trade aspects—including the World Trade Organization (WTO) rule set—must in other words be incorporated into the concrete solutions or by use of internationally agreed measures. The General Agreement on Tariffs and Trade (hereinafter referred to as the GATT) mandates equal treatment of like products. The mandatory sustainability criteria that hinder or prevent the use of non-sustainable biofuels, which are 'like' sustainable biofuels, are potentially conflicting at least with GATT Article I, which sets out provisions in respect to the principle of the most favoured nations (MFN), according to which the Parties must extend any advantage granted to a product from one WTO member to 'like' products from all other WTO members. The WTO and its Appellate Body makes it possible to assess the competitive consequences of the initiatives under the GATT and the other WTO rules that regulate the market access. The question is whether the Appellate Body can assess the three cases described supra in Sects. 7–9, as the WTO system only aims at delimiting trade restrictions that nations impose on transboundary trade in goods and services.

10.1 The Inclusion of Non-EU Flights in the EU ETS

One of the objectives of the extraterritorial effect of the EU ETS rules on flights, described *supra* in Sect. 7, is to limit the risks of 'carbon leakage' and to take care of competitive consequences of the costs of allowances for the national flight services. The requirements established under the 2008 amended ETS Directive can only be characterised as covered by the WTO rules, if the marketing of emission units can be defined as services under the Agreement on Government Procurement (AGP) listed in its Annex 4, Appendix 1. However, this is not the case. The allowances under the EU ETS exist only as a consequence of the legally binding commitments on mitigation of GHG emission, and the transfer of allowances as well as the transfer of units created by projects (JI of CDM projects) do not create a market

in neither goods nor services. Therefore, trading of the rights to emission among actors in the EU under the EU ETS is not covered by the WTO rules (Voigt, 288f.). The sector-based measures on flights may not be appealing to many non-Annex I parties. Especially the LDC and the SIDC do not have the capacity to take part in reliable sector-based measures, and emphasis on such measures is risks is of diverting attention from the financial commitments by the Annex II-parties under the UNFCCC/Kyoto Protocol, and shift the focus further from public to private financing.

10.2 The EU Initiatives on International Maritime Emissions

The MRV obligations proposed by the EU as a regulation on ships, described *supra* in Sect. 8, are not product- or service-related measures—they are procedural instruments meant to ensure transparency. Transparency matters also for the ability of citizens to hold their governments accountable—and thereby to ensure the private involvement. The work of the WTO itself includes obligations for the parties on monitoring and reviewing of national trade policies, which is considered to be of fundamental importance for increasing the transparency and understanding of the parties' trade policies and practices. The EU's proposal on the MRV obligations for the shipping industry also explicitly takes the WTO criteria on discrimination into account. A potential disadvantage as to the EU proposal on a regulation establishing MRV obligations on shipping industries is the administrative costs, which could become proportionally higher for LDC and the SIDC than for other countries, cf. *supra* Sects. 5.1 and 5.2.

10.3 The EU Sustainability Criteria on all Imported Biofuels

Biofuels is subject to tariff cuts and discussion on trade and environment under the Doha Round (WTO 2012) following negotiations on agricultural market access. Producer countries of biofuels could consider that the sustainability criteria infringe the GATT, since they restrict export of biofuels to the EU.

The agricultural-environmental criteria under the RED (mentioned *supra* in Sect. 9.1.) apply only to biofuels produced from raw materials originating in the EU. The reasoning behind this jurisdictional limitation of the extraterritorial effects of these sustainability criteria is the obligation of the EU and its Member States to comply with the WTO provisions.

The other mandatory sustainability criteria (mentioned *supra* in Sect. 9.1.) could be regarded as non-conformant, if sustainable and non-sustainable biofuels can be regarded as 'like' products. However, Article XX of GATT provides exceptions, if the criteria can be justified as necessary to protect human, animal or plant life or health, or can be justified as necessary because of the need for conservation of

exhaustible natural resources and the measures are made effective in conjunction with restrictions on national production or consumption. Another question is whether the sustainability criteria based on non-product-related criteria are covered by the Agreement on Technical Barriers to Trade (TBT Agreement) that aims to ensure that technical regulations and standards and uniform assessment procedures do not create unnecessary obstacles to international trade. The TBT Agreement seems only applicable to measures that regulate products' characteristics or their processes and production methods (Pelsy 2008, at p. 132). The sustainability criteria are not product-related measures, and they do not guarantee the quality, safety and functionality of biofuels as a product.

11 Some Concluding Remarks

This article examines the extent to which the EU is taking part in the development of the emerging 'global environmental law' as this legal order is described *supra* in Sect. 3. As the background for the examination of the role of the EU the special legal order of the EU is described *supra* in Sect. 4. The EU's ratification of the Aarhus Convention—and the implementation into EU law of the procedural rights—is one of the important means used by the EU to ensure that the vertical harmonization of international, EU and national environmental law is also supported by private activism—the use of its market power is another mean used by the EU.

There are several weaknesses in the effectiveness of the state-based multilateral negotiations on the agreements on the needed measures to tackle the negative consequences of globalisation, see *supra* Sect. 2. The integration and harmonization of the interests of all sovereign states are very complicated, especially within policy areas—such as climate policy—with a very long-term perspective and with interdependency between the sector-based policies. These weaknesses are visible in the actual negotiations on global mandatory measures on mitigation of GHG. It is stated *supra* in Sects. 5and 6 that the EU plays a very active role as a key participant in multilateral rule setting in the UNFCCC and the Kyoto Protocol as a member to these multilateral agreements, and that the EU is playing the same active role in the ICAO and the IMO without having the legal status as a member. The specialization at the international level has consequences for the functional set-up and effectiveness of the multilateral agreements on mitigation of GHG. The specialization makes it difficult to ensure solutions that are effectively handling the many interrelated challenges caused by climate changes. In the absence of a new post 2012 agreement on binding mitigation obligations for all Annex I parties under the UNFCCC there is a growing uncertainty over the future of global 'cap and trade' schemes—and consequently also of the EU ETS established as part of a future global carbon market. These challenges are forcing the EU to act.

In addition to relying on its legal position, the EU has exercised directional leadership by means of examples—as described *supra* in Sects. 7–9. It is explained how the European Court of Justice as well as the EU legislation is filling out

the gaps in international environmental law on climate changes by its 'internal' legislation with extraterritorial effects. The EU is acting as a global actor both in the negotiation activities and outside these regimes as a marked actor when exercising its extraterritorial effects by effectively imposing its standards and criteria on all market actors that want to trade with the Member States. The EU's activism inside as well as outside its legal order is one of many ways in which 'global environmental law' can evolve away from a system controlled by nation-state actors defending state sovereignty as a paramount principle.

References

Barendrecht, M., Raic, D., Janse, R., & Muller, S. (2012). *Trend report Rulejungling. When lawmaking goes private, International and informal*. Hiil Report.

Basse, E. M. (2013). The legal design of sustainable criteria on biofuels used by the European Union. In *Environmental reviews and case studies*. Cambridge University. doi:10.1017/S1466046612000488

Basse, E. M., & Gaines, S. E. (2010). Common but differentiated responsibilities and respective capabilities' as Part of the Post-2012 climate regime. In Nordisk Miljörättslig Tidskrift (2, pp. 189–203).

Birnie, P., Boyle, A., & Redgwell, C. (2009). *International law & the environment* (3rd ed.). Oxford: Oxford University Press.

Burrell, A. (Ed.). (2010). *Impacts of the EU biofuel target on agricultural markets and land use. A comparative modelling assessment*. JRC Reference Reports, Institute for prospective Technological Studies (EUR 24449EN).

Edgeworth, B. (2002). *Law, modernity, postmodernity. Legal change in the contracting state* (Applied legal philosophy). London: Ashgate.

European Environmental Agency (EEA). (2011). *Greenhouse gas emission trends and projections in Europe 2011. Tracking progress towards Kyoto and 2020 targets*. EEA Report no. 4.

ICS. (2009, March). *MBI Analysis report*. Prepared by Giraffe Innovation Ltd on behalf of ICS. - ICS Executive Committee Working Group on Greenhouse Gas Emissions and Marked Based Instruments.

Jans, J. H., & Vedder, H. H. B. (2012). *European environmental law* (4th ed.). The Netherlands: European Law Publishing.

Koskenniemi, M. (2006). *Fragmentation of international law: Difficulties arising from the diversification and expansion of international law*. Report of the Study Group of the International Law Commission under the U.N. General Assembly, A/CN.4/L.682 of 13 April 2006.

Meltzer, J. (2012). Regulating CO_2 emissions from aviation in the EU. *American Society of International Law*, *16*(27). https://www.asil.org/insights/volume/16/issue/27/regulating-co2-emissions-aviation-eu

Nengye, L., & Frank, M. (2012, April). Legal constraints to the European Union's accession to the International Maritime Organization. *Journal of Maritime Law & Commerce*, *43*(2), 279–291.

Nielsen, R. (2012). Towards an interactive comparative method for studying the multi-layered EU Legal Order. In U. Neergaard & R. Nielsen (Eds.), *European legal method—in a multi-level EU legal order* (pp. 105–106). Copenhagen: DJOF Publishing.

Oberthür, S., & Kelly, C. R. (2008, September). EU Leadership in International climate policy: Achievements and challenges. *The International Spectator*, *43*(3), 35–50.

Pelsy, F. (2008). The European Commission 2008 directive proposal on biofuels: A critique. *Law Environment and Development Journal, 4*(2), 121–134.

Percival, R. V. (2010). Liability for environmental harm and emerging global environmental law. *Maryland Journal of International Law, 25*, 37–63.

RICARDO-AEA. (2013, January 13). *Support for the impact assessment of a proposal to address maritime transport greenhouse gas emissions.* Report for European Commission.

Rose, G. L. (2011, October 12–13). National and global environmental laws: Dichotomy and interlinkages as examined through the implementation of multilateral environmental agreements. UNEP, First Preparatory Meeting of the World Congress on Justice, Governance and Law for environmental Sustainability.

Semmelmann, C. (2012). The public-private divide in European Union law. In U. Neergaard & R. Nielsen (Eds.), *European legal method—in a multi-level EU legal order* (pp. 183–223). Copenhagen: DJØF Publishing.

Sommer, T. (2013). *Can law make life (too) simple? From gene patents to the patenting of environmentally sound technologies*. Copenhagen: DJØF Publishing.

UNEP. (2011). *Towards a green economy. Pathways to sustainable development and poverty eradication*.

Voigt, C. (2006, September). *Sustainable development as a principle of integration in International law. Resolving potential conflicts between WTO law and climate change mitigation measures*. University of Oslo, Faculty of Law (unpublished dissertation).

World Commission on Environment and Development (WCED). (1987). *Our common future*. New York: Oxford University Press.

WTO. (2012). *The Doha round*, Geneva. http://www.wto.org/english/tratop_e/dda_e.htm

Yang, T., & Percival, R. (2009). The emergence of global environmental law. *Ecology Law Quarterly, 36*, 615–664.

Product Market Integration, Tax Distortions and Public Sector Size

Torben M. Andersen and Allan Sørensen

Abstract The implications of product market integration for public sector activities (transfers and public consumption) are considered in a standard setting. The analysis supports that a larger public sector (higher tax rate) tends to increase wages and worsen wage competitiveness. However, the implications of product market integration for the public sector are far from straightforward. The reason is gains-from-trade effects which tend to increase the tax base and decrease the opportunity costs of public consumption (marginal utility of private consumption falls). It follows that the retrenchment view that product market integration inevitably leads to a downward pressure on public sector activities does not get unconditional support in a standard setting. A particularly noteworthy finding is that a country with a large public sector (strong preferences for public consumption) may benefit more by integrating with a country with a smaller public sector (weak preferences for public consumption). This type of comparison could not be made in earlier analysis assuming symmetric preferences and technologies across countries.

1 Introduction

The future of the welfare state is a topical issue in many industrialized countries. In particular it is a widely held view that globalization makes it more difficult to maintain a large and extended welfare state. The public debate centres around the nexus between taxes needed to finance public sector activity and competitiveness. If taxes tend to lead to higher wages and thus a deterioration of competitiveness, and

T.M. Andersen (✉)
Department of Economics and Business Economics, Aarhus University and CEPR, CESifo and IZA, Aarhus, Denmark
e-mail: tandersen@econ.au.dk

A. Sørensen
Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark
e-mail: allans@econ.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_18

if globalization makes competitiveness more important for economic activity,[1] the obvious reasoning seems to be that globalization tends to make it more expensive in terms of efficiency losses to maintain a large public sector. Retrenchment of the public sector therefore must follow (for unchanged preferences) from the inevitable fact that the social costs of maintaining public sector activities become larger.

There is indeed a large body of literature[2] building on an extensive tradition in trade and open macroeconomics supporting that higher taxes [3] or a higher level of public sector activities may harm competitiveness.[4] Empirical analyses have also found support that fiscal policy via a cost channel harms competitiveness [see e.g. Alesina and Perotti (1997), Daveri and Tabellini (2000), Lane and Perotti (2003)]. But does this immediately support the retrenchment view? These arguments seem to neglect the gains from trade which under standard assumptions lead to higher income and private consumption; i.e. there is a benefit side to globalization which has to be considered on par with the cost side before statements on the effects on public sector activities can be made.

In the present paper we merge elements from trade theory, macroeconomics and public finance into a general equilibrium setting allowing for an endogenous determination of production, specialization, and thus trade structure across countries. Ricardian models have recently been widely used to analyse the effects of international integration since this framework allows for an endogenous determination of production, trade and specialization structure, depending on trade frictions as a metric of market integration. Since globalization driven by both political and technological changes lowers trade frictions, it follows that this framework captures essential elements of the changes and effects associated with the globalization process.[5] This literature builds on Dornbusch et al. (1977), and recent contributions build on Eaton and Kortum (2002). We follow this approach modelling product market integration as reductions in trade frictions, which in turn implies that the

[1] Often phrased as production or employment becoming more sensitive to relative prices and wages, i.e. elasticities go up, see e.g. Rodrik (1997), Burda (2000), OECD (2007).

[2] See e.g. Bruce and Purvis (1985) and Marston (1985). Similar effects are found in the "New Open Macroeconomics" literature, see e.g. Botman et al. (2006).

[3] This applies to taxes financing public activities which in no way improve production possibilities.

[4] This holds with both competitive and imperfectly competitive markets. In competitive models, pre-tax wages increase as labour supply contracts provided that substitution effects dominate income effects, as is usually found empirically, see e.g. Evers et al. (2008). In imperfectly competitive labour markets, the tax wedge causes higher wage costs, and this tends to harm competitiveness [see e.g. Layard et al. (2005)].

[5] Tax base mobility is an important aspect of globalization. Mobile tax bases may change location to minimize tax payments, and this creates a specific channel through which taxation is affected by globalization (see e.g. Razin and Sadka (2005)). This has led to an intensive debate on tax competition in relation to taxation of corporations (see e.g. EEAG (2007) for a recent discussion). While an important issue it is of secondary importance to the financing of public sector activities since the revenue accruing from the tax bases becoming mobile is below 5–10 % for all OECD countries, and moreover (tax driven) labour migration is relatively low. Hence, to focus on the main mechanisms migration is disregarded.

non-tradeable sector shrinks and that there is reallocation of production and employment according to comparative advantages and thus gains from specialization. The public sector finances public consumption (service provision) and transfers via an income tax. In this general equilibrium setting, we analyse how fiscal policy affects various key variables, including wage competitiveness, and we consider the optimal determination of both transfers and public consumption. We also analyse how the effects of fiscal policy and the optimal policy (for given policy objectives) change in the wake of further international integration.

In a companion paper (Andersen and Sørensen, 2012)[6] we consider how a welfare state financed by taxation of labour income is affected by product market integration in a general equilibrium setting where labour taxation distorts labour supply and thus competitiveness, goods are traded internationally and trade and production structures are endogenously determined. Under a utilitarian social welfare function optimal policies are determined by a variant of the Samuelson rule in which the marginal utility of private consumption plays a crucial role (part of the opportunity costs of higher public activities). In addition to the "competitiveness"-effect, two other effects are at play and released by gains from trade. First, an increase in employment and income due to market integration tends, for given tax rates, to increase tax revenue and thus the scope for financing public sector activities (or given activities can be financed at lower tax rates). Second, an increase in private consumption tends to lower the marginal benefit of private consumption, which in turn tends to increase the optimal level of public activities by reducing the opportunity costs of such activities. Hence, even though the "competitiveness effect" is present so are "gains from trade", and it follows that the fear of a retrenchment of the public sector may be exaggerated. In particular in Andersen and Sørensen (2012) we showed the following: First, in case of cooperative tax policy the effect of further product market integration (equivalent to a productivity increase in the private sector) on the optimal labour income tax is ambiguous. Second, provided the elasticity of labour supply is constant (or decreasing in wages) marginal costs of public funds in case of a cooperative policy decrease with product market integration for a given revenue constraint. Thirdly, in the non-cooperative case it is more likely that the optimal tax (and hence public sector activities) increases than in the cooperative case.

The abovementioned results are derived in a setting with integration between symmetric countries (similar technologies and similar preferences with respect to public sector activities). However, asymmetries or heterogeneity among integrating countries are prevalent, and it is thus an open question how such asymmetries affect the abovementioned channels. This paper considers these issues in a simplified version of Andersen and Sørensen (2012) using specific functional forms and numerical analysis to address the role of asymmetries across integrating countries. Two questions are of particular interest. If a country with a large public sector (strong preferences for public consumption) integrates with a country with a smaller

[6] See also Andersen and Sørensen (2011).

public sector, is the former country to a larger extent under pressure from the competitiveness-effect? Does this lead to a convergence of public sector size to the lower of the two levels (race to the bottom)? Related, what role do differences in productivities (and thus income levels) play for the effects of product market integration on public sector activities?

In policy debates it is a widespread idea that if taxes harm competitiveness, it is to be expected that countries acting non-cooperatively choose too low taxes (a race to the bottom) and thus the level of public sector activities is too low. However, a very robust result from explicit general equilibrium models is that countries acting non-cooperatively tend to choose too high levels of public activities and thus taxes. The reason is that countries perceive that they can affect the terms of trade to their advantage. This effect is not present in the cooperative case, and therefore there is an upward bias in taxes determined in the non-cooperative case [see e.g. Chari and Kehoe (1990), Devereux (1991), Turnovsky (1988), van der Ploeg (1987, 1988), and Andersen et al. (1996)]. Epifani and Gancia (2009) build on this literature and show in a model with specific functional forms, exogenous labour supply and exogenous production/specialization structures how globalization may increase public sector activity, and they present empirical evidence in support of this finding. In Andersen and Sørensen (2012) we showed in a rather general setting with endogenous labour supply and production/specialization structure that this non-cooperative bias not only applies to public consumption but also to transfers. Below we show that the bias becomes stronger with tighter product market integration; i.e. the non-cooperative tax rate is higher than the cooperative tax rate, and the differences are higher the more markets are integrated.

The rest of the paper is organized as follows: Sect. 2 sets up the basic structure of the Ricardian trade model with trade frictions and a public sector and shows that there exists a unique equilibrium. Section 3 considers the symmetric case paying particular attention to the upward bias in tax policy and how it is affected by product market integration. Section 4 considers the effects of product market integration when countries are asymmetric, and finally Sect. 5 offers a few concluding remarks. Appendices provide further technical material and proofs.

2 The Model

We set up a highly stylized two-country economy applying standard textbook functional forms. We consider a standard Ricardian trade model [see Dornbusch et al. (1977)] with a public sector financed by labour income taxation. The structures of the two countries are identical and therefore we only describe the structure of the home economy. However, parameters may differ across countries, and foreign variables and parameters are denoted with an asterisk ($*$).

2.1 Households

The economy consists of a continuum of homogenous households with unit mass. A household derives utility from private (B) and public (G) utility bundles in the following way

$$U = \frac{1}{1-\beta} B^{1-\beta} + \bar{G} \frac{1}{1-\lambda} G^{1-\lambda}; \quad \beta > 0, \lambda > 0 \tag{1}$$

The parameter $\bar{G}$ weights the relative importance of utility from public activities to the utility from private consumption. The private utility bundle is defined as[7]

$$B = C - kL^{\gamma}, \gamma > 1, k > 0 \tag{2}$$

where L is labour and C is the consumption bundle defined over a continuum of goods as

$$C = \left[\int_0^1 c\,(i)^{\frac{\theta-1}{\theta}}\, di \right]^{\frac{\theta}{\theta-1}}, \theta > 1$$

implying a price index given by

$$P \equiv \left[\int_0^1 q\,(i)^{1-\theta}\, di \right]^{\frac{1}{1-\theta}} \tag{3}$$

where $q\,(i)$ denotes the price of good i. Given the CES preferences over goods, demand for each good $i \in [0, 1]$ is given by

$$c_i = \left[\frac{q\,(i)}{P} \right]^{-\theta} \frac{I}{P} = q\,(i)^{-\theta}\, [P]^{\theta-1}\, I = q\,(i)^{-\theta} \left[\int_0^1 q\,(i)^{1-\theta}\, di \right]^{-1} I \tag{4}$$

where I denotes disposable income.

The household receives labour income (WL), profits (Π) and public transfers (TR). Labour income and profits[8] are taxed at a constant marginal tax rate t, thus

$$I = [1-t]\,[WL + \Pi] + TR \tag{5}$$

[7] This formulation implies no income effect in the labour supply decision and a constant labour supply elasticity of $\frac{1}{\gamma-1}$.

[8] In equilibrium profits are zero due to competitive markets. Hence it is irrelevant whether profits are taxed.

In the present set-up, the transfer part of public sector activities seems needless as the representative household framework precludes any welfare gains from redistribution. Yet, restricting public sector activities to transfers is interesting since it highlights the distortionary effects of taxation. In particular it allows us to identify effects driven by financing public expenditures (supply side effects) without mixing them up with the well-known expansionary bias due to expenditure switching (demand side effects).

To maximize utility (1) subject to the budget constraint (5) the representative household supplies the following amount of labour

$$L^s = \left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}} \tag{6}$$

We normalize the mass of households and thus the size of the home country to 1. The size of the foreign country is set to $n \gtreqless 1$.

2.2 Public Sector

The government may be engaged in providing transfers to households (TR) and in the production of public services (G). Public services are produced by use of labour (L^g) and it is assumed[9] that $G = L^g$ (productivity is constant and for simplicity normalized to one). These activities are financed by a proportional tax levied on income[10] by the rate t, and hence the budget constraint reads

$$t\,[WL + \Pi] = WL^g + TR \tag{7}$$

where L is total employment; i.e. $L = L^p + L^g$, where L^p (L^g) denotes labour used in the private (public) sector.

To allow for an easy way of analysing the two main activities of the government (transfers and public services), it is assumed that a fraction $\xi \in [0, 1]$ of the tax revenue is distributed as lump-sum transfers to individuals and the rest is used for public consumption/employment. For $\xi = 1$ we have a pure tax-transfer scheme without any aggregate demand effects, but only a supply side effect via the tax rate on income. This special case allows an identification of the pure distortion effect without mixing it up with other effects of public sector activities.

[9]The assumption implies that public activities are directed towards a non-tradeable, namely, labour. Notice that the assumption here to a first approximation captures the fact that about 2/3 of public consumption expenditures are wage expenditures.

[10]Observe that there is no profit in equilibrium due to competitive product markets and there is no issue as to whether labour and profit income should be taxed at different rates.

Using the public sector budget constraint and the fact that profits are zero in equilibrium disposable income can be written

$$I = [1-t]\,WL + \Pi + TR = WL^p \tag{8}$$

i.e. disposable income is determined by the income generated in the private sector.[11] Using that profits are zero it also follows from the public budget constraint (7) that

$$L^g = \frac{\left[1-\xi\right]t}{1-\left[1-\xi\right]t} L^p \tag{9}$$

Using (6) and (9) private and public consumption bundles, respectively, are given by

$$B = \left[\frac{W}{P}\right]^{\frac{\gamma}{\gamma-1}} \left[\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}} \left[\frac{\gamma-1}{\gamma}\,[1-t]+\xi t\right]$$

$$G = \left[1-\xi\right]t \left[\frac{W}{P}\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}$$

2.3 Firms

Firms are competitive and produce subject to constant returns production functions with labour as the only input; cf. the standard Ricardian trade model, i.e.

$$Y(i) = A(i)\,L(i) \tag{10}$$

Good specific productivity varies across countries. For each good i, let $A(i)$ and $A^*(i)$ denote domestic and foreign productivity in producing good i, respectively, and let $a_i \equiv \frac{A_i}{A_i^*}$ be relative productivity. Assume without loss of generality that goods are ordered such that a_i is increasing in i. To be specific assume that domestic (foreign) productivity of good i is given by $A(i) = Ae^{\alpha i}$ ($A^*(i) = A^* e^{\alpha[1-i]}$), and hence $a_i \equiv \frac{A_i}{A_i^*} = \frac{Ae^{\alpha[2i-1]}}{A^*}$, which is increasing in i.

[11] Note that the gross domestic product as conventionally measured in national accounts is given as (measured from the factor side)

$$GDP = WL^p + WL^g$$

2.3.1 Trade Structure

Due to trade frictions (Iceberg type) a firm has to ship $z \geq 1$ units in order to supply one unit to the export market. As a slight abuse of language we term z trade frictions in the following. Constant returns to scale and competitive markets imply marginal costs pricing. It is endogenous whether a good is traded.[12] Letting $w \equiv \frac{W}{W^*}$ denote the relative wage, good i is only produced domestically if

$$\frac{zW}{A(i)} < \frac{W^*}{A^*(i)} \Leftrightarrow a_i > wz \Leftrightarrow i > \frac{\ln\left(\frac{A^*}{A} wz\right)}{2\alpha} + \frac{1}{2} \equiv i^E(w,z),$$

only produced abroad if

$$\frac{W}{A(i)} > \frac{zW^*}{A^*(i)} \Leftrightarrow a_i < wz^{-1} \Leftrightarrow i < \frac{\ln\left(\frac{A^*}{A} wz^{-1}\right)}{2\alpha} + \frac{1}{2} \equiv i^H(w,z)$$

and produced in both countries and non-traded if

$$i \in \left[i^H(w,z), i^E(w,z)\right]$$

A higher relative wage implies higher i^E and i^H; that is, domestic firms both produce and export fewer types of goods. The intuition is straightforward since it derives from the worsening of wage competitiveness. The more integrated the markets (lower z), the higher i^H and the lower i^E; i.e. with lower trade frictions fewer goods are produced domestically, but more goods are exported. In other words, the number of non-tradeable goods decreases and production is further specialized in goods in which the country has a comparative advantage. Hence, changes in both wage competitiveness (endogenous) and the trade friction (exogenous) cause a change in the trade and specialization structure.

We will only consider equilibria with international trade; i.e. some goods are tradeable and thus impose the following restrictions[13]:

1. Home must export: $i^E(w,z) < 1 \Leftrightarrow w < \frac{A}{A^*} \frac{e^\alpha}{z}$
2. Home must import: $i^H(w,z) > 0 \Leftrightarrow w > \frac{A}{A^*} \frac{z}{e^\alpha}$

[12] Given the Cobb-Douglas preferences endogeneity of the trade structure is the key to allow for cross-country variation in the size of the private sector.

[13] These restrictions imply that $z < e^\alpha$; i.e. they put an upper bound on trade frictions.

2.4 Equilibrium

The equilibrium is determined by solving for the relative wage which subsequently determines all other real variables. The equilibrium condition arises from inserting product market equilibria conditions into the aggregate labour demand relation and setting this equal to labour supply. This condition is equivalent to the balanced trade condition. See Appendix 1 for the derivation.

Throughout we present outcomes of a simple numerical analysis of the model either to illustrate results or to indicate possible effects where no analytical results can be readily found. We apply the following baseline parameter values[14] unless otherwise stated: $A = \alpha = k = \bar{G} = 1, \beta = \lambda = \theta = 3, \gamma = 4$ for both countries and $n = 1$.

3 Symmetric Countries

We start by considering the simple case where all exogenous variables/parameters of the countries are identical in the two countries except for the good specific productivities and potentially for endogenous policy parameters.

In the following we consider optimal policies under the assumption of a utilitarian social welfare function. We consider both non-cooperative and cooperative policy settings.

3.1 Equilibrium Wages and Competitiveness[15]

As a prelude to the subsequent analysis it is useful to note that lower trade frictions are associated with gains from trade. Consider the symmetric equilibrium where fiscal policies are the same in the two countries; i.e. $t = t^*$ and $\xi = \xi^*$ and thus $w = 1$. In this case a reduction in the trade friction (z) increases welfare for given fiscal policies. The utility gain has two components: increasing private (B) and public (G) utility bundles. The former captures standard gains from trade. The latter arises because one of the gains from integration is higher employment, which in turn means a higher tax base and thus revenue and therefore (under the balanced budget constraint) more public sector activities. Alternatively, to maintain a given level of public services, there would be room for a tax reduction, which in turn would increase the private utility bundle further. In the following we will refer to the mechanism that integration expands the tax base and thus gives room for a tax

[14] These values imply risk aversion measures of 3 and a labour supply elasticity of $\frac{1}{3}$.

[15] For proof of results on this section see Andersen and Sørensen (2012).

reduction as the tax base effect. The gains from trade are also reflected in the fact that the real wage ($\frac{W}{P}$) is decreasing with trade frictions (z).

More interesting for the subject of this paper is the fact that a unilateral increase in the tax rate (t) raises the relative wage (w), which in turn deteriorates wage competitiveness but improves the terms of trade. Intuitively a large domestic (relative to foreign) tax burden causes a lower domestic labour supply, and this tends to increase the relative wage. This confirms the common perception that an increase in the tax rate tends to increase relative wages and thereby worsen wage competitiveness, cf. introduction. The increase in the relative wage (reduction in competitiveness) causes a reduction in the number of goods exported and produced in the domestic country; i.e. i^E and i^H increase, and in line with public concerns production moves abroad. This underlines that the mechanisms often highlighted in public debates (cf. the introduction) are supported by the present analysis. However, the higher relative wage, and thus reduced competitiveness, also implies a terms of trade improvement and allows the country to specialize in goods in which it has stronger comparative advantages.

Cross-country variations in both size and composition of the public sector activities are large. In policy debates it is often taken for granted that a large public sector is tantamount to a worsening of the competitive position, and partial models confirm this. With the present framework, we can analyse two key asymmetries, namely differences in size and composition of the public sector. Consider first size. If $t > t^*$ and $\xi = \xi^*$, i.e. the domestic public sector is more extended than the foreign, but the relative composition (services and transfers) is the same, it follows that home is less competitive than foreign, $w > 1$. Turning to composition we have that if $\xi > \xi^*$ and $t = t^*$, i.e. the domestic public sector is relatively more focused on transfers than on public services compared to foreign, but the size of the public sector is the same, it follows that home is more competitive than foreign, $w < 1$.

Intuitively a large domestic (relative to foreign) tax burden causes a lower domestic labour supply, and this tends to increase the relative wage. When a country focuses more on transfers than on provision of public services, the labour requirement of the public sector is lower (compared to foreign). The lower public labour requirement increases labour supply to the private sector and thereby reduces the relative wage. These results stress that both the size and the composition of the public sector/fiscal policy are important for competitiveness. An important implication is that a country with a relatively large public sector may indeed be competitive provided it uses a relatively high fraction of tax revenue on transfers.[16]

[16]Note that public consumption is assumed to be "pure" consumption having no direct effect on labour supply or productivity.

3.2 Marginal Costs of Public Funds (Pure Transfer Case, $\xi = 1$)

The problem of choosing the optimal tax rate (t) under a constraint that a real revenue of $\widehat{T}$ should be collected can be formulated as the following Lagrange problem

$$\max_t \Upsilon = U(B,0) + \mu \left[tR - \widehat{T} \right],$$

where μ is the Lagrange multiplier measuring the effects on utility of changing the revenue requirement. The first order condition reads

$$U_B B_t + \mu \left[R + tR_t \right] = 0$$

and the marginal costs of public funds measured in monetary equivalents are

$$MCPF \equiv \frac{\mu}{U_B} = -\frac{B_t}{R + tR_t} \tag{11}$$

where R denotes real income generated in the private sector; i.e. $R \equiv \frac{W}{P}L$. The $MCPF$ thus measures how much real income private households lose if the real income going to the public sector increases by one unit. In the cooperative equilibrium marginal costs of public funds are given by (see Appendix 2 for a derivation)

$$MCPF = \frac{t^{\frac{1}{\gamma-1}\frac{1}{1-t}}}{1 - t^{\frac{1}{\gamma-1}\frac{1}{1-t}}}$$

which increases with trade frictions for a given revenue constraint and is invariant to trade frictions for a given tax rate. The $MCPF$ increase in the tax rate. Accordingly, as higher trade frictions imply a lower tax base and thus a higher tax rate for a given revenue constraint, it follows directly that $MCPF$ increase with trade frictions.

In the pure transfer case all public resources are devoted to redistribution of income. Optimal policy may in this case at first seem trivial due to the representative agent framework. However, as we will see, this is not the case due to a terms of trade effect of taxation in the non-cooperative policy case. This case also describes the situation in which public consumption is waste,[17] i.e. $\bar{G} = 0$. Finally this case is a stepping stone to the more complicated case of optimal public consumption where integration also affects taxation due to relative changes in marginal utilities of private and public consumption.

[17] We have that $MCPF^{\text{Publicconsumption}} = MCPF^{\text{Transfers}} + 1$.

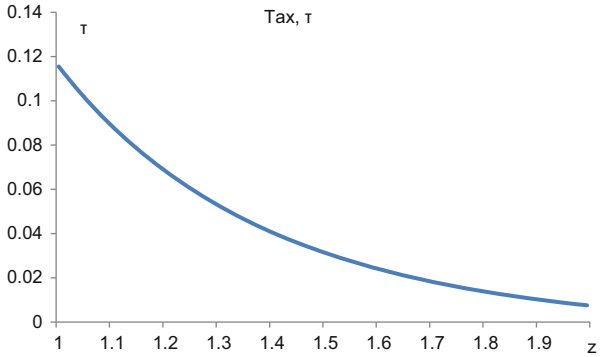

Fig. 1 Tax rate in non-cooperative equilibrium, transfers only

For cooperative policy makers the optimal tax rate is zero as the income tax is distortionary and there are no gains from redistribution (representative agent setting). Non-cooperative policy makers have incentives to impose distortionary labour taxes in order to obtain a terms of trade advantage. In the symmetric equilibrium the terms of trade effect will never be realized for the countries. The optimal cooperative tax rate is zero while the optimal non-cooperative tax rate is positive; hence there is an upward bias in non-cooperative tax policy.

Is the non-cooperative bias larger with more product market integration? Figure 1 displays the optimal tax rate as a function of the trade cost. It is seen that a lower trade friction leads to a higher tax rate in the non-cooperative case, and hence the difference between the cooperative ($t = 0$) and non-cooperative equilibrium widens; i.e. the bias in non-cooperative policies is larger the tighter product markets are integrated.

3.3 Public Consumption ($\xi = 0$)

Now we consider the scenario with no redistribution of income; i.e. all public revenue is devoted to public consumption. If marginal utility of public consumption is sufficiently high, this corresponds to optimal policy where both t and ξ are chosen optimally.

Cooperative Policies

Cooperative policy makers set the tax rate in order to maximize welfare (see Appendix 2 for details). The optimality condition reads

$$\Gamma = B^{-\beta}\frac{\partial B}{\partial t} + \bar{G}G^{-\lambda}\frac{\partial G}{\partial t} = 0$$

$$\Gamma_t < 0$$

We have that the effect of integration on optimal taxation is ambiguous since

$$sign\frac{dt}{dz} = sign\left(\gamma\left[1-\beta\right]-\left[1-\lambda\right]\right)$$

Note that γ is the relative risk aversion measure for the disutility from work, $1-\lambda$ measures the relative risk aversion for public consumption, and $1-\beta$ measures the relative risk aversion for private consumption. Hence, the sign $\frac{dt}{dz}$ depends on the weighting of disutility to work compared to the relative value of public consumption to private consumption. However, the effect of integration on public consumption is more complex as public consumption increases with integration for a given tax rate due to gains from trade. Public consumption increases with integration unless $\beta < 1-\frac{1}{\gamma}\frac{1}{1-t\frac{\gamma}{\gamma-1}} < 1$ where t is the (endogenous) optimal tax rate.[18] It follows that for $\beta \in \left(1-\frac{1}{\gamma}\frac{1}{1-t\frac{\gamma}{\gamma-1}}, 1-\frac{1-\lambda}{\gamma}\right)$ public consumption increases although the tax rate decreases.

3.3.1 The Case of Non-cooperative Policy

In this case the optimality condition reads

$$\Psi = B^{-\beta}\left[\frac{\partial B}{\partial t}+\frac{\partial B}{\partial w}\frac{dw}{dt}\right]+\bar{G}G^{-\lambda}\left[\frac{\partial G}{\partial t}+\frac{\partial G}{\partial w}\frac{dw}{dt}\right]=0$$

As $\Psi = \Gamma + B^{-\beta}\frac{\partial B}{\partial w}\frac{dw}{dt}+\bar{G}G^{-\lambda}\frac{\partial G}{\partial w}\frac{dw}{dt} > \Gamma$, it follows that the tax is higher in the non-cooperative case. Hence as in the pure transfer case we have an upward bias in tax/fiscal policy. It turns out that no clear analytical results on the effect of market integration on the optimal non-cooperative tax are available.

Figure 2 illustrates the optimal tax rate in the cooperative and non-cooperative case.[19] It is seen that the tax is increasing in both the cooperative and the non-cooperative case when markets integrate (lower trade friction z). The non-cooperative bias is also larger, the more integrated product markets are.

[18] Unfortunately we cannot get a closed solution for the optimal tax rate.

[19] Note that other parameter values may imply that the tax is decreasing when the trade cost z decreases, cf. expression above.

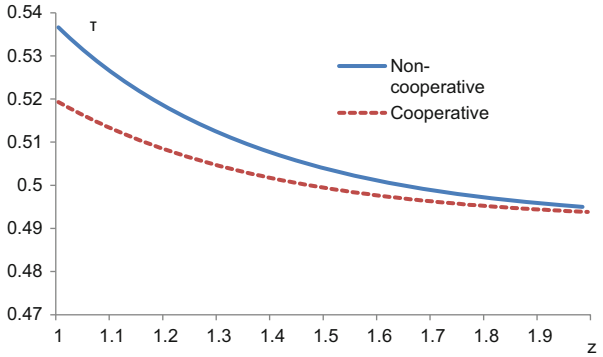

Fig. 2 Tax rate in non-cooperative and cooperative equilibrium, public consumption only

4 Asymmetric Countries

The integration process is proceeding among countries which in various ways differ, and therefore we now turn to a consideration of heterogeneities or asymmetries across countries. For the present purpose it is particularly interesting to consider the implications of asymmetries with respect to preferences for public consumption ($\overline{G}$). In countries—like the Nordic—with a relatively large public sector it is an important question how integration with other countries with a smaller public sector affects the need and scope for public sector activities. We consider this question below and also comment on the implications of asymmetries arising from productivity (A). In case of asymmetric countries we only consider non-cooperative policies. It is not obvious how to distribute gains from cooperative policies and furthermore in the absence of lump-sum transfers side payments distort the economy. Accordingly the cooperative case becomes quite complex. Despite the simplicity of the model only few analytical results can be obtained in the case of country heterogeneities and therefore the following numerical results are suggestive in terms of possible results.

4.1 Preferences for Public Consumption

A straightforward way to model differences in preferences for public consumption and thus the size of the public sector measured in terms of public consumption is via $\overline{G}$. If $\bar{G}^* < \overline{G}$, it follows that the home country has a stronger preference for public consumption relative to private consumption than the foreign country, and vice versa.

To what extent are policies affected by integration with a country with a weaker (stronger) preference for public consumption than the home country? How do foreign preferences affect the home country for a given level of product

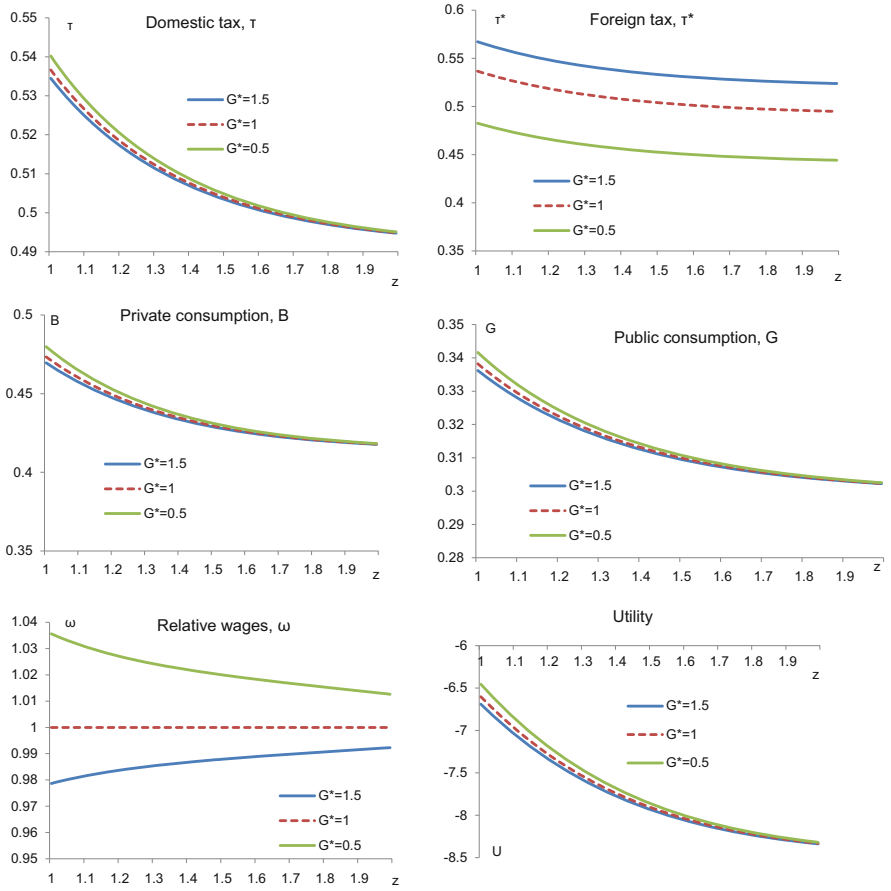

Fig. 3 Non-cooperative equilibrium—the role of preferences for public consumption

market integration, and how are these aspects influenced by further product market integration?

Figure 3 above has the parameter for public consumption in the home country constant ($\overline{G} = 1$) and considers three different levels of foreign preferences for public consumption; weaker ($\bar{G}^* = 0.5$), identical ($\bar{G}^* = 1$), and stronger ($\bar{G}^* = 1.5$).

Consider first the home country implications of foreign preferences for public consumption. It is seen from Fig. 3 that weaker government consumption preferences in the foreign country tend to imply a larger tax and thus level of public consumption in the home country. The levels of public consumption are therefore strategic substitutes. The reason is that a weaker preference for public consumption in the foreign country leads to a lower tax rate and hence a larger labour supply for the private sector. This increases private consumption in the home country, and thus

leads to higher public consumption. This is a striking result since integration with a country with a lower preference for public consumption often in the public debate is taken to imply a "race to the bottom"; that is, a convergence to tax rates and levels of public consumption dictated by the countries with the lowest preferences for public sector activities.

It is seen from Fig. 3 that further product market integration (lower z) leads to a higher tax in both countries. The response is non-linear. As above the tax is fairly insensitive to the product market integration when trade frictions are high, but becomes more sensitive at lower levels of trade frictions. It also means that although product market integration in the past may not have had significant effects on tax rates, it may have so in the future with further product market integration.

It is seen that if the foreign country has a weak preference for public consumption, wage competitiveness of the home country is worse ($\omega > 1$) than if the foreign country has a strong preference for public consumption. The difference is non-linear in the trade-friction, and largest when product markets are most tightly integrated.

The utility consequences are shown in Fig. 3. As expected for highly non-integrated product markets the utility consequences for the home country of the preferences for public consumption in the foreign country are marginal. Even for tightly integrated product markets the consequences are rather small compared to the gains from trade. This indicates that the utility consequences of the spill-over effects running via the public sector (and differences herein across countries) are of marginal importance.

4.2 Heterogeneity in Productivity

An important dimension of integration is that of countries with different income levels which in the present setting can be captured by different productivity levels. Below we consider variation in foreign productivity where the parameter $A^*$is set at 0.85, 1 and 1.15, respectively (home productivity: $A = 1$).

The main results are illustrated in Fig. 4. The tax in the home country is higher, the higher productivity in the foreign country. The reason is that the higher productivity in the foreign country tends to increase foreign production. As a consequence foreign goods become cheaper, and this leads to an increase in private consumption at home. Since private and public consumption are substitutes in preferences, it follows that there is a shift towards public consumption, and hence the result. Therefore, the higher productivity in the foreign country, the larger the level of both private and public consumption at home. Utility at home is higher when foreign productivity is high.

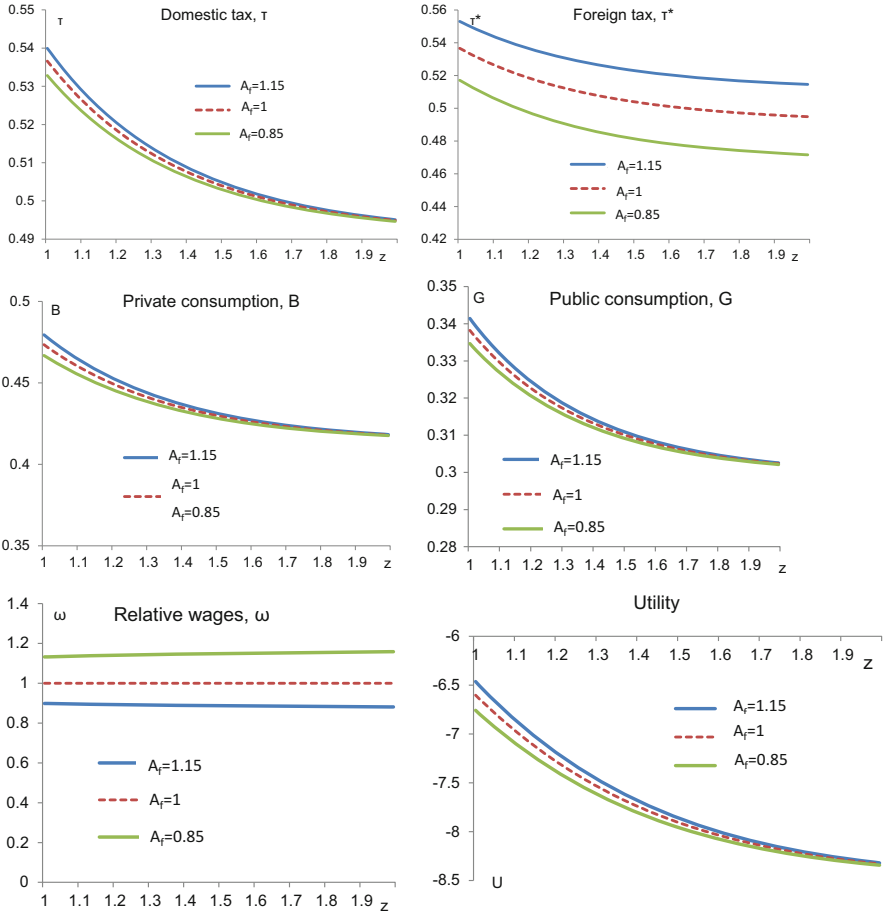

Fig. 4 Non-cooperative equilibrium—the role of aggregate productivity

5 Concluding Remarks

We have considered the role of product market integration for public sector activities (transfers and public consumption). Although the analysis supports that a larger public sector (higher tax rate) tends to increase wages and worsen wage competitiveness, the implications of product market integration for the public sector are far from straightforward. The reason is a "gains from trade" effect which tends to increase the tax base and decrease the opportunity costs of public consumption (marginal utility of private consumption falls). It is worth stressing that the present analysis builds on standard assumptions in trade theory and open macroeconomics. It follows that the retrenchment view that product market integration inevitably

leads to a downward pressure on public sector activities does not get unconditional support in a standard setting.

A particularly noteworthy finding is that a country with a large public sector (strong preferences for public consumption) may benefit more by integrating with a country with a smaller public sector. Likewise integrating with a country with a higher productivity (income) level is not necessarily detrimental to public sector activities. The terms of trade effects of public sector activities are crucial to these effects. The terms of trade effect is also the reason that non-cooperative policies tend to imply more public sector activity (transfers or public consumption) than in the cooperative case, and we show that the non-cooperative bias may be strengthened with further product market integration.

This paper has only considered one aspect of the globalization process, namely, product market integration. Clearly globalization is a wider concept, and other aspects may challenge the public sector through different channels. One such mechanism is increased factor mobility forcing tax reductions on the mobile factors and thus causing a revenue drag.[20] This can be interpreted as a direct threat to the public sector in the sense that the market enforces a change in policies [see e.g. Tanzi (2002) and Razin et al. (2011)]. A significant increase in migration coupled with selection mechanisms such that there is net-inflow of less skilled and net-outflow of high skilled may be a serious threat to a tax financed welfare model. It is an interesting topic for future research to integrate these mechanisms in a framework also highlighting the effects considered in this paper.

Acknowledgements Comments from an anonymous referee are gratefully acknowledged.

Appendix 1

In this appendix we derive the main equations of the model and prove existence and uniqueness of the equilibrium for given policy variables; i.e. for given t, t^*, ξ and ξ^*.

Consumer Prices

Markets are perfectly competitive implying that firms set prices equal to marginal costs and thus home market equilibrium consumer prices are

$$q(i) = \begin{cases} z\frac{W^*}{A_i^*} = \frac{zW^*}{A^* e^{\alpha(1-i)}} & \text{if } i \in M \Leftrightarrow i < \frac{\ln\left(\frac{A^*}{A}w\right) - \ln z}{2\alpha} + \frac{1}{2} = i^H(w,z) \\ \frac{W}{A_i} = \frac{W}{Ae^{\alpha i}} & \text{if } i \in H \Leftrightarrow i \geq \frac{\ln\left(\frac{A^*}{A}w\right) - \ln z}{2\alpha} + \frac{1}{2} = i^H(w,z) \end{cases}$$

[20] For evidence on such downward competitiveness on taxes see e.g. EAAG (2007).

and similar for foreign consumer prices

$$q^*(i) = \begin{cases} \frac{W^*}{A_i^*} = \frac{W^*}{A^* e^{\alpha(1-i)}} \text{ if } i \in H^* \Leftrightarrow i < \frac{\ln\left(\frac{A^*}{A} w\right) + \ln z}{2\alpha} + \frac{1}{2} = i^E(w,z) \\ z\frac{W}{A_i} = \frac{zW}{Ae^{\alpha i}} \quad \text{if } i \in M^* \Leftrightarrow i \geq \frac{\ln\left(\frac{A^*}{A} w\right) + \ln z}{2\alpha} + \frac{1}{2} = i^E(w,z) \end{cases}$$

The consumer price index accordingly becomes

$$P = \left[\int_0^1 q(i)^{1-\theta} di\right]^{\frac{1}{1-\theta}} = \left[\left[zW^*\right]^{1-\theta} \int_0^{i^H(w,z)} \left[A_i^*\right]^{\theta-1} di + W^{1-\theta} \int_{i^H(w,z)}^1 A_i^{\theta-1} di\right]^{\frac{1}{1-\theta}}$$

$$= \frac{zW^*}{A^*} \left[\frac{1}{\alpha\left[\theta-1\right]} \left[e^{\alpha[\theta-1]} \left[1 + \left[\frac{A^*}{A}\frac{w}{z}\right]^{1-\theta}\right] - 2e^{\frac{1}{2}\alpha[\theta-1]} \left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}\right]\right]^{\frac{1}{1-\theta}}$$

Similar for foreign

$$P^* = \frac{zW}{A} \left[\frac{1}{\alpha\left[\theta-1\right]} \left[e^{\alpha[\theta-1]} \left[1 + \left[\frac{A}{A^*}\frac{1}{wz}\right]^{1-\theta}\right] - 2e^{\frac{1}{2}\alpha[\theta-1]} \left[\frac{A}{A^*}\frac{1}{wz}\right]^{\frac{1-\theta}{2}}\right]\right]^{\frac{1}{1-\theta}}$$

Consumption Bundles

Private consumption bundle

$$B = C - kL^\gamma = \frac{I}{P} - kL^\gamma = \frac{W}{P}L^p - kL^\gamma$$

Using $L^p = \left[1 - \left[1-\xi\right]t\right]L$ and $L = \left[\frac{W}{P}\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}$ we have

$$B = \left[\frac{W}{P}\right]^{\frac{\gamma}{\gamma-1}} \left[\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}} \left[\frac{\gamma-1}{\gamma}[1-t] + \xi t\right]$$

Public consumption bundle

$$G = L^g = \frac{\left[1-\xi\right]t}{1-\left[1-\xi\right]t} L^p = \left[1-\xi\right]tL = \left[1-\xi\right]t\left[\frac{W}{P}\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}$$

Similar for the foreign country

$$B^* = \left[\frac{W^*}{P^*}\right]^{\frac{\gamma}{\gamma-1}} \left[\frac{1-t^*}{k\gamma}\right]^{\frac{1}{\gamma-1}} \left[\frac{\gamma-1}{\gamma}\left[1-t^*\right]+\xi^* t^*\right]$$

$$G^* = \left[1-\xi^*\right] t^* \left[\frac{W^*}{P^*}\frac{1-t^*}{k\gamma}\right]^{\frac{1}{\gamma-1}}$$

Consumption and Budget Shares

Given the CES preferences over goods, demand for each good $i \in [0, 1]$ is given by

$$c_i = \left[\frac{q(i)}{P}\right]^{-\theta} \frac{I}{P} = q(i)^{-\theta} [P]^{\theta-1} I = q(i)^{-\theta} \left[\int_0^1 q(i)^{1-\theta}\, di\right]^{-1} I \qquad (12)$$

The budget share of good i accordingly reads

$$e_i = \frac{q(i)\, c_i}{I} = \frac{q(i)^{1-\theta}}{\int_0^1 q(i)^{1-\theta}\, di}$$

The similar expressions for the foreign country read

$$c_i^* = q^*(i)^{-\theta} \left[\int_0^1 q^*(i)^{1-\theta}\, di\right]^{-1} I^*$$

$$e_i^* = \frac{q^*(i)^{1-\theta}}{\int_0^1 q^*(i)^{1-\theta}\, di}$$

We have the share of home income spent on home goods

$$N^H = \int_{i^H}^1 e_i di = \int_{i^H}^1 \frac{q(i)^{1-\theta}}{\int_0^1 q(i)^{1-\theta}\, di} di = \frac{\int_{i^H}^1 q(i)^{1-\theta}\, di}{\int_0^{i^H} q(i)^{1-\theta}\, di + \int_{i^H}^1 q(i)^{1-\theta}\, di}$$

$$= \frac{1}{\frac{\int_0^{i^H} q(i)^{1-\theta} di}{\int_{i^H}^1 q(i)^{1-\theta} di} + 1} = \frac{1}{\frac{\int_0^{i^H} \left[\frac{zW^*}{A_i^*}\right]^{1-\theta} di}{\int_{i^H}^1 \left[\frac{W}{A_i}\right]^{1-\theta} di} + 1} = \frac{1}{\left[\frac{z}{w}\right]^{1-\theta} \frac{\int_0^{i^H(w,z)} \left[A_i^*\right]^{\theta-1} di}{\int_{i^H(w,z)}^1 A_i^{\theta-1} di} + 1}$$

Share of home income spent on foreign goods

$$N^M = \int_0^{i^H} e_i di = \int_0^{i^H} \frac{q(i)^{1-\theta}}{\int_0^1 q(i)^{1-\theta} di} di = \frac{\int_0^{i^H} q(i)^{1-\theta} di}{\int_0^{i^H} q(i)^{1-\theta} di + \int_{i^H}^1 q(i)^{1-\theta} di}$$

$$= \frac{1}{1 + \frac{\int_{i^H}^1 q(i)^{1-\theta} di}{\int_0^{i^H} q(i)^{1-\theta} di}} = \frac{1}{1 + \frac{\int_{i^H}^1 \left[\frac{w}{A_i}\right]^{1-\theta} di}{\int_0^{i^H} \left[\frac{zW^*}{A_i^*}\right]^{1-\theta} di}} = \frac{1}{1 + \left[\frac{w}{z}\right]^{1-\theta} \frac{\int_{i^H}^1 A_i^{\theta-1} di}{\int_0^{i^H} \left[A_i^*\right]^{\theta-1} di}}$$

Consider the derivative w.r.t. the relative wage, w

$$\frac{\partial N^H}{\partial w} = -\left[N^H\right]^2 \left[\frac{z}{w}\right]^{1-\theta} \frac{\int_0^{i^H(w,z)} \left[A_i^*\right]^{\theta-1} di}{\int_{i^H(w,z)}^1 A_i^{\theta-1} di} \left[\frac{\theta-1}{w} + \left(\frac{\left[A_{i^H(w,z)}^*\right]^{\theta-1}}{\int_0^{i^H(w,z)} \left[A_i^*\right]^{\theta-1} di} + \frac{A_{i^H(w,z)}^{\theta-1}}{\int_{i^H(w,z)}^1 A_i^{\theta-1} di} \right) \frac{\partial i^H(w,z)}{\partial w} \right] < 0$$

Now consider the share of foreign income spent on home goods

$$N^{*M} = \int_{i^E}^1 e_i^* di = \int_{i^E}^1 \frac{q^*(i)^{1-\theta}}{\int_0^1 q^*(i)^{1-\theta} di} di = \frac{\int_{i^E}^1 q^*(i)^{1-\theta} di}{\int_0^{i^E} q^*(i)^{1-\theta} di + \int_{i^E}^1 q^*(i)^{1-\theta} di}$$

$$= \frac{1}{\frac{\int_0^{i^E} q^*(i)^{1-\theta} di}{\int_{i^E}^1 q^*(i)^{1-\theta} di} + 1} = \frac{1}{\frac{\int_0^{i^E} \left[\frac{W^*}{A_i^*}\right]^{1-\theta} di}{\int_{i^E}^1 \left[z\frac{W}{A_i}\right]^{1-\theta} di} + 1} = \frac{1}{[zw]^{\theta-1} \frac{\int_0^{i^E(w,z)} \left[A_i^*\right]^{\theta-1} di}{\int_{i^E(w,z)}^1 A_i^{\theta-1} di} + 1}$$

Again consider the derivative w.r.t. the relative wage, w

$$\frac{\partial N^{*M}}{\partial w} = -\left[N^{*M}\right]^2 [zw]^{\theta-1} \frac{\int_0^{i^E(w,z)} \left[A_i^*\right]^{\theta-1} di}{\int_{i^E(w,z)}^1 A_i^{\theta-1} di} \left[\frac{\theta-1}{w} + \left(\frac{\left[A_{i^E(w,z)}^*\right]^{\theta-1}}{\int_0^{i^E(w,z)} \left[A_i^*\right]^{\theta-1} di} + \frac{A_{i^E(w,z)}^{\theta-1}}{\int_{i^E(w,z)}^1 A_i^{\theta-1} di} \right) \frac{\partial i^E(w,z)}{\partial w} \right] < 0$$

For the assumed productivities we have

$$N^H = \left[\left[\frac{z}{w}\right]^{1-\theta} \left[\frac{A^*}{A}\right]^{\theta-1} \frac{e^{\frac{1}{2}\alpha(\theta-1)} - \left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}}{e^{\frac{1}{2}\alpha(\theta-1)} - \left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{\theta-1}{2}}} + 1 \right]^{-1}$$

$$N^M = \left[1 + \left[\frac{w}{z}\frac{A^*}{A}\right]^{1-\theta} \frac{e^{\frac{1}{2}\alpha[\theta-1]} - \left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{\theta-1}{2}}}{e^{\frac{1}{2}\alpha[\theta-1]} - \left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}} \right]^{-1}$$

$$N^{*M} = \left[[zw]^{\theta-1} \left[\frac{A^*}{A}\right]^{\theta-1} \frac{e^{\frac{1}{2}\alpha[\theta-1]} - \left[\frac{A^*}{A}wz\right]^{\frac{1-\theta}{2}}}{e^{\frac{1}{2}\alpha[\theta-1]} - \left[\frac{A^*}{A}wz\right]^{\frac{\theta-1}{2}}} + 1 \right]^{-1}$$

By definition $N^M = 1 - N^H$ and $N^{*H^*} = 1 - N^{*M}$.

Labour Demand

Labour demand in the private sector can be written

$$\begin{aligned} L^{dp} &= \int_{i^H}^{1} \frac{1}{A(i)} c(i)\, di + \int_{i^E}^{1} \frac{z}{A(i)} c^*(i)\, di \\ &= L^p \int_{i^H}^{1} e(i)\, di + \frac{1}{w} L^{*p} \int_{i^E}^{1} e^*(i)\, di \\ &= L^p N^H + \frac{1}{w} L^{*p} N^{*M} \end{aligned}$$

Using that $N^M = 1 - N^H$ and that $L^{dp} = L^p$ we obtain in equilibrium that

$$L^p = L^{dp} = \frac{1}{w} L^{*p} \frac{N^{*M}}{N^M}$$

Real Wage

Using the expression for P the real wage can be calculated

$$\frac{W}{P} = Ae^{\frac{1}{2}\alpha}\left[\frac{1}{\alpha\left[\theta-1\right]}\left[e^{\frac{1}{2}\alpha[\theta-1]}\left[1+\left[\frac{A^*}{A}\frac{w}{z}\right]^{\theta-1}\right]-2\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{\theta-1}{2}}\right]\right]^{\frac{1}{\theta-1}}$$

$$\frac{\partial\frac{W}{P}}{\partial\frac{w}{z}} = \frac{W}{P}\left[\frac{w}{z}\right]^{-1}\frac{\left[\frac{A^*}{A}\frac{w}{z}\right]^{\theta-1}\left[e^{\frac{1}{2}\alpha[\theta-1]}-\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}\right]}{e^{\frac{1}{2}\alpha[\theta-1]}\left[1+\left[\frac{A^*}{A}\frac{w}{z}\right]^{\theta-1}\right]-2\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{\theta-1}{2}}}$$

Note that $e^{\frac{1}{2}\alpha[\theta-1]}-\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}} > 0 \iff e^{\alpha} > \frac{A}{A^*}\frac{z}{w} \iff i^H(w,z) > 0$. Hence, $\frac{\partial\frac{W}{P}}{\partial\frac{w}{z}} > 0$, $\frac{\partial\frac{W}{P}}{\partial w} > 0$ and $\frac{\partial\frac{W}{P}}{\partial z} < 0$

Labour Supply

Labour supply for the private sector. Using $L^g = (1-\xi)\,tL$ and $L^p = L - L^g = \left[1-\left[1-\xi\right]t\right]L$ and

$$L^s = \left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}}$$

It follows that labour supply to the private sector is

$$\begin{aligned} L^{sp} &= \left[1-\left[1-\xi\right]t\right]\left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}} \\ &= \left[1-\left[1-\xi\right]t\right]\left[\left[\frac{1}{\alpha\left[\theta-1\right]}\left(e^{\alpha[\theta-1]}\left[1+\left[\frac{A^*}{A}\frac{w}{z}\right]^{1-\theta}\right]\right.\right.\right. \\ &\qquad \left.\left.\left.-2e^{\frac{1}{2}\alpha[\theta-1]}\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}\right)\right]^{\frac{1}{\theta-1}}\frac{wA^*}{z}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}} \end{aligned}$$

and

$$\frac{\partial L^{sp}}{\partial w} = \frac{1}{\gamma-1}\left[1-\left[1-\xi\right]t\right]\left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}}\frac{1-t}{k\gamma}\frac{\partial\frac{W}{P}}{\partial w} > 0$$

Equilibrium

Rewriting the labour market equilibrium condition we obtain

$$L^{dp} = L^p = L^{sp} \iff L^{dp} = \left[1-\left[1-\xi\right]t\right]L^s \iff \frac{1}{w}L^{*p}\frac{N^{*M}}{N^M} = L^p$$

$$\iff w\frac{\int_0^{i^H} e_i di}{\int_{i^E}^1 e_i^* di}\frac{\left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}}}{\left[\frac{W^*}{P^*}\frac{[1-t^*]}{k\gamma}\right]^{\frac{1}{\gamma-1}}} = n\frac{1-\left[1-\xi^*\right]t^*}{1-\left[1-\xi\right]t}$$

$$\iff w\frac{1-N^H}{1-N^{H*}}\frac{\left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}}}{\left[\frac{W^*}{P^*}\frac{[1-t^*]}{k\gamma}\right]^{\frac{1}{\gamma-1}}} = n\frac{1-\left[1-\xi^*\right]t^*}{1-\left[1-\xi\right]t}$$

Note that the left-hand side is strictly increasing as $\frac{\partial\frac{W}{P}}{\partial w}>0, \frac{\partial\frac{W^*}{P^*}}{\partial w}<0, \frac{\partial N^H}{\partial w}<0, \frac{\partial N^{H*}}{\partial w}>0$, cf. above. The strictly increasing left-hand side combined with the property that $i^H \to 0$ as $w \to \frac{A}{A^*}\frac{z}{e^\alpha}$ and $i^E \to 1$ as $w \to \frac{A}{A^*}\frac{e^\alpha}{z}$ ensure that the equilibrium relative wage exists and is unique. Inserting expressions from above this can be rewritten as

$$w^{1+\gamma}\left[\frac{e^{\alpha[\theta-1]}\left[1+\left[\frac{A^*}{A}\frac{w}{z}\right]^{1-\theta}\right]-2e^{\frac{1}{2}\alpha[\theta-1]}\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}}{e^{\alpha[\theta-1]}\left[1+\left[\frac{A}{A^*}\frac{1}{wz}\right]^{1-\theta}\right]-2e^{\frac{1}{2}\alpha[\theta-1]}\left[\frac{A}{A^*}\frac{1}{wz}\right]^{\frac{1-\theta}{2}}}\right]^{\frac{1}{\theta-1}}$$

$$\times\left[\frac{1+\left[zw\frac{A^*}{A}\right]^{\theta-1}\frac{e^{\frac{1}{2}\alpha[\theta-1]}-\left[\frac{A^*}{A}wz\right]^{\frac{1-\theta}{2}}}{e^{\frac{1}{2}\alpha[\theta-1]}-\left[\frac{A^*}{A}wz\right]^{\frac{\theta-1}{2}}}}{1+\left[\frac{w}{z}\frac{A^*}{A}\right]^{1-\theta}\frac{e^{\frac{1}{2}\alpha[\theta-1]}-\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{\theta-1}{2}}}{e^{\frac{1}{2}\alpha[\theta-1]}-\left[\frac{A^*}{A}\frac{w}{z}\right]^{\frac{1-\theta}{2}}}}\right]^{\gamma-1}$$

$$=\left[n\frac{1-\left[1-\xi^*\right]t^*}{1-\left[1-\xi\right]t}\right]^{\gamma-1}\frac{A\left[1-t^*\right]}{A^*\left[1-t\right]}$$

This determines the equilibrium relative wage

$$w = w\left(A, A^*, t, t^*, \xi, \xi^*, z, n\right)$$

which in turn determines all the other endogenous variables for given policy parameters.

Appendix 2

This appendix considers the case of symmetric countries

Marginal Cost of Public Funds (Pure Transfer Case, $\xi = 1$)

Marginal cost of public funds

$$MCPF \equiv \frac{\mu}{U_B} = -\frac{B_t}{R + tR_t}$$

Revenue generated by the private sector

$$R = \frac{W}{P} L^p = \frac{W}{P}\left[\frac{W}{P}\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}} = \left[\frac{[1-t]}{k\gamma}\right]^{\frac{1}{\gamma-1}}\left[\frac{W}{P}\right]^{\frac{\gamma}{\gamma-1}}$$

Consumption bundle

$$B = \left[\frac{W}{P}\right]^{\frac{\gamma}{\gamma-1}}\left[\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}\left(1-\frac{1-t}{\gamma}\right)$$

Consider the cooperative equilibrium ($w = 1 \implies \frac{\partial w}{\partial t} = 0$)

$$B_t \equiv \frac{\partial B}{\partial t} = -\left[\frac{W}{P}\right]^{\frac{\gamma}{\gamma-1}}\left[\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}\left[t\frac{1}{\gamma-1}\frac{1}{1-t}\right]$$

$$R + tR_t = \left[\frac{W}{P}\right]^{\frac{\gamma}{\gamma-1}}\left[\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}\left[1-t\frac{1}{\gamma-1}\frac{1}{1-t}\right]$$

Inserting into the expression for marginal cost of public funds we find

$$MCPF|_{cooperative} = \frac{t\frac{1}{\gamma-1}\frac{1}{1-t}}{1-t\frac{1}{\gamma-1}\frac{1}{1-t}}$$

It follows that

$$\frac{d\ MCPF|_{cooperative}}{dz} = \frac{\partial\ MCPF|_{cooperative}}{\partial t}\frac{\partial t}{\partial z} > 0$$

where $\frac{\partial t}{\partial z} > 0$ follows from the tax base effect.

Fiscal Consumption ($\xi = 0$)

Coordinated Public Policy

Assume $n = 1$. In a symmetric equilibrium $w = 1$ and hence the optimality condition reads

$$\begin{aligned}\Gamma &= B^{-\beta}\frac{\partial B}{\partial t} + \bar{G}G^{-\lambda}\frac{\partial G}{\partial t} = 0\\ \Gamma_t &< 0\end{aligned}$$

where

$$\begin{aligned}\Gamma &= -B^{1-\beta}\frac{\gamma}{\gamma-1}\frac{1}{1-t} + \bar{G}G^{1-\lambda}\left[\frac{1}{t} - \frac{1}{\gamma-1}\frac{1}{1-t}\right]\\ &= \bar{G}G^{1-\lambda}\frac{1}{t} - \left[\gamma B^{1-\beta} + \bar{G}G^{1-\lambda}\right]\frac{1}{\gamma-1}\frac{1}{1-t}\\ &= -\left[\frac{W}{P}[1-t]\right]^{\frac{\gamma[1-\beta]}{\gamma-1}}\left[\frac{1}{k\gamma}\right]^{\frac{1-\beta}{\gamma-1}}\left[\frac{\gamma-1}{\gamma}\right]^{1-\beta}\frac{\gamma}{\gamma-1}\frac{1}{1-t}\\ &\quad + \bar{G}t^{1-\lambda}\left[\frac{W}{P}\frac{1-t}{k\gamma}\right]^{\frac{1-\lambda}{\gamma-1}}\left[\frac{1}{t} - \frac{1}{\gamma-1}\frac{1}{1-t}\right]\end{aligned}$$

and where $\frac{1}{t} - \frac{1}{\gamma-1}\frac{1}{1-t} > 0$ in optimum, cf. the Laffer-curve. From this we have

$$\frac{dt}{dz} = -\frac{\Gamma_z}{\Gamma_t}$$

and thus

$$sign\frac{dt}{dz} = sign\Gamma_z$$

since $\Gamma_t < 0$, cf. the second order condition. We have that (using the expression for $\Gamma = 0$)

$$\Gamma_z = \left[\frac{W}{P}\frac{1-t}{k\gamma}\right]^{\frac{1-\lambda}{\gamma-1}}\frac{1}{\frac{W}{P}}\bar{G}t^{1-\lambda}\left[\frac{1}{t} - \frac{1}{\gamma-1}\frac{1}{1-t}\right]\frac{1}{\gamma-1}\left[-\gamma\left[1-\beta\right] + 1 - \lambda\right]\frac{d\frac{W}{P}}{dz}$$

Hence

$$sign\frac{dt}{dz} = sign\left[\left[-\gamma\left[1-\beta\right] + 1 - \lambda\right]\frac{d\frac{W}{P}}{dz}\right] = sign\left[\gamma\left[1-\beta\right] - \left[1-\lambda\right]\right]$$

and optimal taxation may be increasing or decreasing in product market integration depending on preferences and the labour supply elasticity. Note if $\gamma\left[1-\beta\right]-\left[1-\lambda\right]=0$, then t is independent of $\frac{W}{P}$ and accordingly of z. We have that an implicit solution to the optimal tax rate is given by

$$\left[\frac{W}{P}\left[1-t\right]\right]^{\frac{\gamma[1-\beta]-1+\lambda}{\gamma-1}}\left[\frac{1}{k\gamma}\right]^{\frac{\lambda-\beta}{\gamma-1}}\left[\frac{\gamma}{\gamma-1}\right]^{\beta}\frac{1}{1-t}=\bar{G}t^{1-\lambda}\left[\frac{1}{t}-\frac{1}{\gamma-1}\frac{1}{1-t}\right]$$

Regarding public consumption we have

$$\begin{aligned}sign\frac{dG}{dz}&=sign\frac{d\left[t\left[\frac{W}{P}\frac{1-t}{k\gamma}\right]^{\frac{1}{\gamma-1}}\right]}{dz}\\&=sign\left(t\left[\frac{W}{P}\right]^{\frac{1}{\gamma-1}}\left[1-t\right]^{\frac{1}{\gamma-1}}\left[\frac{1}{t}\frac{dt}{dz}-\frac{1}{\gamma-1}\frac{1}{1-t}\frac{dt}{dz}+\frac{1}{\gamma-1}\frac{1}{\frac{W}{P}}\frac{d\frac{W}{P}}{dz}\right]\right)\\&=sign\left(\frac{1}{t}\gamma\left[1-\beta\right]-\left[1-\beta\right]\frac{\gamma}{\gamma-1}\frac{1}{1-t}-\frac{1}{1-t}-\frac{1}{t}\right)\end{aligned}$$

Hence

$$\frac{dG}{dz}>0\Leftrightarrow\beta<1-\frac{1}{\gamma}\frac{1}{1-t\frac{\gamma}{\gamma-1}}$$

References

Alesina, A., & Perotti, R. (1997). The welfare state and competitiveness. *American Economic Review, 87*(5), 921–39.

Andersen, T. M., Rasmussen, B. S., & Sørensen, J. R. (1996). Optimal fiscal policies in open economies with labour market distortions. *Journal of Public Economics, 63*, 103–117.

Andersen, T. M., & Sørensen, A. (2011). Globalisation squeezes the public sector—Is it so obvious? *International Tax and Public Finance, 18*, 369–382.

Andersen, T. M., & Sørensen, A. (2012). Globalization, tax distortions, and public-sector retrenchment. *Scandinavian Journal of Economics, 114*, 409–439.

Botman, D., Laxton, D., Muir, D., & Romanov, A. (2006). *A New-Open-Economy-Macro Model for Fiscal Policy Evaluation*. IMF Working paper 06/45.

Bruce, N., & Purvis, D. D. (1985). The specification of goods and factor markets in open economy macroeconomic models. In R.W. Jones & P. B. Kenen (Eds.), *Handbook of international economics* (Chap. 16, Vol. II). Amsterdam: North-Holland.

Burda, M. (2000). European labour markets and the euro: How much flexibility do we really need? In Deutsche Bundesbank (Ed.), *The monetary transmission process*. New York: MacMillan.

Chari, V. V., & Kehoe, P. J. (1990). International coordination of fiscal policy in limiting economies. *Journal of Political Economy, 98*, 617–636.

Daveri, F., & Tabellini, G. (2000). Unemployment growth and taxation in industrial countries. *Economic Policy, 30*, 47–104.

Devereux, M. B. (1991). The terms of trade and the international coordination of fiscal policy. *Economic Inquiry, 29*, 720–736.

Dornbusch, R., Fischer, S., & Samuelson, P. A. (1977). Comparative advantage, trade and payments in a Ricardian model with a continuum of goods. *American Economic Review, 67*, 823–839.

Eaton, J., & Kortum, S. (2002). Technology, geography and trade. *Econometrica, 70*, 1741–1779.

EEAG (2007). *The EEAG report on the European economy*. Munich: CESifo.

Epifani, P., & Gancia, G. (2009). Openness, government size and the terms of trade. *Review of Economic Studies, 76*, 629–668.

Evers, M., De Moij, R. A., & van Vuuren, D. J. (2008). The wage elasticity of labour supply: A synthesis of empirical estimates. *De Economist, 156*, 25–43.

Lane, P. R., & Perotti, R. (2003). On the importance of the composition of fiscal policy: Evidence from different exchange rate regimes. *Journal of Public Economics, 87*, 2253–2279.

Layard, R., Nickell, S., & Jackman, R. (2005). *Unemployment: Macroeconomic performance and the labour market* (2nd ed.). Oxford: Oxford University Press.

Marston, R. C. (1985). Stabilization policies in open economies. In R.W. Jones & P. B. Kenen, *Handbook of international economics* (Chap. 17, Vol. II). Amsterdam: North-Holland.

OECD (2007). *Employment outlook*. Paris.

Razin, A., & Sadka, E. (2005). The decline of the welfare state: demography and globalization. CESifo book series. Cambridge: MIT.

Razin, A., Sadka, E., & Suwanskiri, B. (2011). *Migration and the welfare state. Political-economy formation*. Cambridge: MIT Press.

Rodrik, D. (1997). *Has globalisation gone too far?* Washington, DC: Institute for International Economics.

Tanzi, V. (2002). Globalization and the future of social protection. *Scottish Journal of Political Economy, 49*, 116–127.

Turnovsky, S. J. (1988). The gains from fiscal cooperation in the two-commodity real trade model. *Journal of International Economics, 25*, 14–27.

van der Ploeg, R. (1987). Coordination of optimal taxation in a two-country equilibrium model. *Economics Letters, 24*, 279–285.

van der Ploeg, R. (1988). International policy coordination in interdependent monetary economies. *Journal of International Economics, 25*, 1–23.

The WTO's Effect on Trade: What You Give is What You Get

Tristan Kohl

Abstract Do countries trade more when they participate in the World Trade Organisation (WTO)? After Rose's (Am Econ Rev 94:98–114, 2004) initial "non-effect", the literature has developed in several ways to re-examine this unexpected result. This paper gives a detailed overview of the developments and exposes the main biases that plague previous contributions. Using a dataset covering 181 countries for the period 1948–2007, we show that zero trade flows are best incorporated using (zero-inflated) negative binomial maximum likelihood estimation. We find that formal members gained more than non-member participants, and the level of WTO participants experienced gains go hand in hand with the extent of their multilateral liberalisation commitments. Developed nations gain more than developing or least developed countries, although poor countries do benefit from trading under the Generalised System of Preferences (GSP). We also correct for selection bias with respect to economic integration agreements and find, overall, that regionalism has a lower trade-promoting effect than WTO membership.

1 Introduction

The World Trade Organisation (WTO) is a forum for intergovernmental negotiations on trade regulation and non-discriminatory liberalisation. The multilateral trade system's ever-expanding membership base, combined with the steady growth of world trade has led to the conventional belief that WTO participation plays an important role in advancing international trade.

Curiously, an initial contribution by Rose (2004) does not support the notion that WTO membership stimulates world trade.[1] This unexpected finding is central to

[1]The change in international trade volumes induced by WTO participation is what will be called the "WTO effect" throughout this paper.

T. Kohl (✉)
Faculty of Economics & Business, University of Groningen, P.O. Box. 800, 9700 AV Groningen, The Netherlands
e-mail: t.kohl@rug.nl

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_19

the inquiry presented in this paper: Does participation in the WTO promote world trade? An extensive survey of the literature in Sect. 2 will be used to argue that Rose's results are biased and that there is a need to improve (1) the quality of the data and (2) the empirical strategy.

Ad 1. Several data improvements have already been put forward in various contributions in the literature. However, a strikingly large number of economic integration agreements (EIAs) is excluded from analysis. For example, Rose (2004) only uses ten EIAs out of a possible 200+. It will be shown that this omission introduces selection bias in the EIA parameter estimates and leads to incorrect inferences about the effect of policy on international trade.

Ad 2. Typical datasets on international trade include a large number of zeros. This poses a methodological problem for studies such as Rose (2004) which estimate a log-linear specification of the gravity equation of international trade with ordinary least squares (OLS) regression. In doing so, a substantial number of observations is ignored. More recent contributions to the WTO effect literature have adopted Tobit or Poisson maximum likelihood estimation (MLE) to include these zero trade "flows". However, these methods are vulnerable to estimation bias and it will be argued that negative binomial (NB) and zero-inflated negative binomial (ZINB) MLE are more suitable alternatives.

Details on how these two shortcomings are solved are addressed in Sect. 3, along with a presentation of the updated dataset. Section 4 revisits Rose's findings and shows that his remarkable results are not robust. Next, a full-fledged examination of the WTO's effect on trade is performed with the empirical tools called for in this chapter. Strikingly, our main finding is that the gains of WTO membership go hand in hand with the extent to which participants commit to extensive multilateral trade liberalisation. Our empirical evidence demonstrates that developed and developing nations have gained from WTO membership, while least developed countries (LDCs) have benefited from preferential market access under the Generalised System of Preferences (GSP).[2] Regionalism is also found to play a significant role in fostering international trade, although its effect is less extreme than suggested by earlier studies. Finally, Sect. 5 discusses the main findings in greater detail and concludes the paper.

2 Literature

Studies on the effect of WTO membership on world trade have only emerged after almost 60 years of multilateralism. Table 1 provides an overview of the main publications and their findings.

[2]Following the WTO's practice, this study considers LDCs and developing countries to be two distinct groups of countries.

Table 1 Results in the WTO effect literature

Variable	(1) Rose (2004)	(2) Subramanian and Wei (2007)	(3) Tomz et al. (2007)	(4) Felbermayr and Kohler (2006)	(5) Liu (2009)	(6) Herz and Wagner (2011)
Both in WTO	−0.04		0.54	0.17/0.05	0.47	0.62
	(0.05)		(0.06)	(0.06)/(0.02)	(0.04)	(0.02)
One in WTO	−0.06		0.27	0.13/0.06	0.21	
	(0.05)		(0.05)	(0.09)/(0.02)	(0.03)	
• Industrial		1.87				
		(0.14)				
• Developing		− 0.31				
		(0.04)				
– Importer						0.27
						(0.02)
– Exporter						0.39
						(0.02)
GSP granter						
• Importer	0.86	2.15	0.19		0.06	− 0.04
	(0.03)	(0.15)	(0.03)		(0.02)	(0.01)
• Exporter					0.22	0.01
					(0.02)	(0.01)
EIA	1.20	1.65	0.77	0.14/0.11	0.32	0.34
	(0.11)	(0.11)	(0.07)	(0.06)/(0.02)	(0.01)	(0.00)
Observations	234,597	76,094	234,597	42,542	1,119,372	697,223
R^2	0.65	0.75	0.85	0.45	N.A.	N.A.
Fixed effects	Year	Importer, exporter & year	Dyad & year	Importer, exporter & year	Dyad & year	Dyad & year
Method	OLS	OLS	OLS	Tobit (IM/EM)	Poisson	Poisson

Levels of significance are not reported due to missing data. Robust standard errors in brackets. IM means "intensive margin", EM means "extensive margin" and N.A. means "not applicable". Other estimates are omitted to save space

Rose (2004) performed the first comprehensive econometric study of the WTO effect, using a standard gravity model to specifically account for the effect of formal WTO membership and participation in GSP schemes on average trade. The outcome of Rose's study is a negative, not significant relationship between trade and formal WTO membership. Although the inclusion of country fixed effects yields a small, positive WTO effect, the paper does not provide robust statistical evidence that WTO membership has played a significant role in creating trade (column 1).

Subramanian and Wei (2007) use a gravity equation with time, importer and exporter fixed effects. Contrary to Rose (2004), these authors also take the following three asymmetries into account: (1) asymmetries between developed and developing countries (including LDCs); (2) asymmetries between developing countries that joined the WTO before and after the Uruguay Round; and (3) asymmetries between sectors in which the WTO has (not) liberalized trade barriers.

Ad 1: The WTO was designed as a two-tier organisation, with more liberalisation obligations for developed than developing country members. Non-tariff barriers (NTBs, especially quantitative restrictions) were not allowed. However, following the principle of special and differential treatment (S&D), developing nations had fewer obligations to liberalise. Furthermore, article XVIII:B (GATT) permitted developing nations to use NTBs and quotas for balance of payments reasons. These instruments were frequently used and resulted in diverging trade policies between developed and developing member states. It should be noted here that, compared to developed and developing nations, LDCs enjoy the most generous exemptions in terms of trade liberalisation commitments in the multilateral trade system.

Ad 2: Developing countries joining the WTO after the Uruguay Round received fewer exemptions than incumbent developing countries in order to close the gap between developed and developing countries in terms of their respective obligations to liberalise trade. Hence, there is an asymmetry between the effects of the WTO on (developing country) members acceding before and after the Uruguay Round.

Ad 3: This asymmetry arises from the fact that developed countries were not required to liberalise sectors such as agriculture, textiles and clothing. These countries were allowed to maintain higher tariffs and to implement quotas, amongst others, in the Multi-Fibre Arrangement (MFA).

Subramanian and Wei (2007) find that the WTO strongly increases merchandise imports for industrialised countries and developing nations that joined the multilateral trade system after the Uruguay Round by $e^{1.87} - 1 \approx 550\,\%$ (column 2). However, developing nations that joined the WTO before the Uruguay Round experienced small to no gains from membership due to the lack of their liberalisation commitments.

The authors also argue that the WTO does not by design aim at liberalising all industries. Membership of the WTO has the smallest impact on trade in protected industries such as agriculture and textiles, as opposed to larger gains in more

liberalised industries.[3] They conclude that dealing with institutional asymmetries and using a better econometric specification of the gravity equation solves Rose's "mystery" of not finding evidence that the WTO increases trade.

Tomz et al. (2007) point out that formal, de jure WTO membership is not the same as informal, de facto participation. In other words, Rose (2004)'s data on WTO membership assume that multilateral rules only applied to formal (de jure) members that were listed on the WTO membership roster. However, several de facto participants—whose names were not registered on that roster—also had rights and obligations under these rules and therefore need to be included in the WTO membership variable as well. These participants can be divided into three categories: colonies, newly independent states (NIS) and provisional members.[4]

The first category of de facto participation involves colonies. According to Article XXVI:5(a) GATT, every contracting party signed the Agreement on behalf of itself and the territories for which it had international responsibility (although exemptions were possible). Belgium, Denmark, the Netherlands, Portugal, Spain, the United Kingdom and the United States all applied GATT to virtually all their colonies upon signing the Agreement (Tomz et al., 2007, p. 2007).

The second category of de facto participation has bearing on decolonisation. What would happen with a colony's de facto participation upon becoming a NIS? The GATT provided two options: either become a formal member under the terms agreed to by the NIS' former metropolitan government, or attempt to become a formal member by negotiating new terms. Pending their decision, the GATT participants abode by the existing Agreement as long the NIS did the same. It was only upon the creation of the WTO in 1995 that NISs were forced to choose between acceding as formal members (e.g., Burkina Faso) or withdrawing (e.g., Kiribati and Tonga) (Tomz et al., 2007, p. 2008).

The final category of de facto participation involves provisional members. A number of states were treated as contracting parties even though negotiations for full accession had not yet been concluded. Provisional members who enjoyed this special arrangement before the transition to the WTO include Argentina, Colombia, Egypt, Iceland, Israel, Japan, the Philippines, Switzerland, Tunisia and Yugoslavia (Tomz et al., 2007, p. 2008).

Tomz et al. (2007) create a new WTO dummy variable that includes de facto participation and also corrects for a number of incorrect de jure specifications in Rose (2004). A standard gravity equation with time and country-pair (dyad) fixed effects is then estimated using OLS. The authors find a positive WTO effect.

[3] Unfortunately, the aggregate nature of the trade data in the present dataset does not allow for industry-specific analyses. However, studies focusing on a select number of countries, years and industries have found supporting evidence for WTO membership with respect to capital-intensive commodities (Engelbrecht and Pearce, 2007) and trade excluding agriculture, textiles and oil (Kim, 2010).

[4] The terms membership and participation are used interchangeably throughout this paper, but the context will be sufficiently specific when distinctions between de jure, formal membership and de facto, informal participation are necessary.

Specifically, the increase in trade is at least as large for non-member participants as it is for formal members. This result mainly stems from the fact that most non-member participants have historically had strong colonial ties with other WTO members/participants. Column 3 shows that if both trade partners are WTO participants, their trade is $e^{0.54} - 1 \approx 70\,\%$ higher than if neither country participates, all else constant.

An issue that had thus far been ignored is the incorporation of zero trade "flows". OLS estimates require a log-linear specification of the gravity equation, which means that the log of the dependent (trade) variable is also needed. However, as will be shown in Sect. 3.2, a significant number of trade observations is zero. The log-linear specification is problematic because $ln(0) = n.d.$, so these observations are dropped from the regression analysis and thereby introduce estimation bias.

The first study in the WTO effect literature to specifically address this problem is by Felbermayr and Kohler (2006), who use a corner-solutions Tobit method. They find a WTO effect at the intensive (extensive) margin of 19 (5) % if both trading partners are WTO members and 14 (6) % if only one is a member (column 4).[5]

Liu (2009) and Herz and Wagner (2011) use Poisson quasi-MLE to determine the WTO effect. Liu (2009) disentangles the WTO effect in terms of the intensive and extensive margins of trade and finds that if both trade partners are WTO members, their trade is 60 % (39 % intensive, 21 % extensive) higher than if neither is a member, all else constant. If only one of them is a WTO member, trade still is 23 % (8 % intensive, 15 % extensive) higher than if both are non-members (column 5).

Herz and Wagner (2011) are the first to follow Subramanian and Wei (2007) in distinguishing between importers and exporters. Moreover, they incorporate Tomz et al. (2007)'s improved WTO dummy variable and adopt Liu (2009)'s approach in using Poisson quasi-MLE to include the zeros in their analysis. They find that a country-pair's trade volume rises by 86 % if both countries are WTO participants. If the importer is a WTO participant and the exporter is not, trade increases by 48 %. Non-participants import 31 % more if their exporters are WTO participants, all else constant (column 6).

Despite these contributions to the literature, it will be shown that two issues still lead to potentially biased estimates. Table 2 presents an overview of the methodological developments in the WTO effect literature. Almost all studies use datasets with extensive country coverage and annual data for most of the multilateral trade system's history, with the most recent datasets being the least fragmented (rows "Year" and "Interval").

However, the earliest studies fail to deal with the drawbacks associated with how the dependent variable is defined. This gives rise to potential estimation bias due to misspecification of the dependent variable (row "Trade: Specification"). Rose

[5]Trade at the intensive margin studies the WTO effect on country-pairs that already have a trade relationship. However, WTO membership may also lead to the creation of trade relationships that countries would otherwise not have had, which is referred to as trade at the extensive margin.

Table 2 Developments in the WTO effect literature

Variable	(1) Rose (2004)	(2) Subramanian and Wei (2007)	(3) Tomz et al. (2007)	(4) Felbermayr and Kohler (2006)	(5) Liu (2009)	(6) Herz and Wagner (2011)
Year	1948–1999	1950–2000	1948–1999	1965–2004	1948–2003	1953–2006
Interval	Annual	5 years	Annual	5 years	Annual	Annual
Trade specification	Arithmetic average	Unidirectional imports	Arithmetic average, exports and imports	Unidirectional exports	Unidirectional imports	Unidirectional imports
Zero trade	Ignored	Ignored	Ignored	Included	Included	Included
WTO variable	Only de jure	Only de jure	De jure and de facto	Only de jure	Only de jure	De jure and de facto
EIA variable	Limited	Limited	Limited	Limited	Limited	Limited
Method	OLS	OLS	OLS	Tobit	Poisson	Poisson

Shaded cells indicate problematic issues that may yield biased parameter estimates

(2004) defines average trade as an arithmetic average or log of sums. Nevertheless, Baldwin and Taglioni (2006) point out that the correct specification must be a *geometric* average because Rose's estimated model is derived from a multiplicative specification.[6] Moreover, Anderson and Wincoop (2003)'s and Feenstra (2004)'s theoretical derivations of the gravity equation show that the dependent variable must be unilateral trade (imports *or* exports) and not aggregate trade (i.e., imports *plus* exports, as in Rose's study). Compared to aggregate trade flows, the advantage of unidirectional trade flows is that the researcher can identify potential asymmetries between importers and exporters.

The first few studies also fail to incorporate zero trade "flows", which is also likely to yield biased results (row "Zero trade"). Moreover, Tomz et al. (2007)'s improvement of the WTO participation variable has only been incorporated by Herz and Wagner (2011). As such, all other studies suffer from biased WTO effect estimates because the independent variable of interest is misspecified (row "WTO variable").

This study incorporates the piecemeal improvements that are scattered over various studies in the literature. Moreover, it addresses the two remaining issues that have so far been ignored. As indicated in the lower rows of Table 2, the proposed improvements deal with (1) the specification of the EIA dummy variable and (2) the econometric method used to estimate the gravity equation with zero trade "flows". The reasons why these issues warrant closer inspection and how they improve upon the existing literature are discussed in Sects. 3.1 and 3.2, respectively.

3 Methodology

3.1 Regionalism

The WTO effect literature acknowledges that regionalism, i.e. economic integration governed by means of EIAs, may have trade-creating effects. Gravity equation specifications usually allow for EIAs to be included as a binary variable that is 1 if trade between two countries is regulated by an EIA and 0 otherwise. So far, the literature's evidence on regionalism's effect on trade has been mixed. It can be argue that these estimates are biased for at least two reasons.

First, Rose (2004) only uses ten EIAs to construct his EIA dummy variable, which is limited to ANZCERTA, ASEAN, CACM, CARICOM, EEC/EC/EU, Mercosur, NAFTA, PATCRA, SPARTECA and US-Israel FTA. These EIAs do not qualify as an appropriate representation of all the EIAs that actually exist. For an overview of the 200+ EIAs that could and should have been included, see Table 3. Using such a small sample of EIAs to represent the entire population is certain to

[6]The former is calculated as $ln[\frac{1}{4}(export_{ij} + export_{ji} + import_{ij} + import_{ji})]$, the latter as $ln[(export_{ij} \times export_{ji} \times import_{ij} \times import_{ji})^{\frac{1}{4}}]$.

Table 3 EIAs by year of enforcement

Year	EIA
<1950	Belgium-Netherlands-Luxembourg Customs Union (Benelux) (1957), Southern African Customs Union (SACU).
1958	European Community (EC).
1960	European Free Trade Association (EFTA).
1961	Central American Economic Integration Agreement (CAEIA) (1966)
1966	Central American Common Market (CACM) (1970)
1968	Association of Southeast Asian Nations (ASEAN) Free Trade Agreement (AFTA).
1971	EC-Malta (2004), EC-Overseas Countries and Territories (EC-OCT).
1973	Caribbean Community (CARICOM), Costa Rica-Panama, EC-Cyprus (2004), EC-Egypt, EC-Iceland, EC-Norway, EC-Switzerland-Liechtenstein, Honduras-Panama.
1976	Asia Pacific Trade Agreement (APTA), Australia-Papua New Guinea (PATCRA), EC-Algeria.
1977	EC-Syria.
1981	Gulf Cooperation Council (GCC), India-Maldives, Latin American Integration Association (LAIA), South Pacific Regional Trade and Economic Cooperation Agreement (SPARTECA/PACER).
1982	Australia-New Zealand (ANZCERTA), Organisation of Eastern Caribbean States (OECS).
1984	China-India.
1985	Economic Community of Central African States (ECCAS), Economic Cooperation Organisation (ECO), Israel-US.
1988	Andean Community (Cartanega).
1989	Arab Maghreb Union (AMU), Canada-US Free Trade Agreement (CUSFTA) (1994).
1990	Central American Common Market (CACM) (revival)
1991	EC-Andorra, India-Nepal, Laos-Thailand, Mercado Comun del Sur (MERCOSUR).
1992	Armenia-Russia, Belarus-Ukraine, Czech Republic-Slovak Republic (2004), EC-Czech Republic (2004), EC-Hungary (2004), EC-Poland (2004), EC-Slovak Republic (2004), EFTA-Czech Republic (2004), EFTA-Slovak Republic (2004), EFTA-Turkey, Faroe Islands-Norway, Kyrgyz Republic-Russia.
1993	Armenia-Moldova, Bolivia-Chile, CARICOM-Venezuela, Chile-Venezuela, EC-Bulgaria (2007), EC-Romania (2007), Economic Community of West African States (ECOWAS), EFTA-Bulgaria (2007), EFTA-Hungary (2004), EFTA-Israel, EFTA-Poland (2004), EFTA-Romania (2007), Russia-Ukraine.
1994	Armenia-Kyrgyz Republic, Armenia-Ukraine, Baltic Free Trade Area (BAFTA) (2004), Bolivia-Mexico, Central European Free Trade Agreement (CEFTA), Common Market for Eastern and Southern Africa (COMESA), Commonwealth of Independent States (CIS), Costa Rica-Mexico, European Economic Area (EEA), Faroe Islands-Switzerland, Georgia-Russia, Kazakhstan-Ukraine, Melanesian Spearhead Group (MSG), Moldova-Romania (2007), North American Free Trade Agreement (NAFTA), Turkmenistan-Ukraine, Ukraine-Uzbekistan, West African Economic and Monetary Union (WAEMU).

(continued)

Table 3 (continued)

Year	EIA
1995	Armenia-Cyprus (2004), Armenia-Georgia, Armenia-Iran, Armenia-Turkmenistan, Association of Caribbean States (ACS), Azerbaijan-Ukraine, CARICOM-Colombia, Colombia-Mexico-Venezuela, EC-Estonia (2004), EC-Latvia (2004), EC-Lithuania (2004), EFTA-Slovenia (2004), Estonia-Ukraine (2004), Georgia-Ukraine, Kazakhstan-Kyrgyz Republic, Kyrgyz Republic-Moldova, Kyrgyz Republic-Ukraine, South Asian Preferential Trade Agreement (SAPTA) (2006).
1996	Azerbaijan-Georgia, Canada-Chile, Canada-Israel, Croatia-Macedonia, Czech Republic-Estonia (2004), Czech Republic-Israel (2004), Czech Republic-Latvia (2004), Czech Republic-Lithuania (2004), EC-Turkey, EFTA-Estonia (2004), EFTA-Latvia (2004), EFTA-Lithuania (2004), Estonia-Slovak Republic (2004), Estonia-Slovenia (2004), Georgia-Turkmenistan, Israel-Turkey, Kyrgyz Republic-Uzbekistan, Latvia-Slovak Republic (2004), Latvia-Slovenia (2004), Lithuania-Poland (2004), Lithuania-Slovak Republic (2004), Lithuania-Slovenia (2004), Macedonia-Slovenia (2004).
1997	Armenia-Canada, Croatia-Slovenia (2004), Czech Republic-Turkey (2004), EC-Faroe Islands, EC-PLO, EC-Russia, EC-Slovenia (2004), Estonia-Faroe Islands (2004), Estonia-Turkey (2004), Eurasian Economic Community (EAEC), Georgia-Kazakhstan, Hungary-Israel (2004), Hungary-Turkey (2004), Israel-Poland (2004), Israel-Slovak Republic (2004), Lithuania-Turkey (2004), MERCOSUR-Bolivia, Mexico-Nicaragua, Romania-Turkey (2007), Slovak Republic-Turkey (2004).
1999	Armenia-Kazakhstan, Bulgaria-Macedonia (2007), Chile-Costa Rica, Chile-El Salvador, Chile-Guatemala, Chile-Honduras, EFTA-Morocco, EFTA-PLO, Macedonia-Turkey, Poland-Turkey (2004).
2001	Albania-Macedonia, Bosnia & Herzegovina-Macedonia, Bosnia & Herzegovina-Slovenia (2004), Bulgaria-Estonia (2004), Bulgaria-Israel, Bulgaria-Lithuania (2004), Canada-Costa Rica, Central America-Mexico, EC-Macedonia, EFTA-Macedonia, EFTA-Mexico, Israel-Romania (2007), Macedonia-Ukraine, Mexico-Northern Triangle, Northern Triangle, Tajikistan-Ukraine, Vietnam-US.
2002	Albania-Bosnia & Herzegovina, Albania-Bulgaria (2007), Albania-Croatia, Armenia-Estonia (2004), Bosnia & Herzegovina-Romania (2007), Bosnia & Herzegovina-Serbia & Montenegro, Bosnia & Herzegovina-Turkey, Bulgaria-Latvia (2004), CARICOM-Dominican Republic, Croatia-Turkey, EC-Croatia, EC-Jordan, EC-San Marino, EFTA-Croatia, EFTA-Jordan, El Salvador-Panama, Guatemala-Panama, Japan-Singapore, Pakistan-Sri Lanka.
2003	Afghanistan-India, Albania-Moldova, Albania-Romania (2007), Albania-Serbia & Montenegro, ASEAN-China, Australia-Singapore, Bosnia & Herzegovina-Bulgaria (2007), Bosnia & Herzegovina-Moldova, Bulgaria-Moldova (2007), Bulgaria-Serbia & Montenegro (2007), Chile-South Korea, Chile-US, China-Hong Kong, China-Macao, Croatia-Lithuania (2004), Croatia-Moldova, Croatia-Serbia & Montenegro, Dominican Republic-Panama, EC-Chile, EC-Lebanon, EFTA-Singapore, Mexico-Uruguay, Moldova-Serbia & Montenegro, Moldova-Ukraine, Pacific Island Countries Trade Agreement (PICTA), Romania-Serbia & Montenegro (2007), Singapore-US.
2004	Australia-Thailand, Australia-US, Bosnia & Herzegovina-Croatia, CARICOM-Costa Rica, Common Economic Zone (CEZ), EC-Fiji-Papua New Guinea, EC-Serbia & Montenegro, EFTA-Chile, Japan-Mexico, Jordan-Singapore, Macedonia-Moldova, Macedonia-Romania (2007), Morocco-Turkey, Morocco-US, Panama-Taiwan, Syria-Turkey, Tunisia-Turkey.

Table 3 (continued)

Year	EIA
2005	Bahrain-US, Bhutan-India, CARICOM-Cuba, Chile-China, EFTA-Tunisia, Egypt-Turkey, Faroe Islands-Iceland, India-Singapore, Japan-Malaysia, MERCOSUR-Colombia-Ecuador, MERCOSUR-Peru, MERCOSUR-SACU, PLO-Turkey, New Zealand-Thailand, Singapore-South Korea.
2006	ASEAN-South Korea, Bangladesh-India, Central America-Dominican Republic-US, Chile-India, Chile-Peru, EC-Albania, EFTA-South Korea, Guatemala-Taiwan, MERCOSUR-Mexico, Panama-Singapore, South Asian Free Trade Agreement (SAFTA), Trans-Pacific Strategic Economic Partnership (TPSEP).
2007	Chile-Japan, China-Pakistan, EFTA-Egypt, EFTA-Lebanon, Japan-Thailand, Mauritius-Pakistan.

Years of expiration or transition to another EIA are in brackets. *Sources*: McGill (2009), Tuck (2009), WorldTradeLaw.net (2009), World Bank (2011c) and WTO (2011a)

introduce bias in the EIA point estimates. This problem is persistent in the WTO effect literature, especially because a number of papers re-use Rose's EIA variable (see Table 6 on p. 477).

Second, the literature has almost exclusively relied on the WTO website as its only source of information about EIAs. Although WTO members have an obligation to notify all existing and newly ratified EIAs to the WTO Secretariat, in practice it may take several months, or even years, before the Secretariat receives the required information and publishes it online. The problem is that the researcher who only looks at this source of information does not get the most up to date information on all EIAs that are or have been active on the world trade scene.

A related complication is that the WTO website only has data on the most recent version of an EIA. For example, the EC and Middle Eastern countries had established several bilateral trade relationships by the 1970s. These treaties underwent some changes after the turn of the millennium and their renewed content was notified to the WTO Secretariat. Since new treaties typically replace older versions, only data regarding the latest version, including the year of ratification, are reported online. In the case of EC-Middle Eastern trade agreements, the naive researcher will wrongly infer from the WTO website that these agreements (signed after 2000) are first of their kind and that their predecessors simply did not exist.

These forms of selection bias are solved by exploiting several databases that provide information on EIAs. This has the advantage that data on EIAs can be included even if they are no longer listed on the WTO's website because they have been replaced by newer versions or have ceased to exist. There are also EIAs that have not been notified to the WTO, but which have been enforced by their respective governments nonetheless.[7]

[7]There are no provisions in international law that require governments to notify their EIAs to the WTO before they can be enforced.

In all cases, the EIAs concerned are comprehensive agreements that cover trade in (almost) all goods, and not just a limited number of sectors. In addition, every effort was made to verify each EIA's date of ratification, the participating countries and when they joined or left the EIA. This involved several additional sources such as the text of the original treaties, government records documenting the parliamentary discussions and final ratification of the treaties, newspapers and professional magazines.

By including EIAs that were ignored by Rose (2004) and others, more than 40 % of the EIA dummy variables in the present dataset come from sources other than the WTO website. The number of country-pairs registered as having an EIA increases by more than 220 % compared to Rose's original dataset.

The next section discusses developments in the methodology used in estimating the WTO effect and argues why a (ZI)NB model is more appropriate than OLS and Poisson (quasi-)MLE when zero trade "flows" are to be included in the analysis.

3.2 Econometric Estimation

The standard tool to model bilateral trade flows in the international trade literature is the gravity equation, which borrows its name from Newton's law of universal gravitation (for a review, see van Bergeijk and Brakman, 2010). A simple form of the gravity equation is:

$$T_{ijt} = \frac{Y_{it}^{\beta_1} Y_{jt}^{\beta_2}}{D_{ij}^{\beta_3}}, \tag{1}$$

where T_{ijt} is bilateral aggregate import by importer i from exporter j in year t, Y is a proxy for the respective country's economic size (GDP) and D_{ij} is the physical distance between i and j. The gravity equation is usually rewritten in a log-linear form to estimate the β's:

$$ln(T_{ijt}) = \beta_0 + \beta_1 ln(Y_{it}) + \beta_2 ln(Y_{jt}) - \beta_3 ln(D_{ij}) + \beta_4 ln(z_{ijt}) + \epsilon_{ijt}, \tag{2}$$

where β_0 is a constant, z_{ijt} represents a host of variables that control for factors such as countries' physical size, the number of landlocked and island states in a dyad, whether countries share a common border or language and their colonial history, and ϵ_{ijt} the error term. A shortcoming of this specification is that it suffers from omitted variable bias, as it fails to account for unobserved price indexes. Anderson and Wincoop (2003) formalise the so-called "multilateral resistance term" (MRT), which takes into account that trade between two trading countries is affected by their bilateral trade barrier relative to their average trade barriers *vis-à-vis* all their trade partners.

Feenstra (2004, Chap. 5) shows that a convenient way to include the multilateral resistance term in regression estimates is by including country fixed effects. This

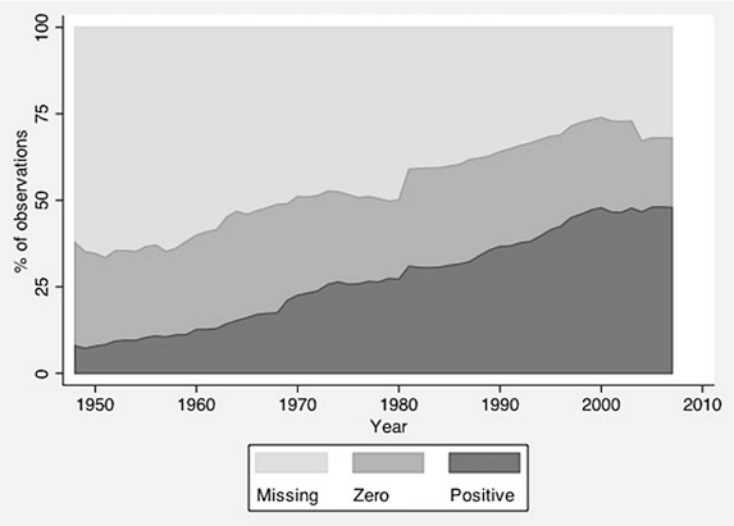

Fig. 1 Distribution of missing, zero and positive values per year

yields a log-normal fixed effects specification that is commonly accepted in the literature:

$$ln(T_{ijt}) = \beta_0 + \beta_1 ln(Y_{it}) + \beta_2 ln(Y_{jt}) + \beta_3 z_{ijt} - \beta_4 ln(D_{ij}) + \gamma_i F_i + \delta_j F_j + \zeta_t F_t + \epsilon_{ijt}, \tag{3}$$

where F_i represents fixed effects for the importing country and F_j represents fixed effects for the exporting country. Time fixed effects, F_t, are included to correct for common trends and shocks.[8]

Missing and zero values account for a large part of a typical dataset on international trade, including the one used in the present study. The balanced panel covers 181 countries and 60 years. Overall, almost half of the dependent variable consists of missing values. An additional 25 % of the observations involve zero trade, which reflects the fact that either there is no actual trade between countries, or that the level of trade is so small—with small depending on pre-defined levels determined by the statistical agency collecting the data—that they are rounded to zero. Missing values, on the other hand, are points in the dataset that remain unobserved. There may be actual ("positive") trade, or there may really be no trade at all ("zero"). In either case, the information has gone unreported. Figure 1 shows

[8] Baldwin and Taglioni (2006) and Baier and Bergstrand (2007) argue that these "traditional," time-invariant MRT should actually be time-varying to correct for endogeneity bias, and that it can be included in linear models by estimating importer-time (it), exporter-time (jt) and country-pair (ij) effects. However, the large number of dummies leads to an incidental parameter problem. Although this can be addressed in linear models by de-meaning the data, in this case cancelling the dyad dummies, there is as of yet no straightforward solution for non-linear models.

the distribution of missing, zero and positive values per year. The increasing share of positive trade flows suggests that countries have become, on average, engaged in a larger number of trading relationships over time. The share of the zeros has nevertheless remained fairly constant. An important drawback of estimating the gravity equation in its log-linear form is that it ignores these zero trade "flows" because the natural logarithm of zero is not defined. Five ways are discussed in which this matter can be dealt with.

The first approach is to simply drop all observations with zero trade "flows" (as in Rose, 2004; Subramanian and Wei, 2007; Tomz et al., 2007). However, this means that a substantial amount of valuable data is ignored and renders analyses to capture the effects of WTO and EIA membership at the extensive margin impossible. A second approach is to increase all zeros by a small constant, yet this approach tends to introduce even more bias in the regression estimates when the zero values are not randomly distributed, which is often the case (Bosker, 2008).

A third approach is to use a Tobit procedure (as in Felbermayr and Kohler, 2006), which can be used to isolate membership in institutional arrangements at both the intensive and extensive margin of trade. In contrast, Santos Silva and Tenreyoro (2006) argue that this method's stringent assumptions of homoscedasticity and normality are easily violated, which is likely to yield even worse results than OLS.

Instead, Santos Silva and Tenreyoro (2006) propose using Poisson MLE as a fourth alternative to incorporate zero trade "flows". Although Poisson models are commonly used with count data, Wooldridge (2002) shows that they can also be easily used with non-negative continuous variables. Liu (2009) and Herz and Wagner (2011) estimate the WTO effect using Poisson quasi maximum likelihood estimation (QMLE), which is favourable to the above-mentioned OLS and Tobit procedures because (1) it solves the zero trade problem by using a multiplicative, non-linear specification of the gravity equation (i.e., there is no need to calculate the natural logarithm of the dependent trade variable) and (2) it is not subject to strict assumptions of homoscedasticity and normality. For example, Siliverstovs and Schumacher (2009) show that a non-linear multiplicative Poisson specification of the gravity model performs better than traditional OLS estimates of a log-linear gravity equation.

Nevertheless, a drawback of standard Poisson models is the assumption of equidispersion, i.e., the conditional mean and conditional variance are equal. This assumption is easily violated in applied work. In the event of overdispersion, Poisson MLE yields a biased model fit and underpredicts the standard errors of the estimates (Hilbe, 2007, Chap. 4). Although Cameron and Trivedi (2009, p. 316) show that the Poisson quasi-MLE is capable of providing consistent estimates even in the case of overdispersion (provided that the conditional mean function is correctly specified,) the more important question is whether excess zeros are modelled correctly. Goodness-of-fit statistics may indicate that other models are more appropriate. In this regard, STATA's `countfit` package is a useful tool. See Table 8.

Burger et al. (2009) extend the empirical gravity model proposed by Santos Silva and Tenreyoro (2006) and show that the problem of overdispersion can also be addressed by so-called modified Poisson models, specifically (ZI)NB MLE. In addition to the above-mentioned benefits pertaining to Poisson models, (ZI)NB MLE does not require equidispersion and may be better in modelling the data than a standard Poisson distribution. Until now, this alternative approach to dealing with zero trade "flows" has remained unexplored in the WTO effect literature. Hilbe (2007) provides the theoretical framework for (ZI)NB MLE, while Burger et al. (2009) apply ZINB on traditional gravity equations and Brakman et al. (2010) and Swart and Marrewijk (2009) do so for gravity equations involving cross-border mergers and acquisitions.

NB MLE is used to model overdispersed data, but it may predict fewer zeros for a given mean than the actual number of observed zeros. This result arises when there is an excessive number of zeros in the dependent variable. In this case, the negative binomial's zero-inflated variant may be more appropriate.[9]

ZINB MLE is a two-part model with the density function:

$$g(T_{ijt}) = \begin{cases} f_1(0) + (1 - f_1(0)) f_2(0) & \text{if} T_{ijt} = 0, \\ (1 - f_1(0)) f_2(T_{ijt}) & \text{if} T_{ijt} \geq 1. \end{cases} \tag{4}$$

The first part, $f_1(\cdot)$, is a binary process that estimates the gravity equation—specified in (5) below—with a binary regression model such as logit. The second part, f_2, is a count process that estimates (5) with either Poisson or negative binomial regression. The trade flows are split into "passive" and "active" groups. When the outcome of the logit is 0, with probability $f_1(0)$, the trade flows are zero ($T_{ijt} = 0$) and are "passive". If the outcome is 1, with probability $f_1(1)$, these "active" trade flows take on count values $0, 1, 2, \ldots$ from the second part, $f_2(\cdot)$ (see also Cameron and Trivedi, 2009, p. 681). Note that the ZINB procedure does not require unique sets of independent variables for each part of the model. This is a useful feature in the present context of international trade, because it is difficult to imagine two distinct sets of independent variables, both of which are trade-related, but with the first set specifically explaining "zero" trade and the second set focusing on "non-zero" trade. For example, trade can be "zero" if distance is sufficiently great, while "non-zero" trade is fostered by geographic proximity at the same time.

Another useful feature of modified Poisson models is that they include an overdispersion parameter, α. If $\alpha = 0$, the conditional mean is equal to the conditional variance and a standard Poisson model is the most appropriate. If $\alpha > 0$, there is evidence of overdispersion in the data and the negative binomial is preferred to Poisson. If $\alpha > 0$, the next step is to determine whether the negative binomial or its zero-inflated variant must be used. This is determined by using a Vuong (1989) test. Negative values favour the negative binomial, while positive values favour its zero-

[9]Zero-truncated models are not considered because they only allow for positive trade flows, which does not help to address the zero trade flow problem.

inflated variant. As will be shown in Sect. 4, the α test statistics repeatedly indicate a preference for NB to Poisson. These test criteria suggest that it is more appropriate to estimate the WTO effect using (ZI)NB MLE than OLS or standard Poisson models previously employed in the literature. Moreover, positive Vuong z-statistics are often found, implying that ZINB MLE is preferable to the non-inflated variant. If the Vuong z-statistic is negative, the NB's results are automatically calculated and reported.

The gravity model that is estimated with this (ZI)NB MLE procedure is:

$$T_{ijt} = \beta_0 \times Y_{it}^{\beta_1} \times Y_{jt}^{\beta_2} \times P_{it}^{\beta_3} \times P_{jt}^{\beta_4} \times D_{ij}^{\beta_5} \times Both_{ijt}^{\beta_6} \times One_{ijt}^{\beta_7} \times GSP_{ijt}^{\beta 8} \times EIA_{ijt}^{\beta 9} \times z_{ijt}^{\beta 10} \times F_i^{\gamma_i} \times F_j^{\delta_j} \times F_t^{\zeta_t} \times \epsilon_{ijt}, \quad (5)$$

where P is population and $Both_{ijt}$ is a dummy variable that is 1 (0) if both (none of the) countries in the country-pair are WTO participants. One_{ijt} is a dummy that is 1 if only one country in the country-pair is a WTO participant and 0 otherwise. GSP_{ijt} is a binary variable that is 1 if the GSP applied to a dyad's trade relationship and 0 if not. EIA_{ijt} is a binary variable that is 1 if both countries in a dyad belong to the same EIA and 0 otherwise. z_{ijt} is a vector of additional controls, described below. The multilateral resistance term is incorporated by using dummies that represent individual country effects (F_i and F_j), while common shocks and trends are captured by the year dummies (F_t). This estimation strategy is recommended to obtain reliable estimates with modified Poisson models (Allison and Waterman, 2002).

3.3 Data

The panel dataset covers 181 countries and contains observations for the period 1948 (the GATT's founding year) to 2007. Table 4 lists the countries included in the dataset. The panel is arranged by country-pair and year, regardless of missing or zero values. Each country-pair is represented twice, once as *ij* and once as *ji*. This is done because bilateral *imports* are used as the dependent variable, which avoids Baldwin and Taglioni (2006)'s critique on the specification of the dependent variable (see p. 466). There is a maximum of $181 \times 180 \times 60 = 1{,}954{,}800$ observations and approximately 98 % of world trade is covered by the present dataset.

Bilateral trade data (imports c.i.f. and exports f.o.b. in US$ millions) are from IMF (1995, 2008). The dependent variable of choice is bilateral imports. In case of missing values, the country's trade partner's bilateral exports are used as a proxy of that country's bilateral imports. Following Liu (2009), a 10 % c.i.f. rate is assumed when exports are used to replace missing imports. Baldwin and Taglioni (2006)

Table 4 Countries in the dataset

Afghanistan*, Albania, Algeria, American Samoa, Angola*, Antigua & Barbuda, Argentina, Aruba, Australia, Austria, Bahamas, Bahrain, Bangladesh*, Barbados, Belgium, Belize, Benin*, Bermuda, Bhutan*, Bolivia, Botswana*, Brazil, Brunei, Bulgaria, Burkina Faso*, Burundi*, Côte d'Ivoire, Cambodia*, Cameroon, Canada, Cape Verde*, Cayman Islands, Central African Republic*, Chad*, Chile, China, Colombia, Comoros*, Costa Rica, Cuba, Cyprus, D.R. Congo*, Denmark, Djibouti*, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea*, Eritrea*, Ethiopia*, Faeroe Islands, Falkland Islands, Fiji, Finland, France, French Polynesia, Gabon, Gambia*, Germany, Ghana, Gibraltar, Greece, Greenland, Grenada, Guam, Guatemala, Guinea*, Guinea-Bissau*, Guyana, Haiti*, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Kiribati*, Kuwait, Laos*, Lebanon, Lesotho*, Liberia*, Libya, Luxembourg, Macao, Madagascar*, Malawi*, Malaysia, Maldives*, Mali*, Malta, Marshall Islands, Mauritania*, Mauritius, Mexico, Micronesia, Mongolia, Montserrat, Morocco, Mozambique*, Myanmar*, Namibia, Nauru, Nepal*, Netherlands, Netherlands Antilles, New Caledonia, New Zealand, Nicaragua, Niger*, Nigeria, North Korea, Norway, Oman, Pakistan, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Republic of Congo, Romania, Rwanda*, St. Helena, St. Kitts & Nevis, St. Lucia, St. Pierre-Miquelon, St. Vincent & Grenadines, Samoa*, São Tomé & Príncipe*, Saudi Arabia, Senegal*, Seychelles, Sierra Leone*, Singapore, Solomon Islands*, Somalia*, South Africa, South Korea, Spain, Sri Lanka, Sudan*, Suriname, Swaziland, Sweden, Switzerland, Syria, Tanzania*, Thailand, Togo*, Tonga, Trinidad & Tobago, Tunisia, Turkey, Tuvalu*, Uganda*, United Arab Emirates, United Kingdom, United States, Uruguay, Vanuatu*, Venezuela, Vietnam, Yemen, Zambia*, Zimbabwe.

* indicates LDCs

argue that deflating trade data with a common price index may bias the regression estimates, but that time fixed effects may sufficiently address this issue. Since time fixed effects are included in the regression estimates, there is no problem with deflating trade by the US Consumer Price Index (All Consumer Goods, $1983-4=100$) obtained from Bureau of Labor Statistics (2008).

Data on GDP (in 1990 international dollars) are from Maddison (2007). Additional data are from World Bank (2011b) using the GDP in 2000 international dollars series, which was reconverted to be consistent with Maddison's data. Data on population were also obtained from Maddison (2007). Population data for 1948–1949 are from World Bank (1951). Several variables are from CEPII (2008): simple geodesic distance (in kilometres), country size (in square kilometres), whether countries share a common major/official language, a border, the number of countries in the dyad that are islands or landlocked, whether the countries in the dyad used to be one country, and details on their colonial history.

Details on the EIA variable are provided in Sect. 3.1. Data on countries' WTO participation status are from Tomz et al. (2007). A number of updates were necessary, mostly for a number of countries that became formal WTO members in the period 2000–2007. Updates were obtained from the WTO website.

GSP was initiated in 1969 to allow developed nations to unilaterally grant preferential market access to exporters from developing countries. Rose's GSP dummy is only based on GSP decisions published in 1974, 1979 and 1984. In addition to using data from these years, data have also been obtained data for several others (UNCTAD, 1974, 1975, 1979, 1981, 1985, 2005, 2006, 2008). The published preferences are assumed to have remained valid during years for which data could not be obtained. Thus, data for 1973 also apply to preceding years, 1974 also applies to 1975–1976, 1977 also applies to 1978, 1979 also applies to 1980–1983, 1984 also applies to 1985–1994, 2004 also applies to 1995–2003 and 2005 also applies to 2006. 1995 is considered a "break" year in the long gap between data for 1984 and 2004 because GSP schemes are likely to have been altered during the Uruguay Round.

Descriptive statistics of the dataset are provided in Table 5, while a summary of the data sources used in previous studies and in the present dataset is provided in Table 6.

Table 5 Descriptive statistics

Variable	Observations	Mean	Std. deviation	Minimum	Maximum
Imports	1,054,520	119.06	1589.26	0.00	214,440
Both in WTO	1,954,800	0.33	0.47	0.00	1.00
One in WTO	1,954,800	0.43	0.50	0.00	1.00
GSP	1,954,800	0.12	0.32	0.00	1.00
EIA	1,954,800	0.04	0.19	0.00	1.00
ln GDP importer	1,494,000	9.52	2.19	3.53	16.04
ln GDP exporter	1,494,000	9.52	2.19	3.53	16.04
ln population importer	1,483,380	1.66	1.89	−3.69	7.18
ln population exporter	1,483,380	1.66	1.89	−3.69	7.18
ln land area importer	1,911,600	10.99	3.04	1.95	16.12
ln land area exporter	1,911,600	10.99	3.04	1.95	16.12
ln distance	1,869,120	8.84	0.75	2.35	9.90
Number landlocked	1,869,120	0.29	0.50	0.00	2.00
Number of islands	1,869,120	0.68	0.67	0.00	2.00
Common land border	1,954,800	0.01	0.11	0.00	1.00
Common language	1,869,120	0.21	0.41	0.00	1.00
Ever colony	1,954,800	0.01	0.10	0.00	1.00
Common coloniser	1,954,800	0.14	0.34	0.00	1.00
Currently colonised	1,954,800	0.00	0.03	0.00	1.00
Common country	1,954,800	0.01	0.09	0.00	1.00

Table 6 Data sources in the WTO effect literature

Variable	Rose (2004)	Subramanian and Wei (2007)	Tomz et al. (2007)	Felbermayr and Kohler (2006)	Liu (2009)	Herz and Wagner (2011)	This study
Year	1948–1999	1950–2000	1948–1999	1965–2004	1948–2003	1953–2006	1948–2007
Interval	Annual	5 years	Annual	5 years	Annual	Annual	Annual
Trade	IMF DOTS	IMF DOTS	Rose (2004)	IMF DOTS	IMF DOTS, WED, WTF	IMF DOTS	IMF DOTS
CPI	US CPI all urban consumers, freelunch.com	Rose (2004)	Rose (2004)	N/A	US CPI	IMF IFS, WBDI	US CPI all urban consumers, Bureau of Labor Statistics (2008)
Pop. & GDP	IMF IFS, PWT, WBDI	Rose (2004)	Rose (2004)	IMF IFS	IMF IFS, Maddison (2007), PWT, WBDI	IMF IFS, Maddison (2007), PWT, WBDI	Maddison (2007), World Bank (1951, 2011b)
WTO variable	WTO website	Rose (2004)	WTO documentation	N/A	WTO website	Tomz et al. (2007)	Tomz et al. (2007) & WTO website
EIA variable	WTO website	Rose (2004)	Rose (2004)	Baier and Bergstrand (2007)	WTO website	WTO website	McGill (2009), Tuck (2009), WorldTradeLaw.net (2009), World Bank (2011c), WTO (2011a)
GSP variable	UNCTAD 1974, 1979, 1984	Rose (2004)	Rose (2004)	N/A	UNCTAD 1973–1975, 1977, 1979, 1984, 2001	UNCTAD 1973–1986, 2001, 2005	UNCTAD 1973–1974, 1977, 1979, 1984, 2005–2006, 2008
Control variables	CIA WFB	Rose (2004)	Rose (2004)	N/A	CIA WFB	PWT	CEPII (2008)

The longest time-series were used. Gaps were filled using trends obtained in alternative data sources: CIA World Factbook (WFB), IMF Direction of Trade Statistics (DOTS), IMF International Financial Statistics (IFS), Penn World Tables (PWT), World Bank Development Indicators (WBDI), World Export Dataset (WED) and World Trade Flows (WTF)

4 Results

First, Rose (2004)'s calculations are re-examined in Table 7, with his original results presented in column 1. The results reported in the top panel only include year effects, while those in the lower panel also include country effects.[10]

The results in columns 2 to 9 provide an updated version of Rose's estimates. Our underlying dataset is different in terms (1) the data sources, (2) coverage, including more countries and years, (3) imports rather than the arithmetic average of trade as dependent variable (see Baldwin and Taglioni, 2006), (4) relying on Tomz et al. (2007)'s unbiased WTO participation data and (5) this study's EIA variable with extensive coverage instead of Rose's biased indicator. Columns 2 to 8 show statistical output based on new data, while keeping the sample, WTO and/or EIA data the same as Rose's. The results presented in column 9 are obtained from using data that overcome all the drawbacks discussed above.

Remarkably, following Rose's approach does not yield his mysterious "non-effect". According to the estimates presented in column 2, the WTO has a positive effect on the trade flows between its members. Why do the results differ from the benchmarks? Note that there are minor differences in the (versions of) data sources employed in the construction of both datasets (see Table 6). Also notice that the number of observations is substantially larger. The present dataset is arranged such that one country features twice a year in each country-pair, once as importer and once as exporter, which yields a balanced panel. Rose's data, on the other hand, constitute an unbalanced panel. This means that a country is sometimes featured as an importer, sometimes as an exporter, without any explanation as to why this is the case. This difference in how the data are organised largely explains why the number of observations differ between columns 1 and 2.

In changing the sample, dependent variable, WTO and EIA indicators one by one, any one change can yield a different outcome. Changing the sample or dependent variable does not alter the outcomes very specifically. Then again, there are cases when switching to an improved variable yields a more pronounced and different outcome, e.g. positive WTO effects in column 5 or less pronounced EIA effects in column 6 and 9.

[10]Rose's model specification is slightly different from (5). In particular, he uses the natural logarithm of the product of the importer and exporter's real GDP. In the remainder of this study, these variables are included separately.

Table 7 Rose (2004) revisited. OLS with country and/or time fixed effects

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fixed effects: year									
Both in WTO	−0.039	0.420***	0.371***	0.361***	0.482***	0.415***	0.220***	0.522***	0.488***
	(0.053)	(0.039)	(0.030)	(0.048)	(0.043)	(0.038)	(0.036)	(0.053)	(0.053)
One in WTO	−0.058	−0.137***	−0.164***	−0.200***	0.206***	−0.123***	−0.225***	0.179***	0.150**
	(0.049)	(0.024)	(0.024)	(0.027)	(0.042)	(0.024)	(0.026)	(0.051)	(0.051)
EIA	1.112***	1.294***	1.215***	1.400***	1.302***	1.009***	1.360***	1.353***	0.831***
	(0.108)	(0.089)	(0.070)	(0.084)	(0.089)	(0.044)	(0.071)	(0.071)	(0.040)
GSP	0.852***	1.091***	0.941***	0.749***	1.067***	1.108***	0.639***	0.619***	0.634***
	(0.032)	(0.025)	(0.022)	(0.028)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
ln distance	−1.110***	−1.129***	−1.140***	−0.940***	−1.133***	−1.080***	−0.927***	−0.932***	−0.891***
	(0.022)	(0.018)	(0.016)	(0.020)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
ln GDP	0.916***	1.121***	1.114***	1.013***	1.121***	1.118***	0.968***	0.966***	0.965***
	(0.010)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
R^2	0.649	0.672	0.660	0.624	0.672	0.674	0.587	0.588	0.588
Fixed effects: country & year									
Both in WTO	0.130**	0.149***	0.150***	0.144***	0.471***	0.147***	0.047	0.207***	0.156***
	(0.046)	(0.035)	(0.027)	(0.040)	(0.040)	(0.035)	(0.032)	(0.048)	(0.047)
One in WTO	0.034	−0.162***	−0.170***	−0.145***	0.116**	−0.153***	−0.175***	−0.042	−0.079
	(0.043)	(0.020)	(0.019)	(0.025)	(0.035)	(0.020)	(0.023)	(0.045)	(0.044)
GSP	0.519***	0.749***	0.598***	0.484***	0.757***	0.769***	0.349***	0.352***	0.359***
	(0.029)	(0.023)	(0.020)	(0.024)	(0.023)	(0.023)	(0.022)	(0.022)	(0.022)
EIA	1.332***	0.728***	0.616***	0.985***	0.736***	0.719***	0.880***	0.884***	0.630***
	(0.103)	(0.102)	(0.080)	(0.092)	(0.102)	(0.039)	(0.078)	(0.078)	(0.034)
ln distance	−1.306***	−1.300***	−1.287***	−1.099***	−1.305***	−1.254***	−1.064***	−1.066***	−1.025***
	(0.023)	(0.019)	(0.017)	(0.021)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)

(continued)

Table 7 (continued)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln GDP	0.219***	1.000***	0.939***	1.425***	0.996***	1.008***	1.428***	1.424***	1.432***
	(0.049)	(0.033)	(0.029)	(0.041)	(0.033)	(0.033)	(0.035)	(0.035)	(0.035)
R^2	0.715	0.745	0.730	0.704	0.745	0.746	0.663	0.664	0.664
Observations	234,597	418,008	551,280	341,343	418,008	418,008	436,775	436,775	436,775
Data	Rose	New	New	New	New	New	New	New	New
Sample	Rose	Rose	New	Rose	Rose	Rose	New	New	New
Trade	Rose	Rose	Rose	New	Rose	Rose	New	New	New
WTO	Rose	Rose	Rose	Rose	New	Rose	Rose	New	New
EIA	Rose	Rose	Rose	Rose	Rose	New	Rose	Rose	New

The regressand is either the natural logarithm of the arithmetic average of import plus export when Trade is marked "Rose", otherwise the natural logarithm of imports. Estimates marked ***/**/* are significant at the 1/5/10 percent level. Robust standard errors (clustered by country-pair) are in brackets. Estimates for intercepts, other controls and fixed effects are omitted to save space

Our findings in column 9 are noteworthy for two reasons. First, the estimates that only rely on year fixed effects turn Rose's "non-effect" upside-down with respect to both countries and one country in a country-pair participating in the WTO. However, there now is a large literature demonstrating that year fixed effects are certainly not sufficient to incorporate the multilateral resistance term (see Anderson and Wincoop, 2003; Feenstra, 2004; Baier and Bergstrand, 2007). In other words, ignoring country effects yields biased estimates. Second, when we turn to the estimates in column 9 that do include country effects, we see an outcome highly similar to that of Rose with respect to WTO membership, i.e. joint membership fosters trade, while membership of only one partner does not. However, the robustness of these results is questionable because the underlying model fails to account for the presence of another problem discussed earlier: the presence of zero trade "flows".

Having demonstrated the importance of using the appropriate measures of WTO and EIA participation, the question that remains is how the (updated and improved) data has to be modelled in the presence of zero trade "flows". Analysis of the improved trade variable reveals that it is overdispersed, given that the conditional mean and variance are not equal. Moreover, goodness-of-fit measures provide further evidence in favour of (ZI)NB MLE. Indeed, the average predicted probabilities for the Poisson model fare worse than those provided by (ZI)NB models. Table 8 displays the penalised log-likelihood-based statistics (AIC and BIC), where the BIC penalises model complexity (i.e., the number of estimated parameters) more severely than the AIC. Both indicate that NB is strongly preferred to estimating our gravity equation with a Poisson distribution. Moreover, all of the tables presenting (ZI)NB MLE output (Tables 9, 10 and 11) show positive and significant α's for all estimates, which is evidence of overdispersion and means that a negative binomial distribution is superior to a Poisson distribution. Additional goodness-of-fit tests indicate overdispersion, suggesting that a Poisson model is not suitable to estimate our model. If the Vuong z-statistic is negative, the first step (logit) in the ZINB MLE procedure is automatically skipped and only results from the second step (NB MLE) are reported. In these cases, NB MLE is preferable to the zero-inflated version. Overall, the goodness-of-fit statistics, the positive and significant values for α, as well as the favourable Vuong z-statistics show strong support for the (ZI)NB MLE procedure applied in this study.

Using the updated dataset and improved EIA variable, the WTO effect is estimated with (ZI)NB MLE.[11] The results are presented in Table 9. The results for the (ZI)NB analyses are subdivided into two columns, one for the "active" group

[11] MLE is performed using the DFP algorithm (Gould et al., 2006, Chap. 1).

Table 8 Model selection

Model	Statistic		Difference	Prefer	Over	Evidence
Poisson	BIC	7.77e+7				
	AIC	112.16				
Poisson versus NB	BIC	−6.92e+6	8.46e+7	NB	Poisson	Very strong
	AIC	4.79	107.38	NB	Poisson	
	LRX2	8.46e+7		NB	Poisson	$p = 0.00$

STATA output generated with the `countfit` package

(using negative binomial regression) and one for the "passive" group (using logit). A variable's coefficient in the "passive" column shows the extent to which that variable contributes to the chance of being in the passive group. For example, the EIA variable in Table 9 has a negative coefficient in the "passive" column, which means that if a pair of countries participates in an EIA, this reduces (increases) the probability that the pair belongs to the passive (active) group.

In Table 9, column 1, results are shown for the following variables of interest: both countries in a dyad participate in the WTO, only one country in a dyad participates in the WTO, GSP applies to the trade relationship, and both countries belong to the same EIA. In column 2, it is investigated whether asymmetries arise between importers and exporters in case either is the sole WTO participant in the country-pair. Differences between GSP beneficiaries and GSP benefactors are also accounted for. Finally, Tomz et al. (2007) are followed by disentangling formal WTO members from informal participants so as to investigate potential asymmetries between these two groups in column 3.

Interestingly, all three columns indicate that the WTO has net trade-promoting effects for country-pairs where both the importer and exporter are members. Positive WTO effects are also found when only one of the trade partners is a member. What is most striking about these findings is that trade between members increases more compared to trade between a member and non-member. Moreover, formal members gain more than informal participants. So, the results suggest that there is a relation between the experienced gains and the extent to which nations have made multilateral commitments.

Robustness checks are performed to ensure the consistency of these findings. (ZI)NB MLE is performed for countries that have been grouped by their development status. The results are presented in Table 10. In addition, the effects vary by WTO negotiation round, as shown in Table 11.

Table 9 (ZI)NB results, 1948–2007

	(1)		(2)		(3)	
Variable	Active	Passive	Active	Passive	Active	Passive
Both in WTO	0.643*** (0.05)	−0.40	0.581*** (0.05)	−0.40		
– Formal & formal					0.759*** (0.07)	−0.24
– Formal & informal					0.757*** (0.06)	−0.12
– Informal & informal					0.607*** (0.05)	−0.04
One in WTO	0.342*** (0.05)	0.03				
– Importer			0.225*** (0.05)	0.04		
– Exporter			0.370*** (0.05)	−0.02		
– Formal & outsider					0.493*** (0.06)	0.03
– Informal & outsider					−0.209*** (0.05)	0.10
GSP	0.229*** (0.02)	−0.36				
– Beneficiary			0.283*** (0.03)	−0.20	0.268*** (0.03)	−0.20
– Benefactor			0.177*** (0.04)	−0.17	0.176*** (0.04)	−0.17
EIA	0.216*** (0.04)	−0.08	0.218*** (0.04)	−0.08	0.205*** (0.04)	−0.08
ln distance	−1.108*** (0.02)	−2.12	−1.109*** (0.02)	−2.12	−1.109*** (0.02)	−2.12
ln GDP						
– Importer	0.754*** (0.03)	−3.96	0.757*** (0.03)	−3.96	0.754*** (0.03)	−3.97
– Exporter	1.008*** (0.03)	−4.94	1.006*** (0.03)	−4.94	1.005*** (0.03)	−4.94
ln population						
– Importer	0.160*** (0.06)	−1.12	0.130*** (0.06)	−1.12	0.122*** (0.06)	−1.12
– Exporter	0.385*** (0.05)	−1.87	0.400*** (0.06)	−1.87	0.406*** (0.06)	−1.87
ln land area						
– Importer	0.127 (0.07)	−3.46	0.160* (0.07)	−3.46	0.150* (0.07)	−3.46
– Exporter	−0.368*** (0.07)	−4.02	−0.385*** (0.07)	−4.02	−0.402*** (0.07)	−4.02

(continued)

Table 9 (continued)

	(1)		(2)		(3)	
Variable	Active	Passive	Active	Passive	Active	Passive
Number landlocked						
– One	1.077* (0.49)	0.03	1.158* (0.49)	0.03	1.223* (0.49)	0.03
– Both	2.484* (0.98)	0.01	2.649** (0.98)	0.01	2.775** (0.98)	0.01
Number of islands						
– One	0.569* (0.29)	−0.07	0.594* (0.29)	0.00	0.597* (0.29)	−0.07
– Both	1.306* (0.58)	−0.02	1.355* (0.58)	−0.02	1.364* (0.58)	−0.02
Common land border	0.634*** (0.09)	−0.03	0.635*** (0.09)	−0.03	0.634*** (0.09)	−0.03
Common language	0.433*** (0.04)	−0.07	0.429*** (0.04)	−0.07	0.437*** (0.04)	−0.07
Ever colony	1.503*** (0.10)	−0.04	1.502*** (0.10)	−0.04	1.502*** (0.10)	−0.04
Common coloniser	0.667*** (0.06)	0.01	0.668*** (0.06)	0.01	0.667*** (0.06)	0.01
Currently colonised	−0.384 (0.80)	0.00	−0.378 (0.80)	0.00	−0.385 (0.80)	0.00
Common country	0.887*** (0.11)	−0.01	0.880*** (0.11)	−0.01	0.886*** (0.11)	−0.01
Log pseudo-likelihood	−1,885,759		−1,885,763		−1,885,594	
Observations	787,840		787,840		787,840	
Zeros	351,065		351,065		351,065	
α	3.04***		3.04***		3.04***	
Vuong	−186.37***		184.79***		−85.92***	
Method	NB		ZINB		NB	

(ZI)NB regression with bilateral imports as dependent variable. Estimates marked ***/**/* are significant at the 1/5/10 percent level. Robust standard errors (clustered by country-pair) are in brackets. Estimates for intercepts and fixed effects are omitted to save space. STATA does not report standard errors for coefficients in the "passive" group

Table 10 ZINB results by development status, 1948–2007

Variable	(1) Full	(2) Developed	(3) Developing	(4) LDC
Both in WTO	0.976*** (0.11)	0.836*** (0.10)	0.646*** (0.07)	0.273 (0.15)
One in WTO	0.651*** (0.11)	0.440*** (0.09)	0.331*** (0.06)	0.117 (0.12)
Development				
– Developing	0.453* 0.23)			
– LDC	−0.250 (0.55)			
Both in WTO × Development				
– Developing	−0.385** (0.12)			
– LDC	−0.421** (0.16)			
One in WTO × Development				
– Developing	−0.332** (0.119)			
– LDC	−0.413** (0.153)			
GSP	0.229*** (0.02)	−0.340*** (0.04)	0.147*** (0.04)	0.266*** (0.07)
EIA	0.217*** (0.04)	0.183** (0.06)	0.264*** (0.05)	0.275** (0.08)
ln distance	−1.107*** (0.02)	−0.921*** (0.04)	−1.281*** (0.03)	−1.116*** (0.06)
ln GDP				
– Importer	0.750*** (0.03)	1.085*** (0.08)	0.777*** (0.04)	0.548*** (0.06)
– Exporter	1.009*** (0.03)	1.153*** (0.05)	1.005*** (0.05)	0.822*** (0.11)
ln population				
– Importer	0.205*** (0.06)	−0.339 (0.23)	0.115 (0.09)	−1.273*** (0.26)
– Exporter	0.376*** (0.05)	−0.087 (0.09)	0.529*** (0.08)	0.389* (0.18)
ln land area				
– Importer	0.010 (0.04)	−0.227* (0.10)	0.146 (0.08)	2.277*** (0.43)
– Exporter	−0.329*** (0.032)	0.066 (0.10)	−0.174* (0.07)	−0.075 (0.13)

(continued)

Table 10 (continued)

Variable	(1) Full	(2) Developed	(3) Developing	(4) LDC
Number landlocked				
– One	1.112* (0.48)	0.574 (0.47)	2.074** (0.08)	1.344 (0.87)
– Both	2.551** (0.97)	1.763 (0.96)	4.650** (1.52)	2.938 (1.75)
Number of islands				
– One	0.706** (0.25)	1.837*** (0.22)	1.946*** (0.31)	0.621 (0.47)
– Both	1.581** (0.49)	3.937*** (0.44)	3.996*** (0.63)	1.491 (0.96)
Common land border	0.631*** (0.09)	−0.039 (0.15)	0.241* (0.10)	1.111*** (0.16)
Common language	0.428*** (0.04)	0.295*** (0.08)	0.365*** (0.05)	0.515*** (0.08)
Ever colony	1.512*** (0.10)	1.452*** (0.12)	0.990*** (0.10)	2.317*** (0.21)
Common coloniser	0.675*** (0.06)	−0.149 (0.32)	0.708*** (0.08)	0.615*** (0.10)
Currently colonised	−0.385 (0.80)	−0.493 (0.48)	0.161 (0.85)	(Omitted)
Common country	0.883*** (0.11)	0.028 (0.50)	0.832*** (0.13)	0.790*** (0.21)
Log pseudo-likelihood	−1,885,674	−682,458	−940,570	−233,698
Observations	787,840	173,956	437,107	176,777
Zeros	351,065	43,485	209,364	98,216
α	3.04**	2.05***	3.22**	2.93*
Vuong	N.A.	1.34*	31.17***	32.51***
Method	ZINB	ZINB	ZINB	ZINB

ZINB regression with bilateral imports as dependent variable. Estimates marked ***/**/* are significant at the 1/5/10 percent level. Robust standard errors (clustered by country-pair) are in brackets. Estimates for intercepts, fixed effects and the "passive" group are omitted to save space. Variables are omitted when not applicable. Development status is based on World Bank (2011a)

Table 11 ZINB results by negotiation round

Variable	(1) 1948–1951, until Torquay Round	(2) 1948–1956, until Geneva Round	(3) 1948–1961, until Dillon Round	(4) 1948–1967, until Kennedy Round	(5) 1948–1979, until Tokyo Round	(6) 1948–1994, until Uruguay Round
Both in WTO	−0.083(0.09)	0.028 (0.07)	−0.171** (0.06)	−0.223*** (0.05)	−0.055 (0.06)	0.810*** (0.06)
One in WTO	−0.038 (0.06)	−0.048 (0.055)	−0.127** (0.05)	−0.143*** (0.04)	−0.044 (0.05)	0.430*** (0.05)
GSP	(Omitted)	(Omitted)	(Omitted)	(Omitted)	0.244*** (0.03)	0.375*** (0.03)
EIA	(Omitted)	(Omitted)	−0.096 (0.11)	0.409*** (0.12)	0.363*** (0.08)	0.182** (0.06)
ln distance	−0.851*** (0.04)	−0.830*** (0.04)	−0.812*** (0.03)	−0.800*** (0.03)	−0.885*** (0.03)	−1.185*** (0.02)
ln GDP						
– Importer	0.901 (0.50)	1.366*** (0.16)	1.500*** (0.12)	1.263*** (0.09)	1.193*** (0.07)	0.905*** (0.08)
– Exporter	2.903*** (0.71)	2.119*** (0.20)	2.066*** (0.13)	1.614*** (0.10)	1.526*** (0.06)	1.071*** (0.04)
ln population						
– Importer	1.287 (0.94)	−0.440 (0.36)	0.008 (0.21)	−0.310* (0.16)	−0.175 (0.11)	0.198*** (0.07)
– Exporter	−1.961 (1.08)	−2.220*** (0.42)	−2.205 (0.25)	−1.518*** (0.16)	−0.787*** (0.11)	0.511*** (0.07)
ln land area						
– Importer	4.837*** (0.57)	−0.567** (0.20)	−0.518** (0.19)	−0.108 (0.13)	−0.221* (0.10)	0.08 (0.08)
– Exporter	−0.632 (1.03)	0.524 (0.32)	0.440* (0.21)	0.501*** (0.15)	0.105 (0.09)	−0.278*** (0.07)
Number landlocked						
– One	−0.273 (35.94)	−22.581*** (0.14)	−0.375 (1.289)	−1.546 (1.28)	−0.02 (0.73)	2.750*** (0.66)
– Both	0.053 (71.95)	−45.24 (·)	−0.591 (2.59)	−2.978 (2.55)	−0.04 (1.49)	5.681*** (1.34)
Number of islands						
– One	1.067 (1.04)	1.422*** (0.44)	1.036*** (0.38)	1.477*** (0.37)	1.419*** (0.31)	1.842*** (0.28)
– Both	2.480 (2.07)	3.204*** (0.88)	2.214*** (0.75)	3.049*** (0.73)	3.131*** (0.61)	3.918*** (0.56)
Common land border	0.086 (0.13)	0.335* (0.13)	0.155 (0.11)	0.169 (0.10)	0.181* (0.09)	0.463*** (0.09)
Common language	−0.119 (0.09)	0.282** (0.09)	0.165* (0.08)	0.243*** (0.06)	0.380*** (0.05)	0.457*** (0.05)

(continued)

Table 11 (continued)

Variable	(1) 1948–1951, until Torquay Round	(2) 1948–1956, until Geneva Round	(3) 1948–1961, until Dillon Round	(4) 1948–1967, until Kennedy Round	(5) 1948–1979, until Tokyo Round	(6) 1948–1994, until Uruguay Round
Ever colony	1.908*** (0.16)	1.770*** (0.15)	1.827*** (0.14)	1.771*** (0.13)	1.683*** (0.11)	1.650*** (0.10)
Common coloniser	1.023*** (0.19)	1.294*** (0.18)	1.085*** (0.16)	1.027*** (0.11)	0.749*** (0.08)	0.802*** (0.07)
Currently colonised	(Omitted)	(Omitted)	(Omitted)	(Omitted)	−0.073 (0.84)	−0.638 (0.806)
Common country	0.474 (0.26)	0.641*** (0.23)	0.244 (0.20)	0.382*** (0.16)	0.953*** (0.13)	0.926*** (0.12)
Log pseudo-likelihood	−22,564	−103,927	−178,165	−298,206	−655,659	−1,283,244
Observations	12,790	63,947	108,395	170,437	318,048	570,699
Zeros	7976	42,292	70,671	107,599	175,715	282,803
α	1.11**	1.67**	1.39**	1.41**	1.73**	2.70**
Vuong	48.60***	74.43***	105.64***	114.87***	150.61***	115.14***
Method	ZINB	ZINB	ZINB	ZINB	ZINB	ZINB

ZINB regression with bilateral imports as dependent variable. Estimates marked ***/**/* are significant at the 1/5/10 percent level. Robust standard errors (clustered by country-pair) are in brackets. Estimates for intercepts, fixed effects and the "passive" group are omitted to save space. Variables are omitted when not applicable

5 Discussion and Conclusion

The aim of this study was to inspect Rose (2004)'s surprising finding that WTO membership has no effect on international trade. It has been argued that the specification of the EIA variable and failure to account for zero trade "flows" are potential sources of biased results. What happens when these issues are addressed?

The first step was to replicate Rose's findings with different data to confirm the robustness of his unexpected "non-effect" of WTO membership on trade. As already indicated, doing so yields positive WTO, GSP and EIA effects. Rose's "non-effect" is not robust. Interestingly, the obtained EIA effect is remarkably lower than prior OLS estimates in the WTO effect literature suggest. Where Rose (2004), Subramanian and Wei (2007) and Tomz et al. (2007) find that EIAs increase trade by 230, 420 and 115 % respectively, the present results indicate that this effect is $e^{0.63} - 1 \approx 90\,\%$. These more nuanced findings are line with other major studies in the field of EIAs (see, e.g., Baier and Bergstrand, 2007). In comparison, GSP schemes boost trade by over 40 % and joint WTO participation by about 20 %, all else constant. This confirms that EIA selection bias gives rise to overestimated EIA parameter estimates, but that—beside WTO membership and GSP schemes—regionalism has an influential role in advance international trade.

Using (ZI)NB MLE to incorporate zero trade "flows," re-estimating the WTO effect with our updated dataset yields the following. The results in Table 9 (column 1) show that trade increases by around 90 % if both trade partners are WTO members and by 40 % if only one of them is a member. GSP schemes have an overall trade-increasing effect. Just as with the OLS estimates, the results also ascribe an important role to EIAs in stimulating trade, with EIAs boosting trade by 25 %. Interestingly, this is less than a third of the OLS estimate, which was already considerably lower than those obtained in previous studies. Hence, the improved EIA variable and methodology reduce the bias introduced by earlier estimates of the effect that regionalism has on trade. Note that most of the other gravity equation coefficients are in line with conventional estimates in the literature. The negative "Both in WTO" coefficient in the "passive" column is interpreted as follows: if both countries in a country-pair are WTO participants, it decreases their probability of having zero trade and being in the "passive" group. One could also say that it increases the probability of that country-pair having non-zero trade and belonging to the "active" group.

In Table 9 (column 2), a distinction is made between importers and exporters in the WTO and GSP variables. Non-members gain from trading with WTO members, regardless of the member's status as importer or exporter. The intuition is that members may extend the scope of their import and export liberalisation to members of anon-members alike, thereby allowing non-members to indirectly benefit from other countries' participation in the multilateral trade system (for a comparison, see Herz and Wagner, 2011, p. 1022). The results again confirm that GSP schemes are trade-promoting.

Strikingly, we find that the degree to which participants experience positive WTO effects goes hand in hand with the extent to which they make binding commitments to liberalise trade in the multilateral trade system. This empirical finding makes intuitive sense: according to the historical narrative of the GATT/WTO discussed in Sect. 2, informal participants usually were least developed countries (LDCs) or developing countries. Throughout the multilateral trade system's evolution, LDCs were exempted from most obligations on the trade agenda. These are countries that were able to gain market access for their exports, but because of highly limited activity to liberalise trade with regard to imports, the overall effect of WTO participation on trade was negligible. Developed countries had more commitments to uphold, though not as much as developed members and only from as late as the 1990s (see Barton et al., 2006; Subramanian and Wei, 2007; Tomz et al., 2007)

Consider Table 9, column 3. Here, we use Tomz et al. (2007)'s disaggregated WTO participation data to estimate how the WTO effect may differ among formal members and informal participants. The empirical evidence shows that countries gain most if both trade partners are formal participants. The gains are substantially smaller if only one is a formal participant. Remarkable is that we also find that the gains of formal membership exceeds those of informal participation, which is consistent with what one would expect to find in the context of the historic narrative. In contrast, the only other empirical study to look into this issue finds a counterintuitive and substantially weaker effect for formal membership compared to informal participation (see Tomz et al., 2007, Table 2, column 5).

Another way of exploring the issue is by including a categorical variable for a country's development status. International conventions allow countries to determine their development status based on self-declaration, i.e. whether a country participates as developed, developing or less developed nation is a (largely) political decision. Again, LDCs have the least extensive liberalisation commitments and developed countries the most. Table 10 presents statistical output for the full sample in column 1, followed by subsets based on the development level in columns 2–4.

The negatively-signed and statistically significant interaction terms in column 1 demonstrate that developing countries and LDCs experience smaller gains from WTO participation compared to developed members. The impact is less severe for developing countries than it is for LDCs. This finding holds if both trade partners in a country-pair are WTO participants, but also in the case that only one of them participates. The subsets in columns 2–4 further demonstrate that the gains are larger for developed countries than developing countries, and that there are no significant gains for LDCs. However, LDCs do gain from tailored GSP schemes.

Overall, the empirical results suggest that the more extensive a WTO participant's commitments are to multilateral trade liberalisation, the more it will experience the resulting trade-creating effects.

Allowing the results to vary across time periods reveals that the multilateral trade system did not promote trade in its younger years (see Table 11). This may be explained, on the one hand, by the increasing complexity of negotiations and on the other by the substantial number of informal participants that were not required to implement liberalisation, thereby offsetting potential gains created by

formal members' liberalisation efforts (WTO, 2011b).[12] However, the Kennedy and Tokyo Rounds have been especially influential in creating a positive WTO effect. Indeed, Preeg (1970) notes that the first five postwar multilateral trade conferences were not as important as the Kennedy Round, during which average tariffs on industrial products were cut by 35 to 40 %, with two-thirds of the cuts exceeding 50 %. Although this is not immediately borne out by the results in column 5, the positive WTO effect that participants experienced after the Tokyo Round suggests that significant long-run gains have materialised.

The empirical evidence presented in this chapter highlights three important findings regarding the multilateral trade system.

First, Rose (2004)'s initial "non-effect" of WTO membership on trade is not robust. In fact, it is difficult not to obtain favourable WTO effects. However, merely updating the data is not enough to obtain robust and intuitive results. Using an appropriate modelling strategy is equally relevant. It has been argued that the large proportion of zeros in international trade datasets creates problems for both traditional log-linear and recent approaches using Poisson QMLE alike. (ZI)NB models successfully deal with the problems of overdispersion and excessive zeros. In all cases that have been examined in this paper, the test statistics indicate that (ZI)NB MLE is more suitable than Poisson QMLE when estimating gravity equations with large panel datasets involving zero trade "flows".

Second, WTO membership has a strong trade-promoting effect among trade partners, even if not all of them are WTO participants. However, "what you give is what you get": formal members gain more than non-member participants, suggesting that the extent of WTO participants' experienced trade gains goes hand in hand with their multilateral liberalisation commitments. This result is shown to be surprisingly robust and perfectly reflects the institutional evolution of the multilateral trade system presented in this chapter.

Finally, countries with EIAs trade more compared to those without EIAs, all else constant. Using the extensive and up-to-date dataset on trade and EIAs that corrects for selection bias, this study confirms that regionalism fosters international trade, although the obtained estimates are considerably lower than those previously suggested in the literature.

Acknowledgements The author thanks Rob Alessie, Jeff Bergstrand, Maarten Bosker, Steven Brakman, Bob Ethier, Marco Fugazza, Harry Garretsen, Michael Koetter, Suzanne Kok, Xuepeng Liu, Roger Smeets, Peter van Bergeijk, participants at the Seminar on Regionalism in an Integrating World Economy (Utrecht) and the European Trade Study Group Annual Conference (Lausanne), and an anonymous referee for insightful discussions and helpful comments on previous drafts.

[12]Observing variation in countries that have been members since the GATT's inception is not possible due to data limitations. The trade liberalisation efforts of these countries may have been even more extensive than those of countries that joined later because, being GATT founders, they showed the greatest formal initiative to liberalise trade. This may imply that the obtained estimates of the WTO effect are underestimated, although this seems difficult to verify directly.

References

Allison, P., & Waterman, R. (2002). Fixed effects negative binomial regression models. *Sociological Methodology, 32*, 247–265.

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review, 93*, 170–192.

Baier, S. L., & Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics, 71*, 72–95.

Baldwin, R. E., & Taglioni, D. (2006). *Gravity for Dummies and Dummies for Gravity Equations*. Working Paper 12516. Cambridge, MA: National Bureau of Economic Research.

Barton, J. H., Goldstein, J. L., Josling, T. E., & Steinberg, R. H. (2006). *The evolution of the trade regime: Politics, law, and economics of the GATT and the WTO*. Princeton, NJ: Princeton University Press.

Bosker, M. (2008). *The Empirical Relevance of Geographical Economics*. PhD thesis, Utrecht University.

Brakman, S., Garita, G., Garretsen, J. H., & van Marrewijk, C.. (2010). Economic and financial integration and the rise of cross-border M&As. In P. A. G. van Bergeijk & S. Brakman (Eds.), *The gravity model in international trade*. Cambridge: Cambridge University Press.

Bureau of Labor Statistics. (2008). Consumer price index. http://www.bls.gov/cpi, last accessed November 8, 2008.

Burger, M. J., van Oort, F. G., & Linders, G. J. M. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis, 4*, 167–190.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. College Station, TX: Stata Press.

CEPII. (2008). Geodesic distances. http://www.cepii.fr/anglaisgraph/bdd/distances.htm, last accessed November 8, 2008.

Engelbrecht, H. J., & Pearce, C. (2007). The GATT/WTO has promoted trade, but only in capital-intensive commodities! *Applied Economics 39*, 1573–1581.

Feenstra, R. C. (2004). *Advanced international trade: Theory and evidence*. Princeton, MA: Princeton University Press.

Felbermayr, G. J., & Kohler, W. (2006). Exploring the intensive and extensive margins of world trade. *Review of World Economics, 142*, 642–674.

Gould, W. W., Pitblado, J. S., & Sribney, W. M. (2006). *Maximum likelihood estimation with Stata* (3rd ed.). College Station, TX: Stata Station.

Herz, B., & Wagner, M. (2011). The 'real' effect of GATT/WTO - a generalised approach. *The World Economy, 34*, 1014–1041.

Hilbe, J. M. (2007). *Negative binomial regression*. Cambridge: Cambridge University Press.

IMF. (1995). *Direction of trade statistics historical database (1948–1980) CD-ROM*. Washington, DC: International Monetary Fund.

IMF. (2008). *Direction of trade statistics*. Washington, DC: International Monetary Fund.

Kim, M. H. (2010). Does the WTO promote trade? Further evidence. *Journal of International Trade and Development, 19*, 421–437.

Liu, X. (2009). GATT/WTO promotes trade strongly: Sample selection and model specification. *Review of International Economics, 17*, 428–446.

Maddison, A. (2007). Historical statistics of the world economy: 1–2006 AD. http://www.ggdc.net/maddison, last accessed November 8, 2008.

McGill. (2009). PTA database. http://ptas.mcgill.ca, last accessed October 3, 2009.

Preeg, E. H. (1970). *Traders and diplomats: An analysis of the Kennedy Round of negotiations under the General Agreement on Tariffs and Trade*. Washington, DC: Brookings Institute.

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review, 94*, 98–114.

Santos Silva, J. M. C., & Tenreyoro, S. (2006). The log of gravity. *The Review of Economics and Statistics, 88*, 641–658.
Siliverstovs, B., & Schumacher, D. (2009). Estimating gravity equations: To log or not to log? *Empirical Economics, 36*, 645–669.
Subramanian, A., & Wei, S. (2007). The WTO promotes trade, strongly but unevenly. *Journal of International Economics, 72*, 151–175.
Swart, J., & van Marrewijk, C. (2009). *Cross-border Mergers and Acquisitions: A Piece of the Natural Resource Curse Puzzle*. Discussion paper, Tinbergen Institute, Rotterdam.
Tomz, M., Goldstein, J. L., & Rivers, D. (2007). Do we really know that the WTO increases trade? Comment. *American Economic Review, 97*, 2005–2018.
Tuck. (2009). Trade Agreements Database. http://www.dartmouth.edu/tradedb, last accessed October 3, 2009.
UNCTAD. (1974). Operation and effects of the generalized system of preferences.
UNCTAD. (1975). Operation and effects of the generalized system of preferences.
UNCTAD. (1979). Operation and effects of the generalized system of preferences.
UNCTAD. (1981). Operation and effects of the generalized system of preferences.
UNCTAD. (1985). Operation and effects of the generalized system of preferences.
UNCTAD. (2005). Generalized system of preferences: List of beneficiaries.
UNCTAD. (2006). Generalized system of preferences: List of beneficiaries.
UNCTAD. (2008). Generalized system of preferences: List of beneficiaries.
van Bergeijk, P. A. G., & Brakman, S. (Eds.) (2010). *The gravity model in international trade: Advances and applications*. Cambridge: Cambridge University Press.
Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypothese. *Econometrica, 57*, 307–333.
Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press.
World Bank. (1951). *Demographic yearbook 1949–1950*. Washington, DC: World Bank.
World Bank. (2011a). Classification of economics by income and region. http://data.worldbank.org/about/country-classifications/country-and-lending-groups, last accessed December 18, 2011.
World Bank. (2011b). Development indicators. http://data.worldbank.org/indicator, last accessed December 18, 2011.
World Bank. (2011c). Global preferential trade agreements database. http://wits.worldbank.org/gptad, last accessed December 18, 2011.
WorldTradeLaw.net. (2009). FTA database. http://www.worldtradelaw.net/fta/ftadatabase/ftas.asp, last accessed October 3, 2009.
WTO. (2011a). Regional trade agreements information system. http://rtais.wto.org, last accessed December 18, 2011.
WTO. (2011b). Understanding the WTO. http://www.wto.org/english/thewto_e/whatis_e/tif_e/tif_e.htm, last accessed December 18, 2011.

Part VI
Globalization and Development

The Cycle of Development in Africa: A Story about the Power of Economic Ideas

Martin Paldam

Abstract During the last 60 years development in Sub-Saharan Africa has had three main phases—P1, P2 and P3—divided by kinks in 1972 and in 1994. P1 and P3 had fairly satisfactory growth, but P2 had negative growth. This cyclical growth path has to be explained by variables with a similar path. A set of socio-economic variables representing 11 hypotheses are considered. Some of these hypotheses have been proposed to explain the low growth of Africa, while most are meant to explain the growth tragedy of P2. Most of the variables have paths with no relation to the cycle, but the path corresponds to the shifts in the dominating development strategy. At the end of P1 the main policy-package in Africa became the one of African socialism. It led to large scale rent seeking, inefficiency and economic regression. At the end of P2 policies were adjusted towards a more market based system and growth resumed.

1 Introduction

This essay deals with the economic development of Africa, which is taken to mean Sub-Saharan Africa. 44 countries on the continent are considered. I assume that they have so much in common that it makes sense to treat their development as a set of variations around a joint African path. This path has three distinct phases shown in Figs. 1 and 6 below: (P1) From the start of the data in 1950–1972 growth was decent. (P2) From 1973 till 1993 growth was negative. (P3) Since 1994 decent growth has resumed. This *cyclical* picture is evident in the data for 32 of the countries while

Previous versions of this paper were presented at the Swiss Society of Economics and Statistics Annual Meeting in Luzerne, the Australasian Public Choice Conference, Dunedin, New Zealand, and at the Globalization Conference in Kolding. I am grateful to the discussants and the referees. An earlier version of the paper was published in the *Swiss Journal of Economics and Statistics* 147, 427–59 (December 2011). I am grateful to the journal for permission to publish the present updated and (slightly) revised version. The last revisions are from September 2013.

M. Paldam (✉)
Department of Economics and Business, Aarhus University, 8000 Aarhus C, Denmark
e-mail: mpaldam@econ.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_20

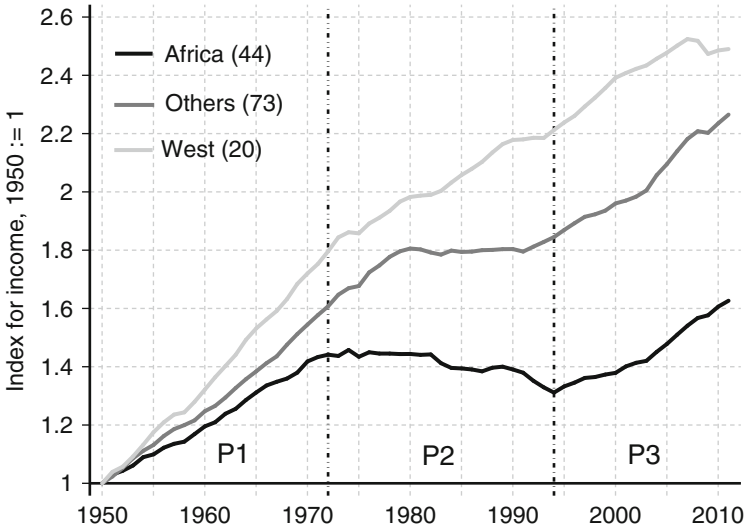

Fig. 1 The growth path of income for three country groups, 1950–2011. The calculations cover 44 *African*, 20 *Western* countries and 73 Other countries, for which data is available for the full period. The data until 2005 are from Maddison (2003, URL ref) and updated with WDI/Penn Word Tables (URL ref). A few missing observations for 2011 are from the CIA World Factbook (URL ref). The data are converted to indices starting at 1 in 1950. It is done by adding 1 and deducting the value in 1950 from all series

3 countries have a different path. The last nine countries follow the cycle for 1–2 periods, but not fully.

Due to the negative growth in P2, the GDP level from the end of P1 was only reached again in 2004, so Africa had 20 years of falling GDP per capita and 30 years of no economic growth. About 14 countries of the world have a GDP per capita that is lower today than in 1950. Of these no less than seven are African (see the left hand panel in Table 2 below). These facts have been known as Africa's *growth tragedy* following Easterly and Levine (1997).

It is covered by a substantial literature. I have searched this literature and found a number of theories used to explain Africa's development.[1] 11 of these theories are operational in the sense that they use an explanatory variable that is covered by a main economic indicator. The methodological idea of the paper is that a development that is visible in the GDP data for 44 countries must be strong. Thus, it is a necessary condition for a theory to be credible that the African average for its indicator variable should have a similar and clearly visible development.

[1]Several broader, less operational, theories are also available, such as the 'primacy of organizations' view; see e.g., Acemoglu and Robinson (2010) and Paldam and Gundlach (2008). My findings may relate to that view, but I do not discuss this in the present paper.

11 theories are a lot to cover, and the data for most of the indicators are incomplete, so I have to assume that the data for the countries covered are typical for all 44 countries. A pattern actually emerges: It appears that only one of these indicators follows the cycle of development in Africa reasonably well. It is the Fraser Index giving the big shifts in the development strategy. Consequently, if one of the theories is true, it is that the cycle is caused by the shifts in the development strategy, corresponding to the changes in the dominating sets of economic ideas in Africa and among the many development economists advocating these ideas. Thus it is a story about the power of economic ideas.

The paper proceeds as follows. Section 2 looks at the GDP data to show what has to be explained. Section 3 looks into the textbooks of economic theory for a general explanation. It discusses three theories most economists automatically reach for when they try to explain the slow growth of Africa. Section 4 considers five specific theories developed to explain period P2, the growth tragedy, and two additional theories that are supposed to apply to all three periods. These theories are all found to be empirically problematic. Section 5 looks at the only theory that is in accordance with the data. Section 6 concludes.

2 What Has to be Explained?

The analysis uses the Maddison (2003, URL ref[2]) data set updated to 2011.[3] It reports complete time series starting in 1950 for 44 countries.[4] The appendix lists the countries and brings a few descriptive statistics. The paper looks at national accounts data in real PPP prices. When GDP is written *gdp* (lowercase) it represents per capita GDP data. *Income* denotes the natural logarithm gdp. The *growth rate* is always based on the gdp-series.

2.1 The Path and the Three Periods: P1, P2 and P3

The income path of the 44 *African* countries is shown in Fig. 1. It is compared with the path for the average *Western* countries and 73 *Other* countries, which are all other countries in the Maddison set with complete series. The three periods—P1, P2 and P3—from the introduction, are indicated by the vertical lines.

Table 1 summarizes the information in Fig. 1. The averages in P1 and P3 do not differ significantly, but growth in P2 is significantly lower than growth in the other

[2](URL ref) means that the references bring the net address (URL).

[3]National accounts for many African countries are weak and some have gaps or jumps. The paper takes the data for granted. They are updated to 2011 as explained in the note to Fig. 1.

[4]The only country I would have included if data had not been missing is Eritrea. By choice Mauritius, the Seychelles and South Africa are excluded. However, South Africa follows the cycle perfectly well.

Table 1 Average growth rates for 1950–2011

Group	Countries	Growth all years		The three periods			Change		N for all
		Per year	Times	P1	P2	P3	ΔP2	ΔP3	periods
Africa	44	1.21	2.1	2.14	0.30	1.84	−2.44	2.14	2684
		(0.23)		(0.31)	(0.46)	(0.44)			
Others	73	2.27	3.9	2.95	1.27	2.59	−1.68	1.32	4453
		(0.17)		(0.29)	(0.32)	(0.26)			
West	20	2.51	4.5	3.72	1.90	1.73	−1.82	−0.17	1220
		(0.15)		(0.26)	(0.23)	(0.25)			

Note: Same data as used for Fig. 1. Growth is in per cent of gdp, and brackets report 2 standard errors. *Source*: Maddison (2003, URL ref)

periods.[5] While the downswing from P1 to P2 is relatively large in Africa, so is the upswing from P2 to P3.

2.2 *The Underlying Political Dynamics of Economic Ideology/Development Strategy*

In a broad historical perspective the *underlying political dynamics* had two mutually reinforcing parts: (i) The learning-to-rule part. The experienced colonial powers managed to set Africa on a nice growth path. The inexperienced new governments led the continent into economic trouble. And then they learned.[6] (ii) *The* reacting-to-colonialism part. The dynamics of independence caused African politicians to promise far more than they could possibly deliver. Also, they ascribed all problems to the misrule and exploitation of the imperialists. If they had pursued different policies, they would have done much better, but it became a tragedy, and time passed so the urge to differ from the old colonial policies weakened and hence the second shift. This political dynamics also has three phases: D1 to D3[7]:

(D1) From 1950 to about 1965 the countries had ***MS***, market systems, as preferred by the main colonial powers.[8] In several countries independence came

[5]The only exception is that in the West growth does not increase from P2 to P3.

[6]It is a problem for this explanation that two of the countries that have followed the cyclical path most closely are Ethiopia and South Africa (not included).

[7]It is difficult to classify development strategies in periods, but the two strategies D2 and D3, chosen by the African countries themselves, were also the strategies advocated by the organizations speaking for Africa, such as the Organization for African Unity/African Union and the UN Economic Commission for Africa.

[8]Prior to World War II the African countries had economic systems that might be characterized as colonial systems, but fairly soon after the war it became clear that most colonies might be independent at some future point in time, and the main colonial powers started policies to create viable economic systems.

later and hereby the possibility to choose another development strategy. (D2) The 20 years 1965/85 was the ***AS***-period of African Socialism, was the typical ***ISI***-package, for Import Substitution Industrialization, with trade protection and ***SOE***s, state owned enterprises. A main purpose of these policies was self-reliance. It was to be reached by insulating the economies from the world market. (D3) From about 1985 the preferred economic system changed back to a ***MS***. To go there most countries implemented a set of ***SA***, Structural Adjustment, of trade liberalizations and privatizations of SOEs.

This zig-zag movement in economic ideology is typical for the LDC world,[9] though it has been relatively strong in Africa. At a first glance, the three Ds do not appear to correspond to the three Ps. Section 5 looks at data showing the cycle in the actual policies. Due to implementation lags the correspondence turns out to be much better than it appears at first.

2.3 Two Exogenous Shocks: The Oil and the Debt Crises

In the 60-year period Africa was hit by two strong international shocks. They affected all three country-groups shown in Fig. 1, but somewhat differently:

The *Oil Crisis* of 1973. It originated in the Middle East, and affected most countries from 1974/75. Part of the downswing in P2 was no doubt due to the Oil Crisis, but as shown on Figs. 1 and 6 the kink started a bit before in Africa. Also, the downswing due to the Oil Crisis ought to have been relatively modest in Africa for two reasons: (i) The policies of self-reliance pursued. (ii) The African countries have relatively large subsistence sectors, which should not be affected at all. However, the fall in the growth rate was relatively large.

The *Debt Crisis* of 1982. From 1973 to 1982 most LDCs borrowed heavily to offset the effects of the Oil Crisis, and in other LDCs the kink in 1973 was small. However, debt burdens rose ominously. In August 1982 this led to a chain reaction of defaults starting in Latin America. For most LDCs the kink came in 1982 and caused one decade of zero growth—known as the *lost decade* in Latin America. In Africa the kink came a decade before and led to the growth tragedy of two decades of negative growth. The effect of debt is further discussed in Sect. 4.6.

2.4 The African Divergence in P3, and a Summing Up

Figure 2 shows the relative distribution of incomes—the σ measure of convergence—of the 44 countries compared with a similar curve for the West. The two country-groups had similar relative income differences in 1950, but while the

[9]LDCs are Less Developed Countries, while DCs are Developed Countries.

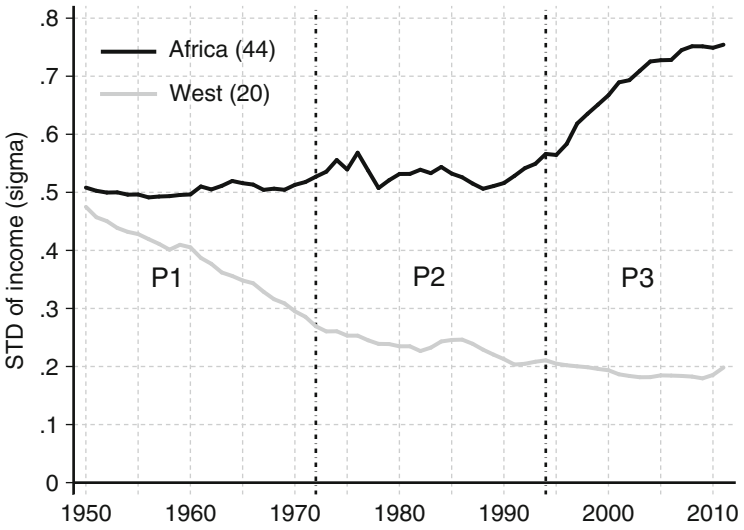

Fig. 2 Convergence/divergence in Africa and the West by the σ-measure. The σ measure of convergence is the standard deviation of income, i.e., of the logarithm to GDP per capita. Data source: Maddison (2003, URL ref)

West converged, Africa stayed at the same level till P3, where a strong divergence started. Maybe it is now leveling out, but it is certainly at a higher level. It follows that while the average African curve on Fig. 1 makes sense till 1992, it is less representative since then. The upswing since 1993 has affected some countries more than others.

Consequently, this section has shown that the African development is unusual in two ways: (1) It has had a relatively strong cyclical path, and (2) a relatively low average growth rate. The two facts interact: Imagine that the downswing in P2 had been only half of what it actually was, i.e., −1.22 percentage points only. Instead of the 1.21 % growth rate reported, this would have caused Africa's growth to have been 1.70 %. It would still have been lower than the one of the two other groups in Table 1, but not much lower. The low growth is thus partly—but not fully—due to the growth tragedy of P2.

Section 3 looks at the low growth of Africa, while the rest of the paper concentrates on the cyclical path. A convincing explanation of the cycle needs to have two characteristics:

(a) It has to work through variables that are time variant over the three periods.
(b) The variables have to be able to affect growth sufficiently to explain the cycle.

3 Looking in the Textbooks for General Explanations

The data has (44 *x* 61 =) 2682 African growth rates, of which 33 % are negative. The theory of economic growth suggests that the low growth of Africa is due to a low level equilibrium trap. A trap means that the growth diagram showing growth as a function of some determinant is non-linear, having a negative section that has to be overcome. To look for traps I consider scatter diagrams with kernel-regressions with a sufficiently low bandwidth.[10]

I cover the two main traps proposed in the literature by asking if the data look as demanded by these theories. The many negative growth rates lead to a third possibility: Maybe a negative shock can put an African economy on a stable negative growth path?

3.1 Malthus' Low Level Trap

The first theory of a low level equilibrium trap goes back to Malthus (see Blaug 1962, Chap. 3). The trap is that increased income causes an extra population growth, wiping out the income per capita increase. Figure 3 shows how income explains the net-population growth rate.

If Malthus' theory explained Africa's development in the last 60 years, the *kernel-curve* curve should be non-linear and have a *positive* slope at the low end in Fig. 3. At the start and the end of the sample the average is only supported by few observations, but from an income level of about 6 to about 8 the curve is well estimated.

The figure shows no signs of a positive slope. The kernel-curve has an insignificant negative slope for most of its range until the last observation is added. If the bandwidth is reduced, the kernel becomes less stable at the ends, but stays the same in the range with many observations. The explanation of Malthus therefore seems irrelevant.[11]

The level of African population growth is high. Over most of the income range, i.e., from 6 to 8 on the horizontal axis, population growth was stable around 2.7 %,

[10]The kernel is a continuous average with a fixed bandwidth. Kernel-curves, such as the ones in Figs. 2 and 3, are rather insensitive to the bandwidth. I have experimented with the range from 0.1 to 2, getting much the same results. Below 0.1 the regression-line becomes very erratic and above 2 it becomes a straight line.

[11]For 1700 and 1820 data exist for 12 Western countries. The annual gdp and population growth rates were 0.12 ± 0.05 and 0.45 ± 0.11 respectively, so it made sense that Malthus discussed if the real growth could be wiped out by small changes in population growth. The numbers for Africa since 1950 are quite different.

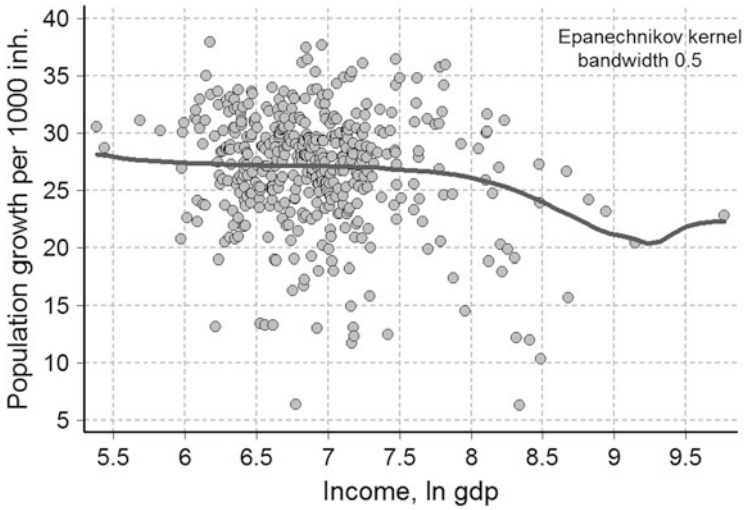

Fig. 3 Looking for Malthus' trap: Population growth over income. The curves are made for 10 observations per country for income and the crude population growth. The income data on the horizontal axis are for the years 1960, 1965, ..., 2005, and the vertical axis considers averages for 1960/64, 1965/69, ..., 2005/08. *Sources*: Maddison (2003, URL ref) and WDI (URL ref). The kernel-curve is robust to bandwidths from 0.2 to 1, where it becomes linear

which is close to the biological maximum of just above 3 %. This gives the African countries population increases of 4.5 times (±0.3) in the last 60 years.[12]

3.2 Solow's Low Level Trap

In standard growth textbooks a version of the Solow model with two equilibria is quite common. As the reader will recall it analyzes a $y = f(k)$-production function, where capital per capita, k, explains production per capita, y. This is drawn in Fig. 4, showing a double-humped production function. The model has two attractors (y_{low}, k_{low},) and (y_{high}, k_{high},), see Aziariadis and Stachurski (2005).

If $0 < k < k_{piv}$, k converges to k_{low}, which gives one attractor (y_{low}, k_{low}).

If $k_{piv} < k$, k converges to k_{high}, which gives another attractor (y_{high}, k_{high}).

The model thus exhibits a pivotal point (y_{piv}, k_{piv}) that is an unstable equilibrium.

The trap property of the model is that if income rises for some reason, but does not exceed the pivot, it falls back to the low equilibrium. When y_{low} and

[12]Maybe Africa could have had the same gross growth with half the population growth. This would have increased the real growth rate per capita from 1.2 % to about 2.5 %. See however Boserup (1965).

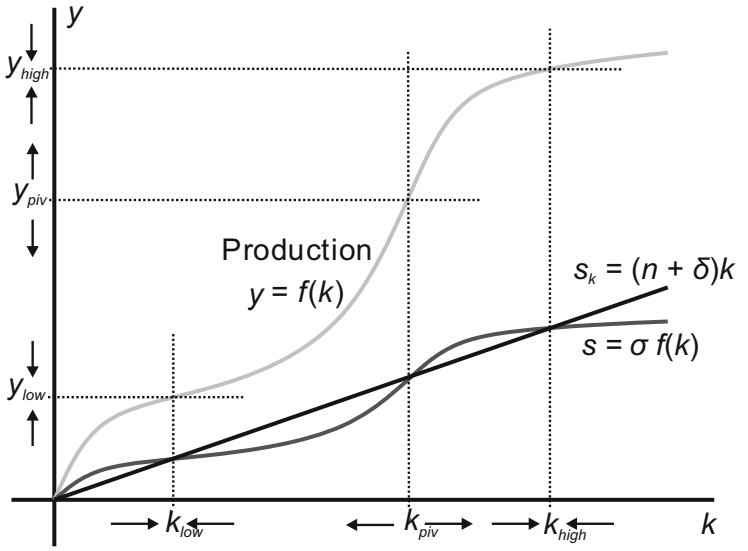

Fig. 4 The Solow model with a low level equilibrium trap. The model contains a well-behaved production function $y = f(k)$ where y and k denote production and capital per capita. The savings-investment function is s, where σ is the propensity to save. The s_k-function models the savings that keep k constant, where depreciation δ and population growth n are covered. For a lucid textbook presentation see Jones (2002)

y_{high} are wide apart, as e.g., the average African and Western countries, and y_{piv} is somewhere in the middle, it becomes difficult for an African country to exceed the pivot. Several explanations can be given why the production function may be double humped. The principle being that it aggregates a dual production structure, where the modern production technology has a large barrier to entry in the form of capital requirements.

The (y_{high}, k_{high},) equilibrium is well known. It defines a production at the 'world' technology level of the West. It suggests that all DCs converge to the same steady state precisely as shown in Fig. 2. This has been known since Baumol (1986).

Imagine that the (y_{low}, k_{low}) equilibrium is, in the same way, determined by a common technology. This appears possible on a continent where all countries border each other and are isolated from the rest of the world by large oceans and wide deserts. Thus, there may be convergence to a common equilibrium. Figure 5 is calculated as Fig. 3. If there is a low level equilibrium, a substantial part of the kernel-curve to the left should have a negative slope. The kernel-curve actually has a negative slope, but the slope is modest, and in the standard convergence regressions

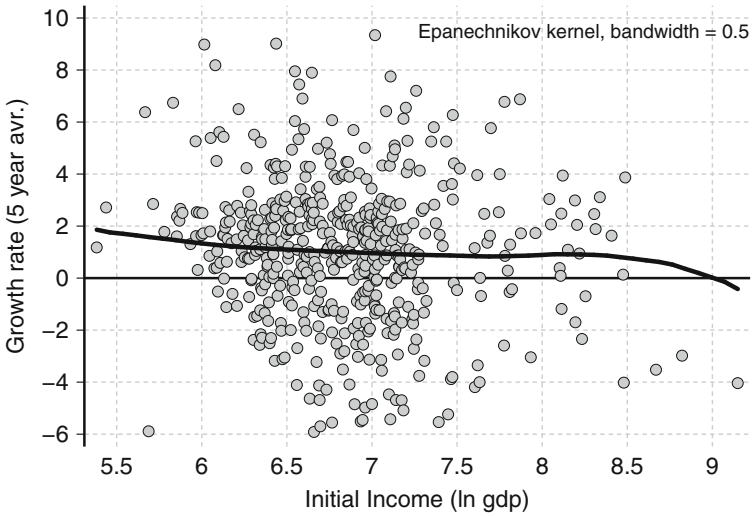

Fig. 5 Looking for Solow's trap: Income growth over income. The data are 12 observations per country for initial income (ln gdp) for the years 1950, 1955, ..., 2005, and the vertical axis considers average growth for 1950/54, 1955/59, ..., 2005/09. Thus, $N = 528$, but 13 observations are outside the frame

it is insignificant.[13] Thus, it does not appear that the African continent has been held back in a low level equilibrium trap.

3.3 A Steady State Negative Growth?

The third possibility is that the African economies can generate a steady-state negative growth rate. Thus maybe a negative shock as the oil crisis of 1973 set the countries on that path. Table 2 shows the worst performers as regards growth.

The table has two sections: The left-hand panel shows the countries from Sect. 1 when the growth tragedy was defined. These countries had some growth in the 1950s and 1960s, but they peaked around 1973, and in the last 38 years they have had falling gdp by an annual rate of 1.6 %. The right hand (gray) panel covers an additional ten countries that had a positive growth over the full 61-year period. In most cases a peak occurs some time ago and the income has been falling since then.

[13]The regression for absolute convergence with $N = 396$ is: $gy = +1.71(0.7) -0.13(-0.4)\ y$ with an $R^2 = 0.0004$, where y is (initial) income, gy is the growth of gdp and brackets hold t-ratios. If the 11 extreme observations are deleted, R^2 increases to 0.012 and the coefficient on y decreases to −0.63 (−2.2), which is significant, but still small. If the four additional extreme observations (two to the left and two to the right) are deleted, the coefficient on income once again vanishes.

Table 2 The 20 countries with the worst economic performance 1950–2011

	Countries with negative or negligible growth					Countries with low growth				
	Country	Growth [a] since 1950	Peak	Since peak Years	Since peak Growth [a]	Country	Growth [a] since 1950	Peak	Since peak Years	Since peak Growth [a]
1	Congo, Ki	-1.2	1974	37	-2.9	11 Togo	0.2	1979	29	-1.8
2	Madagascar	-0.6	1971	40	-1.6	12 Senegal	0.3	1965	43	-0.0
3	CAR	-0.6	1961	50	-1.1	13 Gabon	0.4	1976	32	-0.3
4	Niger	-0.3	1965	43	-1.3	14 Benin	0.5	1982	Has caught up	
5	Djibouti	-0.2	1977	35	-1.3	15 Zimbabwe	0.5	1973	35	-1.0
6	Somalia	-0.1	1974	37	-1.0	16 Zambia	0.5	1965	43	-0.5
7	Comoros	-0.1	1972	39	-1.9	17 Burundi	0.6	1991	20	-1.7
8	Liberia	0.0	1978	33	-1.1	18 Chad	0.7	2005	6	-0.4
9	Côte d'Ivoire	0.0	1980	31	-2.1	19 Uganda	0.8	1969	Has caught up	
10	Sierra Leone	0.2	1981	30	-1.5	20 Angola	0.8	1973	40	-0.1
	Average	-0.3	1973	38	-1.6	Average	0.5	1978	31	-1.2

Several country names are shortened. The first two letters of the capital city distinguish the two Congos. (a) *Growth* is the average annual growth of gdp. *Source*: See Fig. 1

The average peak year is 31 years ago and the fall in income is on average 1.2 % since then.

Thus half of the African countries have had 3–4 decades of falling income. I have analyzed the distribution of the 2684 growth rates. If a stable negative growth rate exists, the distribution should be bimodal, with a second peak in the negative section of the distribution. Growth rates typically have a symmetrical distribution with long tails, and so do African growth rates. I have found no indications of a negative peak in the distribution.

We may also turn to the theory of economic growth and ask: What are the conditions under which the growth rate becomes negative for 3–4 decades? The Solow-model uses a neoclassical production function[14]:

$$Y = F(A, K, L, H) \tag{1}$$

It can only produce negative growth (per capita) under two circumstances:

(1) Investment is so low that the capital stock is constantly falling and the elasticity of substitution is small. In the limiting case of a Harrod-Domar model it might happen if savings are small. But the average African country receives about 14.5 % of GDP in development aid, which in principle is meant to finance development.[15]

[14] Y is GDP; A is technology; K is the real capital stock; L is the labor force; and H is the stock of human capital.

[15] If 8 % of the 14½ % aid finances investments and there are some additional investments, the share of investments (in GDP) in Africa is well above 10 %. This means that K must be growing. This assumes full fungibility; with less than full fungibility the investment share is higher.

(2) Technical progress is negative, which means that techniques are forgotten at a steady rate. This also appears fanciful. Human capital has actually increased in Africa at a rather satisfactory rate, and under these conditions A is unlikely to fall.

Thus it is not easy to explain the long periods of negative growth in many African countries. However, model (1) has often been expanded with additional variables. One of these is *Es*, the economic system of the country[16]:

$$Y = F(A, K, L, H, Es) \tag{2}$$

In the last 60 years two groups of poor countries have had savings/investment rates of twice the ones of the West: East Asia and East Europe. While the former are the East Asian high growth countries that did catch up with the West, the latter were Communist countries where the gap to the West widened. Obviously, the economic system of the communist countries must have been less efficient,[17] so the economic system does matter. Maybe the negative growth in period P2 in Africa was due to a combination of small to moderate investments and the change into a bad economic system that is discussed in Sect. 5.

4 Specific Factors Explaining Africa's Development

As the general explanations seem irrelevant, researchers have looked for *specific* explanations. This section first considers five theories proposed in the literature on P2, the African growth tragedy, and then two theories that are supposed to be more broadly applicable. Each of the seven theories claim that one factor is crucial for the poor growth performance.[18]

To a large extent the specific explanations are a reaction to the theory that the cyclical path of Africa can be explained by the underlying political dynamics as set out in Sect. 2.2. This explanation appeared too easy, and this has generated a body of literature trying to explain the African growth tragedy by more exogenous factors. The five leading ones are:

(1) Bad geography: (a) many landlocked countries, (b) widespread diseases, and (c) uncertain rainfall and poor soils.
(2) Bad history: (a) the slave trade, (b) imperialism, and (c) the colonial legacy.

[16]This can be done in different ways, see e.g., Easterly and Levine (2003).

[17]See Chap. 4 in Paldam (2002b) and the brief summary around Table 3 in Paldam and Gundlach (2008), and Paldam (2003) on East Asia.

[18]Some of the factors may overlap, and several of the factors may work independently at the same time.

(3) The tribal structure of the African countries and the arbitrary borders of the countries.
(4) The relatively high frequency of war/civil war.
(5) Poor governance.

4.1 Bad Geography

The poor geography of Africa has been discussed by, e.g., Gourou (1966) and Kamarck (1967), and recently by Gallup et al. (1999). Geography is basically time invariant, but the importance of a geographical factor may change with technological progress.[19]

The factor of poor location may change with better transport technologies, the fertility of soils may increase with new agricultural technologies, and the high frequency of some diseases in Africa may fall with medical progress and when development changes the organization of society.[20] In poor countries geography is thus more important than in wealthy. However, technical progress gives trends, not cycles, in the importance of the factors.

4.2 Bad History

It has sometimes been argued that the pre-colonial history dominates development today. Nunn (2008) argues that effects of the slave trade are important until this day. The trade was concentrated in the eighteenth century (see Curtin 1969), so it appears a bit far-fetched, and it can certainly not explain the cycle of African development one and a half century later. Also, the areas most affected by the trade appear to be precisely the most developed ones, due to their good location relative to international trade.

The main historical explanation is the imperialism family of ideas, which were popular in the 1960s and 1970s, see e.g., Fanon (1961) and Rodney (1972). These theories claim that it explains African poverty that most of Africa was colonies subject to imperialistic exploitation. It follows that once the yoke of imperialism was lifted at independence, growth should catch up. However, the longer lasting colonies have good geography and better development than the colonies of shorter duration. So causality is difficult to untangle, and if colonial

[19] Apart from microstates Western Europe has only two landlocked countries (Austria and Switzerland). That they are landlocked is no problem today but perhaps it was 200 years ago.

[20] A relevant story is the one of malaria, which is particularly widespread in Africa for geographic reasons. Today it is controllable with enough organization, but the creation of such organization is conditional on development, so malaria has a complex causal relation to development, see Carstensen and Gundlach (2006) for a survey of the discussion and the empirics.

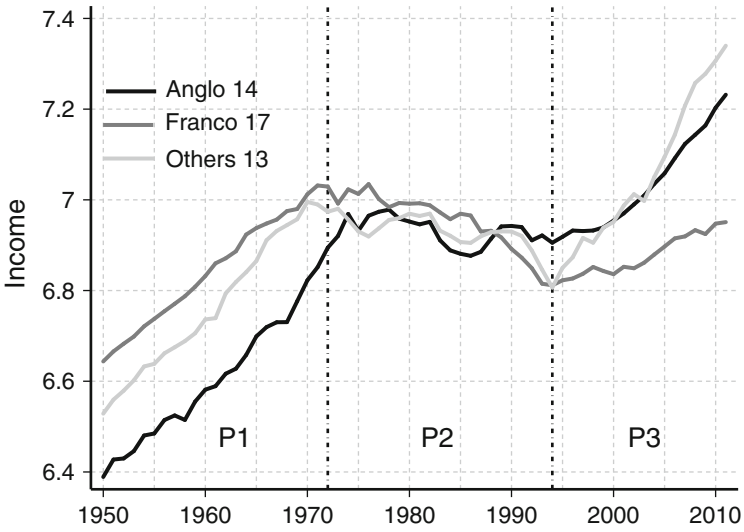

Fig. 6 A division of the 44 African countries by colonial legacy. *Data source*: See Fig. 1

past is a problem, the effect should be falling over time and not have a cyclical path.

The problem for these explanations is that most countries have been colonies in various periods of their development, and the data rather point to colonial past as an advantage; see La Porta et al. (1998) and Grier (1999). These studies, and a handful of later ones, discuss if the different colonial legacies influence future growth. Figure 6 divides the 44 countries into three groups: 14 ex-British colonies, 17 ex-French colonies and 13 others (ex-Portuguese, ex-South African, ex-Spanish and independent). The paths for each group are strikingly similar, though the data give some support to the hypothesis that British legacy is better.[21] However, the colonial legacy of countries is not the decisive factor for the cycle in their development.

Figure 6 has an additional implication. Below a handful of institutional indices are considered. They are rarely complete for all 44 countries or for the full period from 1950 to 2011. Figure 6 suggests that even if the data have gaps, the main trends are so strong that they also appear in random samples of the countries.

[21]The difference between the curves for the Anglo and the Franco countries in Fig. 5 has a strongly significant trend, so that the Anglo group that used to be the poorer is now the richer. However, the group of Others is doing much the same as the Anglo group.

Table 3 Average values of the ethno-linguistic fractionalization index

	Avr	Std	N
Africa	0.66	0.23	45
West	0.28	0.21	25
Others	0.41	0.25	100
All	0.46	0.27	170

Values from the ELF85 index
Source: Roeder (URL ref)

4.3 The Tribal Structure

Tribal diversity is an important fact of life in Africa, and Easterly and Levine (1997) claimed that it is the key to the growth tragedy. The ethno-linguistic fractionalization index is calculated from a classification of the ethno-linguistic groups of each country.[22] The index shows the probability that two arbitrarily chosen people from a country belong to different groups.

Table 3 shows that the average African country is relatively divided. One of the most divided countries in Europe is Switzerland where the index is 0.58. The average African country is even more divided. The literature since Easterly and Levine (1997) has developed a handful of related indices from the underlying data. They might be related to the underlying long-run growth, but the tribal structure changes slowly and cannot explain the cyclical pattern of African development. The tribal structure, partly created by the arbitrary borders, may be the factor behind the next two explanations.

4.4 Many Wars/Civil Wars

A great deal of effort has been put into documenting the extent of war/civil war in the world. The various indices point to Africa as the most war-plagued continent. This certainly has costs in the form of a growth loss. Figure 7 shows the path of conflicts from the PRIO (URL ref) – until 2000 there is a clear upward trend in the two series. Thus the data supported the analysis of Collier and Hoeffler (2004) up to a point. In P1 the various uprisings against the colonial power ceased and things looked well, but then pre-colonial conflicts resumed and new conflicts started, mostly due to tribal tensions. P2 starts with a dip in conflicts, but then they rise steadily. P3 starts with a small dip, but then the level rises—however, after 2000 conflicts fall. Thus, the path of conflicts does not follow the cyclical development in Africa.

[22]The indices should be almost time-invariant, but several classifications of tribes and religions are possible and the available information differs over time—and so do the indices (see Roeder 2001). Also, the indices may reflect religious differences as well.

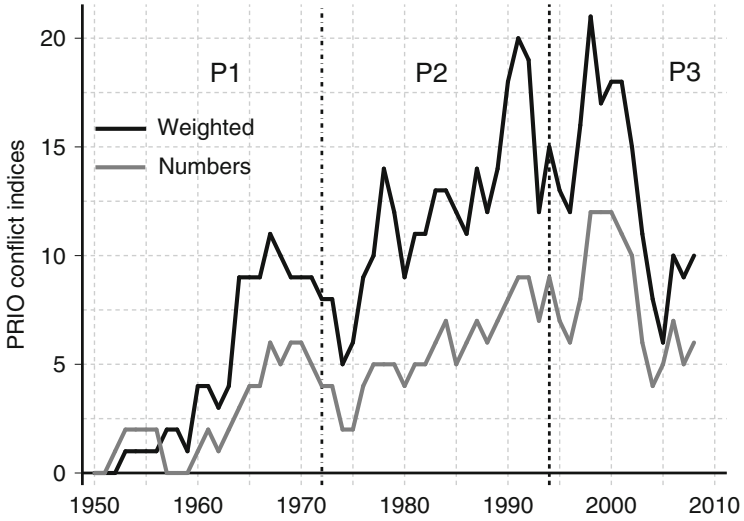

Fig. 7 The paths of the PRIO-conflict indices for Africa. Index is complete for all 44 countries from 1950 to 2008. *Weighted* has two values: 1 if more than 25 are killed, and 2 if more than 1000 are killed. *Number* counts the number of conflicts where more than 1000 have been killed since the conflict started. I have only included the largest conflict in each country. *Source* PRIO (URL ref)

Figure 7 rather looks as if causality is from development to conflicts. Period 1 contains the big liberation wave where countries move out of imperial peace and can fight without interference. In period P2 there is a significant upward trend so negative growth leads to more war. In period P3 the trend is down so that new growth leads to less war.

4.5 Poor Governance

It is easy to tell stories as do Harden (1991) and Naipaul (1975) about bad governance in various African countries, but the world has seen many bad governments, also on other continents. The data do not confirm that African governments are unusually bad. Both the two main democracy indices (Polity and Gastil) and the Transparency International corruption index get an insignificant coefficient to Africa when controlled for income.[23]

The PTS-index (URL ref) gives Amnesty International's evaluation (1976–2009) of the level of government terror directed at the population. The index is scaled

[23] See Gundlach and Paldam (2009) for the Polity democracy index; Paldam and Gundlach (2011) for the Gastil democracy index (from Freedom House); and Paldam (2002a) for the corruption index.

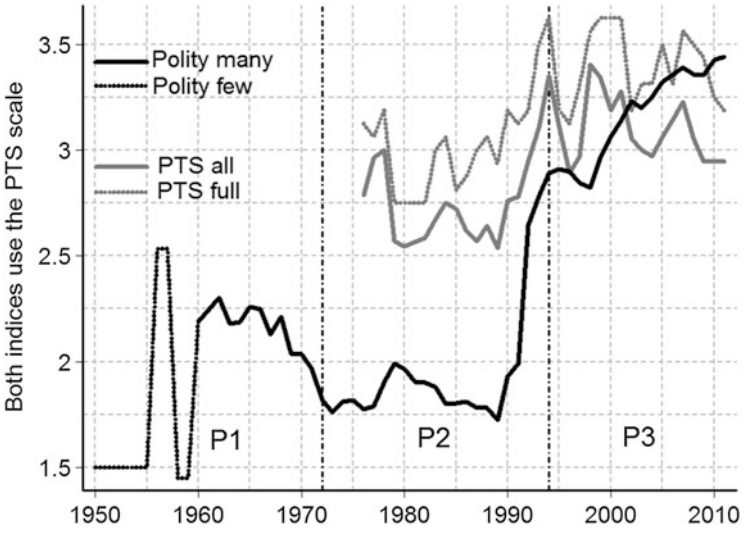

Fig. 8 The path of governance. PTS all is average index for all available observations while, PTS full is the average for the countries that have almost complete series (except for a few interpolations). The Polity index covers very few countries before 1960. *Sources*: PTS and Polity (both URL ref)

from 1 to 5, where 1 is the ideal and 5 is for terror that covers everybody. The West is 1.58 ± 0.02, Africa is 2.91 ± 0.03, and all other countries are 2.90 ± 0.02, so if corrected for income, African countries are less brutally ruled than other countries. When these data are considered in more detail, some country-scores appear surprisingly unfair, but the calculations are done on all 4485 available observations so, they are robust to a great deal of measurement error.

Figure 8 compares the paths of the PTS-index for state terror and the Polity II index (URL ref) rescaled to the PTS-scale. Polity is thus rescaled to 1 for a fully authoritarian regime and 5 for full democracy. The graph shows that the level of democracy increases with a full Polity-point (due to the rescaling it is by 25 % of the full scale) just after 1990 and has even increased by another half point since then. However, at the same time the PTS-index went up, indicating an increase of suppression by half a point. Section 5 argues that the two seemingly contradictory developments may both be true.

From the analysis of governance it is clear that Africa suffers from bad governance, but only to the extent corresponding to the poverty of the continent. Poor countries have a low level of executive capacity.

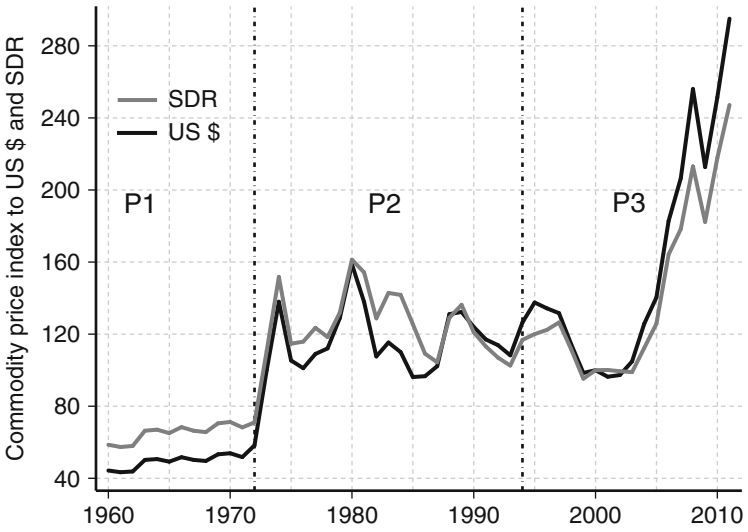

Fig. 9 UNCTADS commodity price index. The two versions show commodity prices relative to the US$ and the SDI. *Source*: UNCTAD

4.6 Hypotheses Related to All Three Periods: Debt and Commodity Price Movements

Several additional explanations have been applied to both the shift from P1 to P2 and the shift from P2 to P3.[24] Two will be discussed:

As mentioned in Sect. 2.3, the African countries have seen a large debt cycle in the period from 1970 and onwards (see Freytag and Paldam 2011). The big upswing in debt was clearly a reaction to the crisis, so it was not causal for the shift from P1 and P2. Also, the big wave of debt forgiveness was in the period from 2003 to 2005 and cannot explain the shift from P2 to P3. While the relation between borrowing and growth is negative, it does not explain the cycle of African development. In fact, debt explains little of the variation in growth rates. It is as if the African governments quickly understood that they would never have to pay back what they borrowed.

It is a tradition in Africa—supported by UNCTAD (URL ref)—to ascribe all major internal economic fluctuations to the world market. UNCTAD has compiled the commodity price index shown in Fig. 9 to support its argument. The downswing

[24]The literature on P3 is modest at present, see Arbache and Page (2010), Johnson et al. (2007) and Beny and Cook (2009).

in the growth rate from P1 to P2 was surely triggered by the first oil price shock in 1973 as already discussed.[25]

The next major change in the development path of the African countries from P2 to P3 was not triggered by commodity price shocks. Several papers, notably Beny and Cook (2009), discuss if the world market had some role to play. They do find some small effect, but if the last decades of period P2 and P3 are compared, there do not seem to be important systematic changes in commodity prices that can explain the shift.

It is interesting to see that the commodity price shock around 2005 is larger than the one in 1973, but the effects on the growth rates (see Figs. 1 and 6) are barely noticeable.

5 The Cycle of Development Strategy

The underlying political dynamics of Sect. 2.2 suggest two major shifts in development strategy: (D1) The colonial powers had installed an MS, Market System, when the countries became independent. (D2) From about 1965, the African countries started to pursue AS, African Socialism. (D3) From the mid 1980s they turned to a policy of SA, Structural Adjustment, returning the countries to an MS. As suggested by the abbreviations, the AS and SA policies are the reverse. AS is a policy of regulation and SA is a policy of deregulation.

5.1 The Path of Economic Regulation

The best data available for the degree of regulation in African are the Fraser Index of economic freedom (URL ref), which is defined as the freedom to run a private business.[26] This is an index of policy outcomes, so it differs from the economic ideology by the implementation lag.

The index measures the distance to the *laissez faire*[27] on a scale where 10 is this ideal, and 0 is as far into regulation as a country can possible get, or a collapse of law

[25]The UNCTAD-data does not cover the commodity price shock in 1951 in connection with the Korea war. It was as large as the Oil shock, but lasted only about 1 year. This commodity shock did not seem to have had much effect on African development.

[26]The index is compiled by an international network of more than 50 liberal think tanks. The information used is well explained and documented in the annual publication, see Gwartney et al. (2012), and the discussion in Paldam (2003).

[27]The laissez faire has law and order and no more regulation—this is an ideal for the group compiling the index. The closest to this ideal is Hong Kong, where the index is around 9, while the lowest value recorded is 2.3 for Nicaragua in 1985. The few recorded values from Eastern Europe in the Communist period were in the range from 3 to 5.

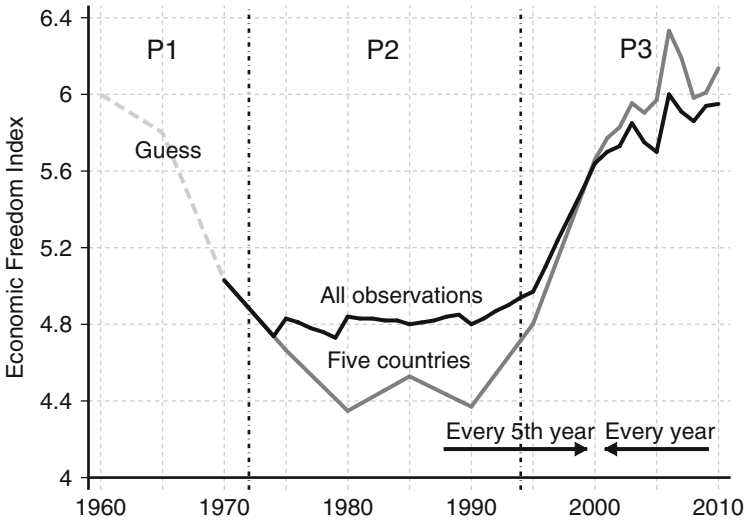

Fig. 10 The Fraser Index for economic freedom. The index covers 33 countries, but not for all years. Only five of these are covered all years. My guess for 1960–70 is very uncertain, but the direction of the move from 1960 to 1970 is clear. *Source*: Economic Freedom Index (URL ref)

and order. The data started in 1970 where only five African countries were covered. Later the African sample increased to 33 countries. The average of all observations for Africa is 5.25, so the level is moderate. The West has 7.15 and others have 6.30.

Figure 10 shows the path over time for these data. The dark gray line covers the five countries with almost full data (with only one interpolation), while the black line is an average of all available observations. Fortunately the two curves have roughly the same shape. The data only starts in 1970. It appears likely that the level in the colonial period before 1960 was about 6.[28] So the data show a cyclical path. It corresponds to the three periods:

P1 1950–72. Gradually after 1965, the AS was introduced. It increased the level of regulation to just below 5 in Africa. The increase tapered off after 1975, but it only peaked around 1985.

P2 1973–93 was the period of the implemented AS-policy regime. One of the main purposes of the AS was to isolate the African economies from the vicissitudes of the world market. Therefore, the countries should have been relatively well-protected when they were hit by the oil crisis, but as already mentioned they were hit harder than everybody else.

P3 after 1993. Here, regulations were released and the economy quickly resumed growth as predicted by the analysis of de Haan and Sturm (2000).

[28] At that time the colonial powers probably had index values a bit below 7, and the legal systems in Africa were less efficient than in the colonial 'motherlands'.

Of all indicators considered in this paper, none correspond so well to the three phases in the development of Africa as the Fraser Index. It 'explains' both the shift from P1 to P2 and the shift from P2 to P3.

It is arguable that the cycle is optimal according to the infant industry argument. Several schools of thought claim that the import substitution strategy such as the one of the AS-package is the ladder to development. The argument is summarized in Chang (2003).[29] In P2 infant industries were created in a protected environment. When they had grown strong enough, protection was gradually lifted, and now the countries have a competitive new industry and grow faster than before the policy started in P1. The main problem with this theory is that Africa grew before the AS-period, and it did not grow faster after the AS period than before. There is no point in going through a period of negative growth if there is no premium at the end.

Hence, I conclude that the explanation of the cyclical development path of Africa is simple: *The shift to negative growth was caused by the introduction of the AS-strategy, and when it was abolished, growth resumed.* It is surely debatable how strong the evidence is, but from now it is taken for granted that the said relation is clear. Also it follows the underlying political dynamics from Sect. 2.2.

5.2 The Shift from Colonial Mixed Economy to the AS Regime

The shift from the colonial mixed systems to AS was justified by three beliefs: (1) Industrialization is the key to development. (2) The main generator of economic variability is the world market. (3) The African countries did not generate enough private savings. Thus the AS-package of policies levied substantial taxes on agriculture (see Bates (1981) and Krueger (1992)) and formed many SOEs (State Owned Enterprises) in the industrial sector which were allowed substantial rents due to tariffs protecting the economy against the world market.

The shift to AS-policies was also affected by the politics of national independence and the Cold War. Independence was supported by left wingers in the West and by the East Block. So in order to balance between West and East and become more independent, African countries wanted to become more socialist.[30] However, to be workable the AS-policies must be implemented under two conditions:

[29]In the 1960s and 1970s many ISI (import substitution industrialization) schools existed. Also, radical proposals for countries to become self-reliant were quite common, see e.g., Amin (1971). The school of African Socialism was typical for these families of thoughts. It is strange that these proposals did not consider the failure of the first major experiment with AS in Ghana under Nkrumah (1958–66). It was the subject to several detailed studies such as Jones (1976) and Killick (1978). Somehow nobody took notice of the economic lesson, yet many noted the great political success of the flamboyant Nkrumah and the rhetoric he used to obtain the success.

[30]Independence was also supported by the USA, and the USSR was the last big colonial power, but somehow these facts were often overlooked.

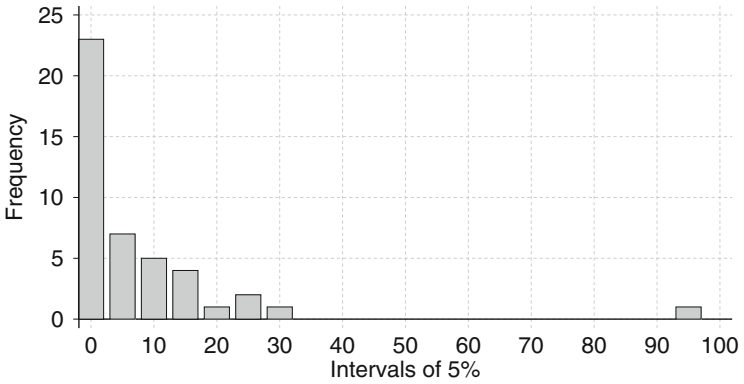

Fig. 11 The size of 44 African economies relative to the Danish economy in 1970. The column over 0 indicates that 23 countries (just above 50 %) had GDP in PPP prices of between 0 % and 5 % of the Danish. The graph for 2008 is similar. Income has grown less than Danish income, but population growth is much higher. *Source*: Maddison (2003)

(1) Enough market size. Modern firms supply goods for a considerable market. Figure 11 reports one measure of the size of the African economies in 1970 when the countries were well into the building of the AS-system. It shows that at that time 23 of the economies were below 5 % of the Danish economy in size.[31] It is difficult to believe that autarchic development of such a small markets is a rational development strategy.
(2) Enough executive capacity. The successful development of SOEs obviously put great strains on the efficiency and honesty of the public management of the firms and on the political climate in which the companies operate.[32]

The dynamics of the AS-package were not to create efficient modern industry, but as documented in the multi volume study of Bhagwati and Krueger (1973–78) and by IBRD (1995) the outcome was to create a dynamic path into inefficiency and political alliances upholding such inefficiency. That is, the path was into the rent seeking society.

[31]The reader may contemplate the development of Denmark if it has opted for autarky with high tax barriers using the tariffs to build a set of SOEs to supply the domestic market.

[32]It has been claimed that the policies of the Asian Tigers were the same as everybody else's, but worked much better as their bureaucracies were more efficient, see e.g., Rodrik (2005). This was not claimed in the 1950s and 1960s, and their policies were actually different, see Paldam (2003).

5.3 The Rent Seeking Society

Consider the dynamics of an SOE in a country with a weak administration, no critical press and a substantial level of corruption. The manager is an ex-bureaucrat appointed by the minister after due consultation with the president, who are the de facto owners. Irrespective of the charter of the company, the manager knows that he has to serve the owners, who can give his firm so much protection that it runs with a surplus. As the owners are politicians they have political agendas. The political agendas are typically to generate rents to finance the political support for the owner by employing his clients, i.e., his kin, his voters and his tribe.

So the typical dynamics of the SOEs is an upward slide into excess employment and consequently an upward slide in the level of protection. So instead of getting more ready to face international competition—as predicted by the infant industry argument—the typical African SOE grew steadily less efficient and more politicized. These dynamics seem to give an adequate explanation of the gradual strangling of the African growth in the late 1960s, and the negative growth in P2.

The oil crisis was a large price shock on the world market, and it generated an excessively large downturn in Africa. I conclude: (i) The policies had the reverse effect of the one desired, and (ii) the main reason for the size of the downswing was structural. The oil crisis was a *trigger* for a crisis that would have occurred anyhow. African countries had installed an economic system that made them unusually inflexible, when they were hit by the oil crisis.[33] My theory is, once again, simple: A rent seeking society is inflexible.

5.4 Adjusting to the New World Order After 1990

The second big strategy change was influenced by the three factors:

(1) The AS-policies, which looked promising when they started, had clearly failed.
(2) Africa had to adjust after the collapse of world socialism and the end of the Cold War.
(3) The Washington Sisters (IMF and IBRD) developed SA-support packages after 1985.

Figure 10 shows that the change from 1985 to 1990 was marginal only, but after 1990 the change became noticeable. Annual data for the period do not exist, but it is likely that the liberalizations correspond to the upswing of P3 from 1994 until now.

[33] Killick (1995) is a conference volume that discusses this theory. It asks why most LDCs adjusted better to the oil price shock than the African countries. It is, e.g., shocking to compare the reaction of the two main copper exporters Zambia and Chile to the dramatic fall in copper prices after the end of the Vietnam War.

The shifts from P1 to P2 and from P2 to P3 are both about 1 point on the Fraser index—this is a substantial change. The transition from communism in Eastern Europe amounts to about 2.5 points in the index over about 10 years. The transition from African socialism is smaller, but African socialism was more moderate than the Russian brand of socialism.

Not surprisingly it was faster to make changes in the political system—after all African countries had democratic systems de jure—so it was just a question of putting a little more content into the existing institutions, such as allowing opposition parties to run for office, and to stop having critical journalists beaten, etc.

To liberalize the economy, overstaffed SOEs must be turned into efficient companies that are viable on market conditions so that they can be sold. This policy inevitably has considerable social costs in the short to medium term. Even when the SA-loans from the Washington Sisters contained social compensations, it was not always enough. Therefore, it demands some increase in social control, and some repression may result. This might explain the seemingly contradictory trends in Fig. 8. However, it is amazing to see that as the SA-policies were implemented, growth resumed.

5.5 The Politics of Policy Change

In many ways Africa is an extreme continent as the data have shown. But this does not mean that Africans are politically extreme.

The World Values Survey has an item asking people about their preferences for private vs public ownership. It has been polled 200 times in altogether 92 countries. Unfortunately only ten of these polls are made in Africa. Bjørnskov and Paldam (2012) use these data to calculate a CS-score that measures the preferences for Capitalism/Socialism for each poll. The theoretical range is +100 for full support for capitalism to −100 for full support for socialism, but the observed range is only from +52 to −36. The average score for Africa is $+1.3 \pm 10.2$, Others are 3.5 ± 3.3, and the West is $+26.9 \pm 3.5$. Thus, the African populations are by no means extreme.[34] Maybe this explains why African Socialism was fairly moderate in most countries. Consequently, deregulation has also been moderate.

The changes showed in Fig. 10 are smaller than one should expect from the debates that have accompanied the changes. The discussion of the costs and benefits of SAs have been mixed up in the discussion of neo-liberalism of the Chicago brand. The path depicted in Fig. 10, and more anecdotal evidence, shows that the actual changes are in no way a movement between extremes. But policy changes often

[34]The CS-score rises with income, so corrected for income the African populations are rather capitalist minded. Note that the only reason that the two times the standard error is so large for Africa is the small number of observations. The standard deviations are much more similar.

take place in a dense ideological haze, and seen from afar only the haze may be visible.[35]

While most African countries have been moderate on a socialism/capitalism scale, a few countries have had periods of extreme socialist policies. It appears that these periods were rather destructive. The Fraser index does not point to countries that have pursued extreme libertarian policies, so it is not known if growth would have been faster if the SA-policies had gone further.

6 Conclusion

The economic development of 44 countries is surely complex. This paper is written on the assumption that the more the development is aggregated, the clearer the basic pattern stands out. And, in fact, a rather clear pattern does appear.

Economic development in Africa has had a cyclical path with three distinct phases: P1 from 1950 (where data start) till 1972, was a period of satisfactory growth; P2 from 1973 to 1993 had negative growth; and P3 from 1994 onwards satisfactory growth has resumed. The growth in P1 and P3 suggests that the underlying growth rate in Africa is about 2 %.

The cyclical path has been confronted with a set of 11 explanations of the development in Africa, where most are described by one variable. It is likely that some of these variables explain why the underlying long-run growth in Africa is 2 % and not, e.g., 4 %. However, most of these indictors have a path that differs from the cycle of African development. Only one fits the cycle:

The development corresponds amazingly well to the changes in the economic development strategy.[36] The African version of the ISI-policy was African Socialism that became the dominant strategy from about 1965 onwards. By the time the package was closest to implementation in the early 1970s, growth turned negative. The downturn was kick-started by the oil-crisis, but then it continued for 20 years.

In the mid 1980s the countries entered a process of deregulation—notably privatizations and foreign trade liberalization—known as structural adjustment. The process started slowly, but by 1995 it was clearly visible in the regulation index, and the process continued until 2000. This corresponds to the upswing in P3, the last phase of the cyclical path.

My interpretation is that this shows that moderate middle-of-the-road policies serve well also in Africa. The analysis suggests that Africa could have grown with about 2 % per year throughout the period. A growth rate of 2 % is too little to catch

[35]The confusion has spread to the moderate common sense rules of thumb known as the Washington Consensus (see Williamson, 1997). They have somehow been politicized so that many think that they are a recipe for libertarian policies (see, e.g., Stiglitz 2002).

[36]The changes are in accordance with the underlying political dynamics from Sect. 2.2.

up, but it is better than the growth achieved. The AS-policies aimed at a higher growth rate, but failed with tragic consequences.

With the zig-zag in the economic system, Africa still had a doubling of GDP per capita in the period from 1950 to 2011. The results presented suggest that if Africa had been economically moderate throughout, GDP per capita would have tripled.

Appendix

Table 4 The countries included in the analysis 1950 and 2010

		Population in 1000 inh.			GDP per capita, i.e., gdp		
	Countries	1950	2010	Growth	1950	2010	Growth
1	Angola	4118	13,072	1.94	1052	1721	0.82
2	Benin	1673	9059	2.86	1084	1416	0.45
3	Botswana	430	2031	2.62	349	4838	4.48
4	Burkina Faso	4376	16,242	2.21	474	1140	1.47
5	Burundi	2363	10,408	2.50	360	506	0.57
6	Cameroon	4888	19,299	2.32	671	1225	1.01
7	Cape Verde	146	588	2.34	450	2913	3.16
8	CAR	1260	5060	2.34	772	537	−0.60
9	Chad	2608	10,551	2.36	476	737	0.73
10	Comoro Islands	148	773	2.79	560	545	−0.04
11	Congo, Br	826	4124	2.72	1198	2388	1.16
12	Congo, Ki	13,569	70,943	2.80	570	263	−1.28
13	Côte d'Ivoire	2860	21,064	3.38	1041	1135	0.14
14	Djibouti	60	1038	4.87	1500	1316	−0.22
15	Equatorial Guinea	211	650	1.89	540	21,695	6.35
16	Ethiopia (& Eritrea)	21,577	93,813	2.48	390	1,023	1.62
17	Gabon	416	1545	2.21	3108	3881	0.37
18	Gambia	271	1822	3.22	607	1,083	0.97
19	Ghana	5297	24,404	2.58	1122	1777	0.77
20	Guinea	2586	10,316	2.33	303	619	1.20
21	Guinea Bissau	573	1565	1.69	289	642	1.34
22	Kenya	6121	40,081	3.18	651	1134	0.93
23	Lesotho	726	1725	1.45	355	1985	2.91
24	Liberia	824	3850	2.60	1055	1017	−0.06
25	Madagascar	4620	21,284	2.58	951	669	−0.58
26	Malawi	2817	16,213	2.96	324	775	1.46
27	Mali	3688	14,664	2.33	457	1175	1.59
28	Mauritania	1006	3205	1.95	464	1392	1.85
29	Mozambique	6250	22,060	2.12	1133	2256	1.15
30	Namibia	464	2130	2.57	2160	4,565	1.26

(continued)

Table 4 (continued)

	Countries	Population in 1000 inh.			GDP per capita, i.e., gdp		
		1950	2010	Growth	1950	2010	Growth
31	Niger	3271	15,881	2.67	617	513	−0.31
32	Nigeria	31,797	152,263	2.64	753	1643	1.31
33	Rwanda	2439	11,337	2.59	547	1066	1.12
34	São Tomé	60	143	1.47	820	1688	1.21
35	Senegal	2654	14,091	2.82	1259	1492	0.28
36	Sierra Leone	2087	4184	1.17	656	715	0.14
37	Somalia	2438	10,113	2.40	1057	998	−0.10
38	Sudan	8051	41,976	2.79	821	1554	1.07
39	Swaziland	277	1584	2.95	721	3120	2.47
40	Tanzania	7935	41,902	2.81	424	780	1.02
41	Togo	1172	6210	2.82	574	642	0.19
42	Uganda	5522	33,404	3.05	687	1080	0.76
43	Zambia	2553	12,060	2.62	661	898	0.51
44	Zimbabwe	2853	11,436	2.34	701	891	0.40

Gdp data are as explained in Table 1. Growth is in % p.a

References

Acemoglu, D., & Robinson, J. A. (2010). Why is Africa poor? *Economic History of Developing Regions, 25*, 21–50.

Aghion, P., & Durlauf, S. N. (Eds.). (2005). *Handbook of economic growth* (Vol. 1A). North-Holland, Amsterdam: Elsevier.

Amin, S. (1971). *Neo-Colonialism in West Africa (From French 1973)*. London: Penguin.

Arbache, J. S., & Page, J. (2010). How Fragile is Africa's recent growth? *Journal of African Economies, 19*, 1–24.

Aziariadis, C., & Stachurski, J. (2005). Poverty traps. Chpt 5, 295–384 in Aghion and Durlauf (2005).

Bates, R. H. (1981). *Markets and states in tropical Africa: The political basis of agricultural policies* (Several later editions). Berkeley, CA: University of California Press.

Baumol, W. J. (1986). Productivity growth, convergence, and welfare: What the long-run data show. *American Economic Review, 76*, 1072–1085.

Beny, L. N., & Cook, L. D. (2009). Metals or management? Explaining Africa's recent economic growth performance. *American Economic Review, 99*, 268–274.

Bhagwati, J., & Krueger, A. O. (1973–78). *The NBER special conference series on Foreign trade regimes and economic development*. Ballinger, Cambridge MA.

Bjørnskov, C., & Paldam, M. (2012). The spirits of capitalism and socialism. A cross-country study of ideology. *Public Choice, 150*, 469–498.

Blaug, M. (1962). *Economic theory in retrospect* (Several later editions). London: Heinemann.

Boserup, E. (1965). *The conditions of agricultural growth: The economics of Agrarian change under population pressure*. London: Allan & Unwin.

Carstensen, K., & Gundlach, E. (2006). The primacy of institutions reconsidered: Direct income effects of malaria prevalence. *World Bank Economic Review, 20*, 309–339.

Chang, H.-J. (2003). *Kicking away the ladder: How the economic and intellectual histories of capitalism have been re-written to justify neo-liberal capitalism*. London: Anthem.

CIA, World Factbook. https://www.cia.gov/library/publications/the-world-factbook/

Collier, P., & Hoeffler, A. (2004). Greed and grievances in civil war. *Oxford Economic Papers, 56*, 563–596.
Curtin, P. D. (1969). *The Atlantic slave trade. A census*. Madison, WI: University of Wisconsin Press.
de Haan, J., & Sturm, J.-E. (2000). On the relationship between economic freedom and economic growth. *European Journal of Political Economy, 16*, 215–241.
Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics, 112*, 1203–1250.
Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: How endowments influence economic development. *Journal of Monetary economics, 50*, 3–39.
Economic Freedom Index (Fraser Institute). http://www.freetheworld.com/
Fanon, F. (1961). *The wretched of the Earth* (From French, many editions). Penguin classics, London.
Freytag, A., & Paldam, M. (2011). *Comparing good and bad borrowing*. P.t. paper for the European Public Choice Conference 2011.
Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and economic development. *International Regional Science Review, 22*, 176–232.
Gourou, P. (1966). *The tropical world* (From French 1953, 4th ed.). London: Longman.
Grier, R. M. (1999). Colonial legacies and economic growth. *Public Choice, 98*, 317–335.
Gundlach, E., & Paldam, M. (2009). Farewell primacy. The political system and the economy. *European Journal of Political Economy, 25*, 340–354.
Gwartney, J., Lawson, R., & Hall, J. (2012). *Economic freedom of the world annual report* (16th ed.). Vancouver, B.C: Fraser Institute.
Harden, B. (1991). *Africa. Dispatches from a Fragile Continent*. London: Harper Collins.
IBRD (World Bank). (1995). *Bureaucrats as managers. The economics and politics of government ownership*. A World bank policy research report. New York: Oxford University Press for the IBRD.
Johnson, S., Ostry, J. D., & Subramanian, A. (2007). *The prospects for sustained growth in Africa: Benchmarking the constraints*. NBER Working Paper 13120.
Jones, C. I. (2002). *Introduction to economic growth* (2nd ed.). New York: Norton.
Jones, T. (1976). *Ghana's first republic: The pursuit of the political kingdom*. London: Methuen.
Kamarck, A. M. (1967). *The economics of African development*. New York: Praeger.
Killick, T. (1978). *Development economics in action*. London: Heinemann.
Killick, T. (Ed.). (1995). *The flexible economy*. London: Routledge (for the odi).
Krueger, A. O. (1992). *The political economy of agricultural pricing policy. A World Bank comparative study*. Baltimore, MA: Johns Hopkins University Press.
La Porta, R., Lopez-de-Silane, F., Shleifer, A., & Vishny, R. (1998). The quality of government. *Journal of Law, Economics, and Organization, 15*, 222–279.
Maddison, A. (2003). *The world economy: Historical statistics.* Paris: OECD, Updated versions until February 2010 available from Maddison homepage. http://www.ggdc.net/maddison/
Naipaul, V. S. (1975, June 26). *A new king for the Congo*. New York Review of Books (Several republications).
Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journal of Economics, 123*, 139–176.
Paldam, M. (2002a). The cross-country pattern of corruption: Economics, culture and the seesaw dynamics. *European Journal of Political Economy*, *18*, 215–240.
Paldam, M. (2002b). *Udviklingen I Rusland, Polen og Baltikum* (The development in Russia, Poland and The Baltic Countries). Aarhus: Aarhus University Press.
Paldam, M. (2003). The economic freedom of Asian Tigers. An essay on controversy. *European Journal of Political Economy, 19*, 453–477.
Paldam, M., & Gundlach, E. (2011). The democratic transition: A study of the causality between income and the Gastil Democracy index. *European Journal of Development Research, 24*, 144–168.

Paldam, M., & Gundlach, E. (2008). Two views on institutions and development: The grand transition vs the primacy of institutions. *Kyklos, 61*, 65–100.
Polity (INSCR, Center for Systemic Peace). http://www.systemicpeace.org/inscr/inscr.htm
PRIO (Peace Research Institute Oslo) conflict data. http://www.prio.no/CSCW/
PTS, Political Terror Scale. http://www.politicalterrorscale.org/
Rodney, W. (1972). *How Europe underdeveloped Africa*. Washington, DC: Howard University Press.
Rodrik, D. (2005). Growth strategies, Chapter 14. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1A, pp. 967–1014). North-Holland, Amsterdam: Elsevier.
Roeder, P. G. (2001). *Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985*. Document and report the ELF-index URL: http://weber.ucsd.edu\∼proeder\elf.htm
Stiglitz, J. E. (2002). *Globalization and its discontents*. London: Allan Lane.
UNCTAD (United Nations Conference on Trade and Development). Statistics at URL: http://www.unctad.org
WDI, World Development Indicators (World Bank), URL: http://databank.worldbank.org/ddp/home.do
Williamson, J. (1997). The Washington consensus revisited. Followed by comments from F. Steward, B., Persaud and T., Yanaghara. In L. Emmerij (Ed.), *Economic and social development into the XXI Century* (pp. 48–80). Washington, DC: Johns Hopkins University Press for the Interamerican Development Bank.

Developing Globalization: How Japanese ODA Created Channels for Globalization Processes, 1954–2012

Annette Skovsted Hansen

Abstract This paper is the first qualitative micro case study of one aspect of globalization: personal networks as a concrete outcome of development assistance spending. The empirical findings related in this paper present circumstantial evidence that Japanese foreign aid has contributed to globalization over the past 60 years by establishing channels for interactions, exchanges, interdependencies, and the circulation of persons. Notably, it analyzes the networks that have been deliberately encouraged by Japanese aid among alumni from Japanese training courses since 1959. The analyses are based on alumni newsletters since 1961, alumni societies since 1964, and alumni conventions since 1986 that distinguish Japanese aid programs by being formal and permanent channels of communication. This is a typical positive externality that is difficult to capture in the usual aid effectiveness studies, because it works in the longer run and in ways that are unrelated to individual projects. The main argument is that the formal channels have provided the alumni with a network, which they can choose to activate for concrete impact.

1 Introduction

This qualitative micro study can be read as a case of growth productivity not only to be measured in production as suggested by Andersen and Dalgaard in their quantitative study on flows of people (Andersen and Dalgaard 2011). The empirical findings related in this paper present circumstantial evidence that Japanese Official Development Assistance (ODA) has contributed to globalization over the past 60 years by establishing channels for interactions, exchanges, interdependences, and the circulation of persons. The contributions to globalization mapped in this paper include channels for interaction such as the training courses since 1959, the alumni newsletters since 1961, alumni societies since 1964, and alumni conventions since 1986 that distinguish Japanese ODA programs by being formal and permanent

A.S. Hansen (✉)
Department of Global Studies, Aarhus University, Aarhus C, Denmark
e-mail: ostash@cas.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_21

channels of communication. The formal channels have provided the alumni with a network, which many have chosen to activate for concrete impact.

Millions of people from more than 100 countries have attended courses offered as part of multilateral and bilateral technical assistance programs in different countries. Among them alumni from courses in Japan have since the 1960s established alumni associations, which have actively prepared new course participants, followed-up on and supported alumni efforts to implement course content at their home institutions. In the case of Japanese foreign aid, the distribution of ODA has been explicitly focused on infrastructure projects to ensure safe transportation of goods in preparation for attracting Foreign Direct Investment (FDI) and enhancing international trade between Japan and developing countries (Rix 1980, 1989; Islam 1991; Orr 1990; Arase 1995, 2005; Aoki 1998). As importantly for globalization, I argue, has been the alumni network of ex-participants in regional and Japan-based training courses financed by Japanese ODA. The empirical focus of the paper will be the Japanese ODA distributed through the Japanese Ministry of Economy, Trade and Industry; METI (formerly Ministry of International Trade and Industry (MITI)) to the Overseas Human Resources and Industry Development Association (HIDA) (from 1959 until 2012, The Association for Overseas Technical Scholarship (AOTS)) for the economic development of the private sector of developing countries through human resource development, since 1959.

The critiques of technical assistance and overseas training courses funded by ODA seem to overlook a very important byproduct of the courses i.e. the personal networks initiated during the course participation and expanded through alumni activities over time. The networks are connected to the host countries, but also encourage south-south exchanges. Although the different host countries have been more or less formal in encouraging alumni activities and exchanges, the individual alumni are connected to alumni from other host countries in many cases. Personal networks are biased according to the mechanism of preferential attachment to generate random scale-free networks (Barabási and Albert 1999), where the host countries selected course participants and trainees among already well-connected elites from developing countries. Basically, hosts prefer elites, because they have a better chance of ensuring impact outcome and provide a more valuable connection to their home country for the host (Levitt 1998: 939).

Historically, the host countries build onto personal networks established through colonial structures of education favoring local elites and people selected for their usefulness for the colonial administration. Similarly, also the trade networks build on the ties of old mercantile hierarchies. Globalization processes depend on long term structures and ties, however, I will still maintain that ODA formalized channels of communication and personal encounters encompass, a wider range of people, who represent a larger number of localities and social backgrounds than other mechanisms save perhaps labor and refugee migration (Blossier 2010; Levitt and Lamba-Nieves 2011; Levitt 1998; Vogt 2012). Therefore, I find it important to inspire more research in the area of networks engendered by overseas training and technical assistance courses funded by ODA.

Impact of personal networks is a typical positive externality that is difficult to capture in the usual aid effectiveness studies (Doucouliagos and Paldam 2009), because it works in the longer term and in ways that is unrelated to individual projects. Furthermore, the study falls into the category of micro-studies seeing positive effects of aid as first mentioned as the 'micro-macro paradox' of aid evaluation studies by Mosley (1987) and cited in McGivillivray et al. (2006: 1031; 1045). McGillivray et al. illustrate the effect on the aid discourse of the World Bank study *Assessing Aid: What Works, What Doesn't, and Why* from 1998 and its conclusion that general aid effectiveness depends on a sound policy environment in the recipient country (2006: 1037; 1041; World Bank 1998). Preferential attachment, colonial legacies, and the process of building an elite over time is hard to measure as also pointed out in the OECD report *New Sources of Growth* (2013: 10). Here I assume that the OECD study will constitute a similar break as *Assessing Aid* by emphasizing the importance of including factors difficult to measure in analyses of impact. This study cannot fully measure the effect, but it can document the size of the efforts of specific networks created.

In the late 1940s, the Danish civil engineer Jørgen Saxild whom the Danish Prime Minister had appointed to head the new Styrelse for teknisk samarbejde med udviklingslandene [Board for Technical Cooperation with Developing Countries] insisted that it was important for Denmark to invite professionals from developing countries to Denmark to attend training courses. He argued that this was necessary for Denmark to establish a network of people with goodwill in order for Denmark to win contracts in developing countries. Saxild ensured that Denmark hosted a series of UN courses on cooperative farming and public administration already from 1953 as the Danish contribution to the United Nations. Since 1962, Denmark has also offered courses bilaterally through systems by different names such as technical assistance, Danida Fellowship Center (1990–), and Building Stronger Universities (2010–). The alumni from these courses are connected to Denmark through informal networks, which are difficult to trace and almost invisible even to the alumni themselves. This has been recognized as a challenge by alumni and Danida Fellowship Centre officials for a long time and on August 20, 2014, the first alumni network of the Building Stronger Universities program will be inaugurated 50 years after the first alumni society established by alumni from training courses in Japan. In the new millennium, also German and French development agencies have begun to retrace alumni from courses hosted by their countries. USA based universities and foundations such as the Ford Foundation, the Carnegie Foundation, and the Rockefeller Foundation have separate alumni associations. In the case of Japan, formal alumni networks have been part of the training program design continuously since the 1960s.

Various Japanese ministries organized training, technical and management courses, beginning with courses offered through the Colombo Plan in 1954. From the early 1960s, Japanese agencies and organizations provided a formal structure for alumni to (re)connect with each other and with Japanese stakeholders through the distribution of newsletters including readers columns and by encouraging the establishment of alumni societies based in the home country of the alumni, and

holding international alumni conventions. The Japanese agencies and organizations argued from the early 1960s that formal structures provided transparency for building friendships throughout the world. This paper will explore the set-up of the structures and some examples of the historical impact on globalization processes of the Japanese Association for Overseas Technical Scholarship (AOTS/now HIDA-AOTS) network. Drawing on experiences from the colonial empires of educating and training an elite in the colonies to qualify them to enter the colonial administration or a trading network (Burton 1998, Hall 2002, and Ogborn 2008), the postwar launching of ODA financed scholarship and fellowship training programs allowed for developing a network including a growing elite in developing countries, which would become central to processes of globalization.

Aid effectiveness studies (Doucouliagos and Paldam 2009) largely employ methods that will not detect the impact of the personal networks introduced here. One of the challenges of this investigation is the measurability of impact of knowledge based capital (KBC) as noted in the OECD report (2013). This paper is mainly based on readings of newsletters, surveys, and evaluations published by AOTS, as well as: a web-based questionnaire distributed to 1500 AOTS alumni in November 2012 and entries in the AOTS facebook group following the March 11, 2011 earthquake in Japan. The sources hold significant bias, because they are all representing a favorable interest in AOTS. Nevertheless, my interest in this material stems from a desire to identify the stories and documentation AOTS has accumulated over a 50 year period in support of their training program and how this ties in with my own observations of the network, which AOTS has engendered among its alumni.

In order to juxtapose Foreign Direct Investment (FDI) and the personal networks of alumni from AOTS training courses, the paper begins with an overview of recent arguments concerning the connection between ODA and FDI followed by a discussion of the expected economic return and expected benefits exemplified by individual stories. In the next section, I present considerations of economic growth tied to alumni activities and global connections of the alumni by mapping the structure and individual cases of use of the structure of the AOTS alumni community. After this follows a section on the way in which the AOTS facebook group and the AOTS alumni newsletter played a role in the wake of the Triple Catastrophe in March 2011, as an illustration of concrete uses of the global channels of communication and money flows.

2 Economic Return on Investments of Money and Time in Training Courses

The expected benefit of the courses include economic return as a result of effectivization through waste reduction, the potential of an increase in individual connections to a global market through the alumni network, and an increased focus

on customer satisfaction after the introduction of improved quality control. A direct causal relationship between aid and growth has proven very difficult to illustrate and Jensen and Paldam (2003) as well as Burke et al. among others conclude that ODA has not brought about economic growth. However, Burke et al. fail to address the fact that the economic growth in their Asian sample took off right after the bilateral aid programs not least with Japanese ODA began in the first years of the 1960s (Burke and Ahmadi-Esfahani 2006: 351). The points about other factors such as effective policies, more non-aid loans, and FDI as the real impetus for growth neglect the variable situation where these other factors are kept constant, but with no aid. McGillivray et al. (2006: 1045) emphasize this point, when they claim that studies of aid and growth, since the World Bank's *Assessing Aid: What Works, What Doesn't, and Why* published in 1998, find that "aid works to the extent that in its absence, growth would be lower." (McGillivray et al. 2006: 1031) Another point made by McGivillivray et al. (2006: 1035) is that "[a] positive and statistically significant impact of foreign aid was found when the sample is restricted to the poorest 30 countries sample and aid is lagged 5 years." The fact that the impact only shows after a 5 year lag suggests the necessity of longitudinal studies for evaluating any effects of aid.

Several studies have concluded that ODA attract FDI (Blaise 2005; Kimura and Todo 2009; Bhavan et al. 2011: 143). In their quantitative study to compare trajectories for aid to Sub-Saharan Africa and South East Asia, Carbonnier et al. show proportions of ODA and FDI in select countries in the two regions comparing FDI net inflows as percentage of GDP and net official development assistance received as percentage of GNI (Carbonnier et al. 2011: figure 18) without discussing how these numbers can be compared, but for this author more problematic is the lack of attention to changes in these numbers and their relative size over time. A longitudinal study would shed light on a possible correlation or sequence in the relationship between ODA and FDI as claimed by Japanese ODA literature (Islam, 1991). The nature or type of ODA may or may not play a significant role (Kasuga 2007). Aid towards transportation and human infrastructure may be specifically designed to prepare for and attract FDI, either because FDI is an advantageous investment for the donor or because the donor believes that economic growth depends on or is not likely to occur until a country can attract or rely on FDI. AOTS designed its training program for the development of the private sector in developing countries already in 1959 to prepare alumni for a global market by offering technical and management training. Included in many of the training courses have been components on quality control, waste reduction, and production efficiency to make the home companies and countries of the trainees attractive for FDI and export goods. It is important to note that HIDA/AOTS selects a large proportion of the course participants upon request from Japanese companies or sub-contractors that are products of FDI, however, the continued ODA investment in the courses and the alumni also attracts new Japanese FDI to the home countries and companies of the alumni.

2.1 A Cost-Benefit Analysis

AOTS quantifies the expected return on the courses in the evaluation report for 2011 activities (AOTS 2012). The report includes a section on economic effects of AOTS training courses measured as expected return. The investment of money and time is multiplied by 2.22 in the case of technical training and 2.72 for management training, because no company, organization, or ministry is expected to invest without reasonable guarantee of some return. The calculations are based on expectations of return on the cost of the training. The statistical material is based on a survey of 1103 respondents: 383 on technical training and 493 on management training. The average training period is 160.4 days for technical trainees and 12.9 days for management trainees. For technical training the average company's costs is 1,901,000 yen including opportunity costs and excluding ODA subsidies (AOTS 2012: 42). It is important to note that around 50 % of the total costs of the training are subsidized by ODA in order to give incentive to more Japanese companies to offer on-the-job training.

The expected return on technical training is 2.22 times, which gives expected direct economic effects as 1,901,000 yen × 2.22 = 4,220,000 yen. In the Japanese financial year of 2011, the number of technical training trainees was 1288 persons, which gave expected total direct economic effects of 4,220,000 yen × 1288 persons = 5,435,360,000 yen (AOTS 2012:43). The direct economic effects of management training, therefore, became 640,000 yen × 2.72 = 1,740,000 yen with the number of trainees being 1326 persons for FY 2011 and the total direct economic effects of management training calculated as 640,000 yen × 1326 persons = 2,385,700,000 yen (AOTS 2012:43). In the report, AOTS states that these preliminary calculations cannot indicate whether the actual return would be higher or lower than these estimates. They conclude that two surveys must be conducted one prior to and the other after the training (AOTS 2012:44; 46).

2.2 Individual Evaluations of Benefits

In attempts to map actual returns on the training, individual alumni testimonies emphasize the indirect or direct influence of their international contacts on the economic return of their training as reflected in the increase in number of trading partners in Japan and other countries for alumni home institutions. This author's question to HIDA/AOTS about measured economic return has inspired the formulation of a web-based questionnaire with 12 questions in order to collect data for this paper.[1] Using a random sampling, HIDA distributed the questionnaire electronically

[1] HIDA owns the rights to the data collected between November 1 and November 18, 2012, from this web-based questionnaire, but because the author of this paper initiated its formulation and distribution, she has been granted the rights to access and use the data.

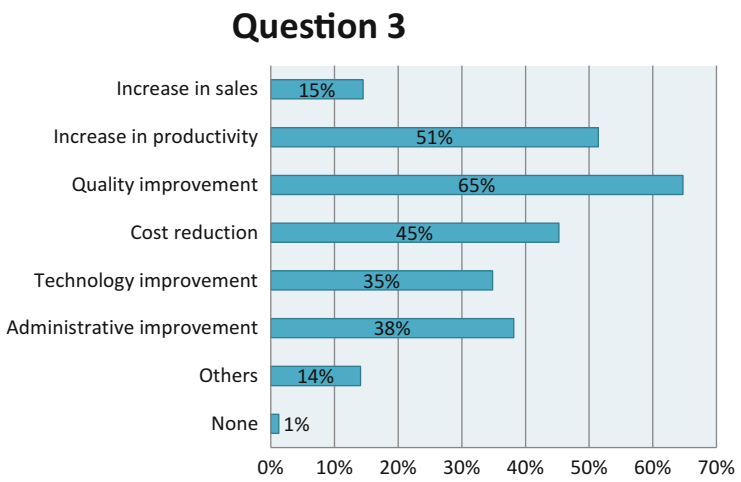

Fig. 1 Question 3—After returning from AOTS training in Japan, what positive contribution did you make to the company's performance? (Please tick all that apply). *Source*: HIDA Questionnaire November 2012

to 1500 alumni in 66 countries, 500 alumni who had attended training courses in Japan during each of the years 2009, 2010, and 2011. HIDA chose the 3-year time span to limit the influence of other variables and it corresponds with the 2.7 year mentioned in the OECD report (2012: 5). The short time span, however, means that the 2011 trainees report on few measurable changes yet. 241 alumni from 25 countries responded and the analyses of the responses have to take the high degree of self reporting into consideration. The responses, however, are some of the only indicators we have on the concrete returns on the training.

The responses to the questions of main interest to this study are shown in Figs. 1, 2, 3, and 4. They can be grouped in two: (1) the improved production that makes the home country attractive for FDI and export goods, because of infrastructure and improved quality of products, and (2) the increase in numbers of countries, with which the company trades. In response to questions not shown here, about 30 % of the respondents report no changes on these two measurements, whereas more than 50 % report personal influence on increased productivity and 65 % on quality improvement as shown in Fig. 1.

According to Fig. 2, 75 % of the respondents ascribes AOTS training a considerable or very considerable role in enabling their positive contributions after their return. The answers to question four may, more than any of the others, be seriously flawed by the fact that the respondents felt obligated to AOTS and had an interest in future training or Japanese FDI. Nevertheless, only 22 % answered 'very considerably' leaving room for individual influence, other inspiration, and other factors.

Figure 3 shows that 25 % of the respondents reports considerable increase in the number of trading partners in Japan.

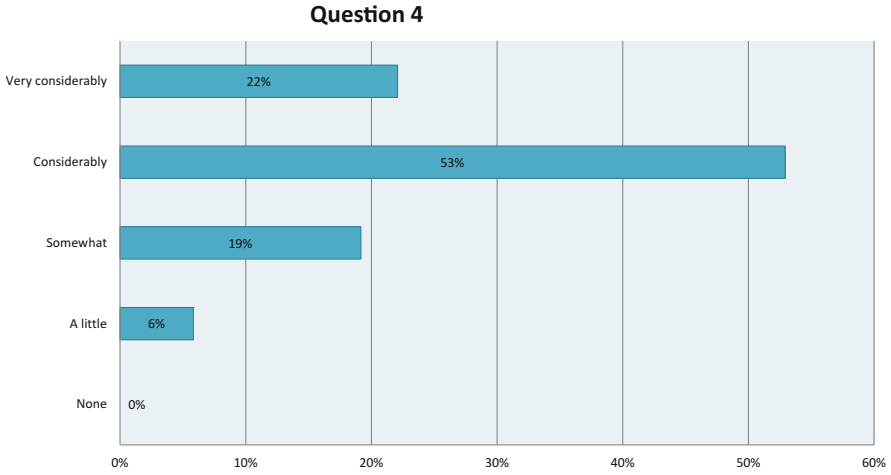

Fig. 2 Question 4—To what degree has your experience of participating in AOTS training in Japan had an impact on the above Q3's contributions to the company's performance? (Please tick one). *Source*: HIDA Questionnaire November 2012

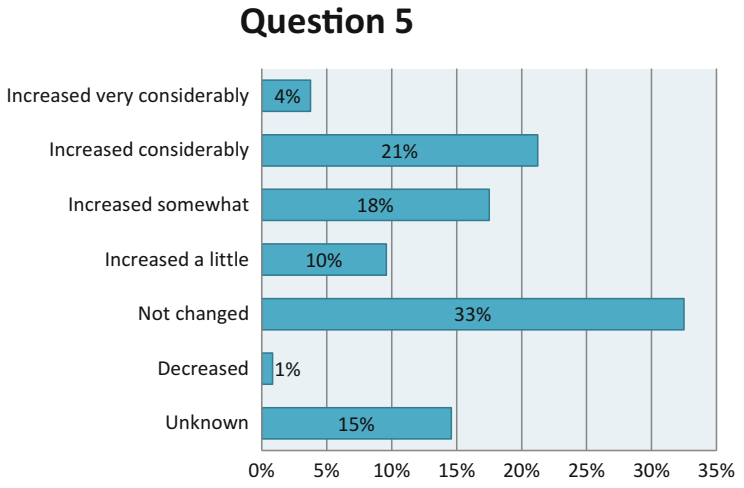

Fig. 3 Question 5—Did the company you work for increase or decrease its trade with Japanese companies after your returning from AOTS training in Japan? (Please tick one). *Source*: HIDA Questionnaire November 2012

Several respondents added comments about how Japanese companies expressed an interest in cooperation with their home companies after their training in Japan. Respondent number 119 wrote

> My company had no experience or awareness of work culture of Japanese companies. We had in the past never interacted nor had we supplied any products to Japanese companies.

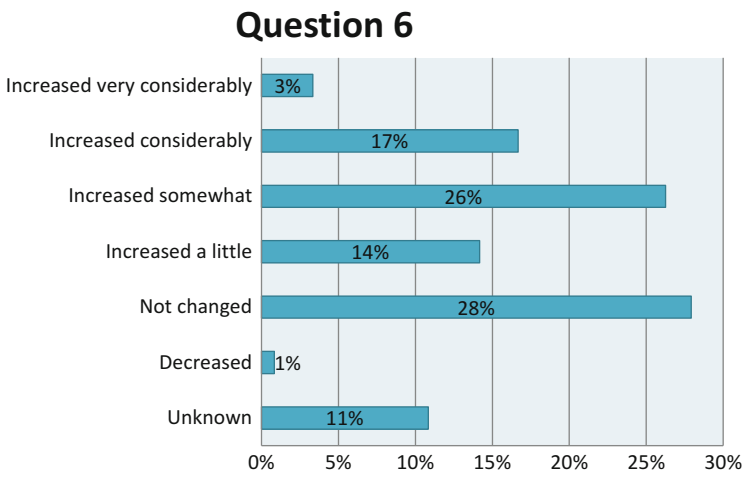

Fig. 4 Question 6—Did the company you work for increase or decrease its trade with other foreign companies after your returning from AOTS training in Japan? (Please tick one). *Source*: HIDA Questionnaire November 2012

> Concentration was always towards European and North American companies. Fortunately, coinciding with AOTS training, a few Japanese companies with presence in India expressed an interest in working with my company. (HIDA Questionnaire November 2012)

60 % of the home companies of the alumni have recorded an increase in the number of countries with which they trade after their return (See Fig. 4).

The more global reach of the training outcome is also accentuated in the so-called success story of Dhingra, who maintained that his company increased the number of countries to which RIL-HZ exports to 80 countries after his return from training, because of increased awareness and focus on customer satisfaction (AOTS 2009: 13). In the comments, respondents emphasize efficiency increase as for example waste reduction, quality improvement as major contributions to significant production increases in their companies after their return from training.

Among indicators of economic gain from training, a series of so-called success stories submitted in connection with the 50th anniversary of AOTS in October 2009 include concrete numbers on economic return of training as local measures of an increase in production and income. The success stories were submitted in three categories: technology transfer, cultural exchange, and social contribution. The stories about technical transfer and social contribution included numbers on economic gain, production increase, and global connections. The Ghanaian case focused on the outcome of six managers' participation in the same training program. GRATIS Foundation Ghana went from being a donor dependent organization to a predominantly self-financed institution. Also the example of Love Green Nepal (LGN), an NGO founded by an AOTS alumni, Amira Dali, shows an increase in numbers of local women who produce export products for a global market (AOTS

2009: 60). Most of the alumni mention in passing the role of their alumni network for the positive and measurable outcome of their training.

2.3 Alumni Networks

The extra cost of network creation is small relative to the costs of the training program. Japanese agencies and associations have established formal channels for an alumni network since 1961 to ensure sustainable economic and other return on the investment in training. They expect alumni networks to enable alumni to access global markets of commerce and education through their formal connections to other alumni from the South and to Japanese hosts and other stakeholders. The AOTS evaluation report of 2011 published in March 2012 mentions globalization, history, and networks as an expected and desired outcome of the training program. The building of personal networks through the AOTS training program begins at one of the AOTS training centers. 67.8 % (799 people) of the respondents to a survey on their impression of the experience of living at an AOTS training Center answer that it offered a good opportunity for exchanges with trainees from other countries as compared to its value for interacting with people from the home country or from Japan with 56.6 and 55.9 % answering in the affirmative, respectively (AOTS 2012: 41). One respondent to the November 2012 questionnaire wrote "The AOTS program created a very useful network of more than 20 entrepreneurs from India who attended the program. We very regularly interact with each other by sharing knowledge as well as doing business with each other. I now have friends across India spread across several cities in India, thanks to AOTS!" (HIDA Questionnaire November 2012)

Since 1961, alumni from AOTS training courses have received a newsletter four times a year with articles on Japanese technology and culture and a readers' column with messages and photos from fellow AOTS alumni representing countries in Asia, Latin America, Africa, and Eastern Europe. All issues of the newsletter published since 1965, when alumni from Singapore established the first AOTS alumni society, include articles on different national alumni association activities. Today there are 71 alumni societies in 43 different countries. In order to encourage South-South alumni interaction, the alumni societies agreed to hold the first international AOTS Alumni Convention in 1986, which has since been followed by more and the decision to establish the World Network of Friendship (WNF) in 1992. WNF is a network facilitating South-South activities organized, promoted, and sponsored by course alumni and their networks. Since 2009, AOTS has also administered a Facebook Group and distributed an e-newsletter. In 2012, AOTS and JODC joined under a new name The Overseas Human Resources and Industry Development Association (HIDA), so the –e-newsletter is now called *Global Interface Japan Newsletter* (HIDA 2012–2016) and includes an "Introduction to a former HIDA/AOTS ex-participant's employer corner,' which "introduces employers who

wishes to do business globally" (*Global Interface Japan Newsletter*). This contact information is for contact between alumni who have not met in person previously.

Alumni have used their networks and their accomplishments gained from training in many different ways. Here I will focus on careers, education, South-South relations, connecting to thematic networks, and economic gain locally or through increased export. Many alumni have made career shifts inspired by their training even if not immediately upon return, but rather when they have proven their ability to implement their learning from the training. Some change careers locally from government employment to private sector, such as Ranganathan in 1992, 20 years after his return to Tamilnadu, India (AOTS 2012: 47). The shifts in career or work relations have also involved links to international organizations such as Darfoor from Ghana working for EU, Divekar from India contributing to and organizing UN events, Ranaganathan cooperate with NORAD (AOTS 2012: 45). The Nepalese woman Amira Dali's sister upon return from Japan gained employment with the Danish Dairy Board that sent her to Aarhus, Denmark as a Danida Fellow on a technical assistance course on dairy. This later qualified her for employment as manager in a chocolate factory in Seattle, which still purchases some product from the Danish dairy industry, thus spanning three countries besides her own on three continents.

Dali activated her AOTS network in connection with her daughter's education and has inspired all her siblings to take training courses in Japan. Many AOTS alumni join international and global thematic networks including large corporations such as the Denso Corporation, which operates in 32 different countries (AOTS 2009 (Songtham Hattha): 6), the art form of Ikebana, which holds international conferences and competitions throughout the world, where some alumni such as Ms. Divekar has held workshops in the US and won prizes. Other alumni use their AOTS training and network to work to combat global pandemics such as HIV/AIDS (Dhingra (AOTS 2009: 17) or address gender and environmental issues (Dali (AOTS 2009: 60–61)). These are just a few personal stories to illustrate the diversity of links fostered by the alumni after their attendance at AOTS training courses.

3 The Triple Catastrophe

In one particular case the connection to Japan inspired alumni to reverse the flow of money from Japan. The March 2011 triple catastrophe in Japan: earthquake, tsunami, and the Fukushima nuclear power plant disaster exemplified concrete financial flows as voluntary altruistic payback of ODA. March 11, 2011 a major earthquake off the coast of Japan caused a severe tsunami that ravaged the northeastern coast of Japan, and also severely damaged the Fukushima nuclear power plant. Within minutes the first expressions of shock and sympathy from AOTS alumni appeared on the AOTS Facebook site. Many entries followed from concerned alumni who offered prayers, notes of deep sympathy, and promises of support. The immediacy of responses reflected the availability of this channel of

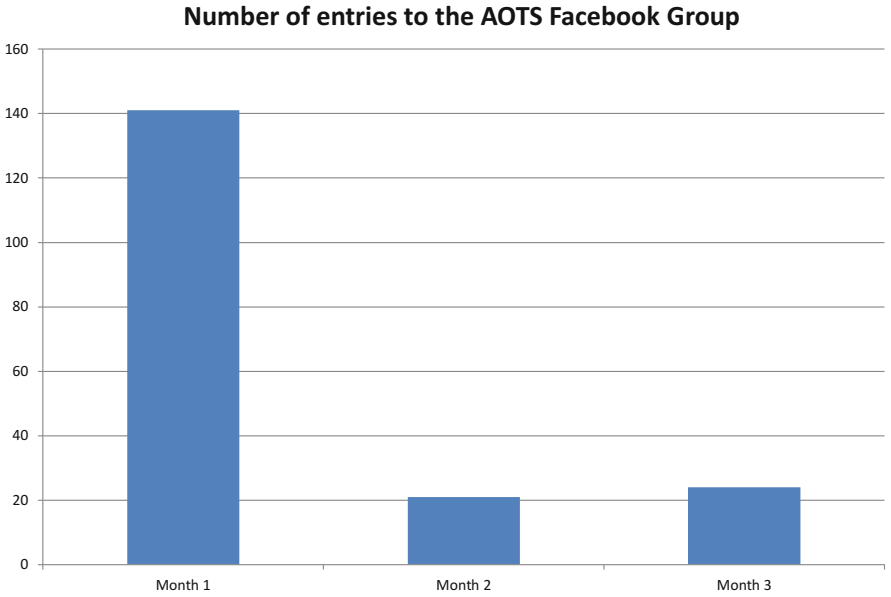

Fig. 5 Number of entries to the AOTS Facebook Group

communication directly to the Japanese host and the awareness of the concern of other alumni throughout the world, who had shared similar experiences in Japan. Figure 5 shows the total number of entries over a three month period following the date of the earthquake.

In Fig. 6, the proportional geographic distribution is divided into three times 1 month time periods, which confirm the consistent Asian majority of entries, but shows a proportional shift with no Middle Eastern or European entries in the two subsequent months and Latin American comments representing more than 30 % and Asia only just over 60 % in the third month.

In the AOTS newsletter a public acknowledgement of gratitude appeared in September 2011:

> **Donations for the Great East Japan Earthquake victims**
>
> AOTS Alumni Societies, AOTS ex-participants and related organizations generously raised 18.6 million yen (as of 12 August 2011) in donations for victims of the Great East Japan Earthquake which was occurred on 11 March 2011. Also, countless words of sympathy and encouragement were sent to AOTS from all corners of the world. A million thanks for their heartfelt messages and donations. (E-Newsletter, Issue 1, 29 September 2011)

The editors of the newsletters and the AOTS representative writing to the AOTS facebook group emphasized the financial contributions from the alumni to Japan and introduced an element of competition between the alumni societies by making the amounts and donators public. As shown in Tables 1 and 2, AOTS made a distinction between the amounts the AOTS alumni societies donated through AOTS in Japan

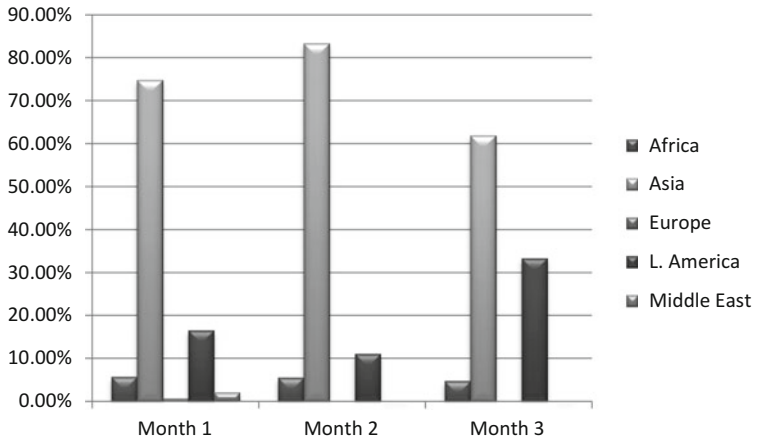

Fig. 6 Proportional geographic distribution originating country of entries to the AOTS Facebook Group by month after the Triple Catastrophe March 11, 2011

Table 1 Alumni Societies donated JPY 13,000,000 (as of August 2011) through AOTS

	Country	Donator	Amount in JPY
1	Sri Lanka	Japan Sri Lanka Technical and Cultural Association (JASTECA)	4 million Yen
2	Thailand	ABK & AOTS Alumni Association (Thailand)	2 million Yen
3	Pakistan	AOTS-ABK Dosokai Lahore Regional Centre	0.7 million Yen
4	Peru	Kenshu Kiokay del Peru	0.5 million Yen
5	China	AOTS Alumni Societies of China	0.5 million Yen
6	India	AOTS Alumni Association of Western India	Below 0.5 million Yen
7	Indonesia	Himpunan Alumni AOTS Indonesia (HAAI)—Jakarta Central Board	Below 0.5 million Yen
8	Malaysia	Persatuan Alumni AOTS Malaysia	Below 0.5 million Yen
9	Nepal	Nepal AOTS Alumni Society (NAAS)	Below 0.5 million Yen
10	Mexico	Ex-Becarios AOTS México-Japón A.C.	Below 0.5 million Yen
11	India	AOTS Alumni Society of Central India (AASCI)	Below 0.5 million Yen
12	Mongolia	AOTS Alumni Society of Mongolia	Below 0.5 million Yen
13	Czech Republic	Czech Alumni Association	Below 0.5 million Yen
14	Bangladesh	Chattagram AOTS Alumni Society (CAAS)	Below 0.5 million Yen
15	India	AOTS Alumni Society, Delhi	Below 0.5 million Yen
16	Paraguay	La Asociación Alumni AOTS de Paragua	Below 0.5 million Yen

Source: *Kenshū* (2011): 4

Table 2 Alumni Societies donated JPY 6,000,000 (as of August 2011) through the Embassy of Japan in their respective countries

	Country	Donator	Amount in JPY
1	Bangladesh	Bangladesh AOTS Alumni Society (BAAS)	2.2 million Yen
2	Vietnam	AOTS Vietnam Alumni Society (AVAS)	2.2 million Yen
3	Singapore	AOTS Alumni Society of Singapore	0.7 million Yen
4	Argentina	Asociación Argentina de Ex-becarios de la ABK y AOTS	Below 0.5 million Yen
5	Ethiopia	Dashen Fujiyama Society	Below 0.5 million Yen
6	India	Alumni Society of AOTS (Kerala (ASA Kerala)	Below 0.5 million Yen
7	India	ABK-AOTS Dosokai Tamil Nadu Centre	Below 0.5 million Yen
8	Kenya	The AOTS Alumni Society of Kenya	Below 0.5 million Yen

Source: *Kenshū* (2011): 4

(Table 1) and through the Japanese embassies in their respective countries (Table 2). This distinction illustrates how most of the combined contributions went through AOTS in Japan and, thereby, showed an obligation to Japan through AOTS and an identification of Japan with AOTS.

Furthermore, The Philippine Cultural and Technical Association of Returned Overseas Scholars (PHILCULTAROS) chose to donate 0.4 million Yen through the Miyagi International Association (*Kenshū* 2011: 4).

Along with the list of donators and the amounts of donations, the newsletter included a dialogue between the AOTS President, Kaneko, and the vice-patron of JESTECA, Dayasiri Warnakulasooriya, who visited AOTS Tokyo Kenshū Center when he was in Japan on business after the triple disaster. He was quoted as saying

> I am proud to hear that the donations will be used to help those in need in the quake-hit areas. As you know, Sri Lanka has received a great deal of physical and material assistance from Japan over the years, including ODA. More than one year has passed since the end of our many years of civil war, and I remember that Japan continued to provide assistance throughout the entire conflict, even during the more difficult times. AOTS provided one form of this assistance through the technical training and management training programs. (*Kenshū* 2011: 2)

During the dialogue Dayasiri confirmed that the donation was due to the friendly relationship between Japan and Sri Lanka and continued to explain how JASTECA hoped to attract more FDI from Japan not least in light of the large number of Chinese and South Korean companies investing in Sri Lanka and inviting Sri Lankans to participate in training courses in South Korea (*Kenshū* 2011: 3). This shows how the Japanese approach to training financed by ODA and preparing for FDI has inspired former Chinese and South Korean trainees in Japan to replicate the practices as they have seen the benefits to tying into global channels of communication through trainee alumni networks.

The triple catastrophe initially shows an outcome in terms of a network where all are tied to Japan, the global aspect appears when we read Dayasiri's concern about South Korea and China as Japan's FDI competitors in Sri Lanka. He is concerned because he has invested in the Japan connection, but has to compete on a playing field with many more actors. Dayasiri in this context refers specifically to Asian actors. However, in the case of most countries the array of foreign interest comes from many countries. In the case of Tanzania, the neighbor of one alumnus from GTZ (German) courses is an AOTS alumnus and they have friends and colleagues who have attended courses in Denmark, the US or other countries and, thereby, mutually link networks centered around specific metropolis. Alumni from courses in different countries exchange views on their experiences and draw on each other for recommendations, connections, and best practices for applying for funding for projects. They act as weak ties making a small world of professionals in the developing world (Granovetter 1983).

The point of this section was to illustrate, on the one hand, how the alumni consider themselves part of a global community with ties to Japan and, on the other hand, how concrete flows of money follow the personal ties of alumni to Japan through their Japanese host organization. Here the effect of the regional cooperation and the competition between regions also become clear.

4 Conclusion

The contributions to globalization mapped in this paper include channels for interaction such as the training courses since 1959, the alumni newsletters since 1961, alumni societies since 1965, and alumni conventions since 1980 that distinguish themselves by being formal and permanent channels of communication. The formal channels have provided the alumni with a network, which individual alumni have chosen to activate for concrete impact. Some concrete outcomes presented as a series of individual experiences include exchanges among alumni within their home country or between different countries in the South, as well as with Japanese stakeholders, career shifts, involvement in global thematic networks, and export increase. Interdependencies are manifold and the concrete example presented was the case of the triple catastrophe of March 2011 in Japan. This illustrates how money flows not only in trade, but also in humanitarian contexts can flow back from developing countries to a former host country, reflecting perhaps a sense of obligation and perhaps a personal sentiment or connectedness to a country of temporary residence.

The material mainly offers evidence of ties between alumni and Japan and gains for Japanese companies. However, there is also evidence that alumni have replicated the channels of communication in their own ODA programs, thereby making the implications of the Japanese practice global. Alumni activate their networks for very different and multiple purposes, which have spurred long-term results of economic gain, social mobility, and involvement in global issues for themselves and their home

communities and countries. The prominence of evidence from the Japan related alumni network is due to the transparency of the formal channels of inter-alumni communication initiated by Japanese stakeholders. The transparency of the Japanese case allows for an insight into the potential and concrete influence of ODA on economic and cultural globalization processes.

Acknowledgments I thank the Overseas Human Resources and Industry Development Association (HIDA) for access to data used in this study. Previous versions of this paper has been presented at the American Historical Association Annual Meeting 2009; the conference Globalization: Strategies and Effects in 2011; at the HIDA Management Performance Evaluation Committee Meeting on November 7, 2012, and at a seminar at the Japanese International Cooperation Agency (JICA) Research Institute on November 8, 2012. I thank participants at these presentations, Martin Paldam, an anonymous referee, and the editors of this volume for invaluable comments. Remaining shortcomings are solely my responsibility.

References

Andersen, T., & Dalgaard, C. (2011). Flows of people, flows of ideas, and the inequality of nations. *Journal of Economic Growth, 16*(1), 1–32.

Aoki, T. (1998). *Kaihatsu enjo ron* [On development aid]. Tokyo: Gakubunsha.

AOTS. (2009, October 27). *AOTS success story convention: "Investing in People for a Better World"*.

AOTS. (2012). *Keizaisangyōjinzaiikuseishienkenshūjigyou: 2011nendojigyōhyōkajōkokusho* [Training Program for Human Resource Development in Economy and Industry: Evaluation Report for 2011 Activities]. Tokyo: AOTS.

Arase, D. (1995). *Buying power: The political economy of Japan's Foreign Aid*. Boulder, CO: Lynne Rienner.

Arase, D. (Ed.). (2005). *Japan's Foreign Aid: Old continuities and new directions*. Routledge Contemporary Japan Series. London: Routledge, Taylor and Francis Group.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Bhavan, T., Xu, C., & Zhong, C. (2011). The relationship between foreign aid and FDI in South Asian Economies. *International Journal of Economics and Finance, 3*(2), 143–149.

Blaise, S. (2005). On the link between Japanese ODA and FDI in China: A microeconomic evaluation using conditional logit analysis. *Applied Economics, 37*(1), 51–55.

Blossier, F. (2010). Migradollars and economic development characterizing the impact of remittances in Latin America. *Washington Report on the Hemisphere*, *30*(10): 1, 6–7.

Burke, P. J., & Ahmadi-Esfahani, F. Z. (2006). Aid and growth: A study of South East Asia. *Journal of Asian Economics, 17*, 350–362.

Burton, A. (1998). *At the heart of empire: Indians and the colonial encounter in Late Victorian Britain*. Berkeley, CA: University of California Press.

Carbonnier, G., Chakraborty, P., & Mulle, E. D. (2011). *Asian and African development trajectories* revisiting facts and figures. Revue internationale de politique de développement. Retrieved August 4, 2014, from http://poldev.revues.org/682

Doucouliagos, H., & Paldam, M. (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys, 23*(3), 433–461.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory, 1*, 201–233.

Hall, C. (2002). *Civilising subjects: Metropole and colony in the english imagination 1830-1867*. Cambridge and Oxford, UK: Polity.
HIDA. (2012–2016). *Global Interface Japan Newsletter*, Issues 1–26.
HIDA web-based questionnaire. (2012, November 1–18).
Islam, S. (Ed.). (1991). *Yen for development: Japanese foreign aid and the politics of burden-sharing*. New York: A Council on Foreign Relations Book.
Jensen, P. S., & Paldam, M. (2003). *Can the new aid-growth models be replicated*? Working No. 2003-17. Institute for Economics, Aarhus.
Kasuga, H. (2007). Evaluating the impacts of foreign direct investment, aid and saving in developing countries. *Journal of International Money and Finance, 26*, 213–228.
Kenshū. (Autumn 2011). *Newsletter*, Issue no. 322.
Kimura, H., & Todo, Y. (2009). Is foreign aid a vanguard of foreign direct investment? A gravity-equation approach. *World Development, 38*(4), 482–497.
Levitt, P. (1998). Social remittances: Migration driven local –level forms of cultural diffusion. *International Migration Review, 32*(4), 926–948.
Levitt, P., & Lamba-Nieves, D. (2011). Social remittances reconsidered. *Journal of Ethnic and Migration Studies, 37*(1), 1–22.
McGillivray, M., Feeny, S., Hermes, N., & Lensink, R. (2006). Controversies over the impact of development aid: It works; it doesn't; it can, but that depends. *Journal of International Development, 18*(2006), 1031–1050.
Mosley, P. (1987). *Overseas development aid: Its defence and reform*. Brighton: Wheatsheaf.
OECD. (2012, May). *New sources of growth: Knowledge-based capital driving investment and productivity in the 21st century.* Interim Project Findings.
OECD. (2013). *New sources of growth: Knowledge-based capital key analyses and policy conclusions.* Synthesis report. Retrieved August 4, 2014, from http://www.oecd.org/sti/inno/knowledge-based-capital-synthesis.pdf
Ogborn, M. (2008). *Global lives: Britain and the World, 1550-1800.* Cambridge: Cambridge University Press.
Orr, R. M., Jr. (1990). *The emergence of Japan's Foreign Aid power*. New York: Columbia University Press.
Rix, A. (1980). *Japan's Economic Aid: Policy-making and politics*. London: Croom Helm.
Rix, A. (1989). Japan's Foreign Aid Policy: A capacity for leadership? *Pacific Affairs, 62*(4), 461–475.
Vogt, G. (Ed.). (2012, July). International labor migration to Japan: Current models and future outlook. *Asian: The German Journal on Contemporary Asia*, *124*, 8–26.
World Bank. (1998). *Assessing aid: What works, what doesn't, and why*. Washington, DC: The World Bank.

Globalization and Female Labor Force Participation in Developing Countries: An Empirical (Re-)Assessment

Konstantin M. Wacker, Arusha Cooray, and Isis Gaddis

Abstract We investigate the impact of foreign direct investment (FDI) and trade, as two measures of globalization, on female labor force participation rate in a sample of 80 developing countries over the time period 1980–2005. Contrary to the mainstream view in the literature, which is mainly based on country-case studies or simple cross-country variation, we find that both, FDI and trade have a generally negative impact on female labor force participation. While the impact is of negligible economic size, it is stronger for younger cohorts, possibly reflecting a higher return to education in open economies. We further find a large degree of cross-regional heterogeneity and that the effect of globalization on female labor force participation depends on the industrial structure, with more positive effects in economies with a higher share of industry in value added. We can thereby explain why country studies find other effects and question the generalization of their results into an overarching globalization tale concerning female labor force participation.

The End of Labor is to Gain Leisure (Aristotle)

1 Introduction

The increase in female labor force participation (FLFP) is one of the most significant global developments of the last decades. There is broad consensus that this is in general a welcome trend since it may contribute to women's economic empowerment and reduce economic costs associated with the underutilization of

K.M. Wacker (✉)
University of Mainz, Mainz, Germany
e-mail: kwacker@uni-mainz.de

A. Cooray
Nottingham University Business School Malaysia, Semenyih, Malaysia
e-mail: Arusha.Cooray@nottingham.edu.my

I. Gaddis
World Bank and University of Göttingen, Göttingen, Germany
e-mail: igaddis@worldbank.org

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_22

women's skills and labor (Klasen and Lamanna 2009; World Bank 2011). However, the determinants of this development are more controversial.

Most contributions to the applied labor economics literature in this context have focused on the United States and some European countries (e.g. Killingsworth and Heckman 1987; Blundell and MaCurdy 1999; Blau and Kahn 2007). Concerning developing countries, one strand in the field argues that there would be a U-shaped relationship between development and the labor force participation of married women (e.g. Goldin 1990 and 1995, Mammen and Paxson 2000): At very low levels of income, agriculture dominates and a large proportion of females are in (often unpaid) agricultural employment. With rising incomes, the introduction of new production technology, and transition to a formal-sector based industrial economy, the income effect (from higher earnings of the spouse) dominates the substitution effect in the labor force participation decision of married women, leading to a fall in the FLFP rate (FLFPR). Women face negative biases against female industrial workers and incompatibility of formal sector employment with traditional care-giving activities at this stage of the development process. As development continues, female education increases and the substitution effect begins to outweigh the income effect, leading to an increase in the FLFPR (Goldin 1990 and 1995). However, Gaddis and Klasen (2014) show that empirical support for the U-shaped relationship between FLFPR and aggregate GDP is not robust across different data sources and specifications, and is particularly weak in non-OECD countries. They also show that agriculture, mining, manufacturing and services generate different dynamics for female employment.

Another line in the literature has argued that increased openness has led to an increase in the FLFPR in developing countries. There are several interconnected channels through which globalization could lead to a feminization of the labor force. Due to existing gender discrepancies, women might be prepared to work long for a low wage and without joining a union. Therefore, exporting and multinational firms are more likely to employ women, especially since most tasks of the industries where developing countries have a comparative advantage are less skill-intensive or a priori expected to be female-intensive (Çağaty and Berik 1990; Anderson 2005). Even if male-intensive sectors benefit most from increased openness, FLFP may rise in equilibrium since men might leave female-intensive industries to take up new jobs in the export sector, thereby opening up employment opportunities for women (Sauré and Zoabi 2014).[1] The process might be accelerated by structural adjustment programs which were often implemented in the course of increased openness, since the accompanied increase in labor market flexibility would make it easier for firms to substitute women for men (Standing 1989; Çağaty and Özler 1995).

Evidence suggests that the initial phase of globalization led to a significant rise in female employment with women accounting for more than a third of the

[1] Similarly, arguments in line with the agricultural linkages literature (Lele and Mellor 1981; Mellor and Lele 1973, 1975) can be built where the openness-induced surge in the male-intensive sector also spills over to the female-intensive sector through production and consumption linkages.

manufacturing labor force in the developing economies and approximately half in some Asian nations. For example, in Sri Lanka, women's share in manufacturing increased from 32 % in 1975 to 61 % in 1992 and in Mauritius from 19 % in 1970 to 60 % in 1992. In Mauritius, where women accounted for 70 % of the labor force in textiles and garments in 1990, employment increased by 344 % since 1980 (Mehra and Gammage 1999). In Morocco and Tunisia, similarly, women's employment in the textile and clothing industry has tripled since 1980 (Mehra and Gammage 1999). In Bangladesh, the number of garment factories increased from four in 1978 to 2400 in 1995; and of the 1.2 million workers employed, 90 % were women below the age of 25. The garment industry represents around three-quarters of female wage occupations in Bangladesh (Sen 1999).

More recently however, some evidence points to a de-feminization of the labor force with an increase in demand for male labor due to technological progress and restructuring of export industries brought about by globalization. In Korea, for example, female employment in manufacturing increased in the early 1980s and then fell over the 1989–1993 period due to the need for specific skills (Mehra and Gammage 1999). Pradhan (2006) finds that in India, technology and capital intensive production have negatively affected female and unskilled workers, but positively affected contract workers. The FLFPR has similarly fallen over the 2000–2009 period from 26.6 % to 23.5 % in Turkey due to structural change (Yenilmez and Isikli 2010). There has also been a trend towards greater "flexibilization" of the labor force with more women working from home (Sen 1999). In Palestine, 90 % of home-based employees in the textile and garment industry are women, of which 35 % are unwaged (Mehra and Gammage 1999). The evidence on the relation between globalization and the FLFPR therefore is not clear-cut. Additionally, the effects of globalization on the FLFPR seem to differ across regions. This is also highlighted by the World Bank (2012: box 1, p. 30 and 2011: 11/12), which emphasizes the role of changing social norms and regulations as well as improved public service provisioning as a main driver of increased FLFP.

The main contribution of this chapter is to put the relationship between FLFP and globalization into perspective by showing that the observed increase in FLFP in developing countries over the last decades cannot be attributed to globalization, as most of the previous literature has suggested. Our results rather suggest that globalization has a negative effect on FLFP, which is more pronounced among younger age cohorts. There also appears to be a large degree of regional heterogeneity in the results and the effect apparently depends on the industrial structure of the economy. Our contribution hence highlights the need to take more country-specifics into account when explaining the economic and social effects of globalization on FLFP.

The remainder of this chapter is structured as follows: We review previous empirical contributions in Sect. 2 (and also summarize them in Table 10 in Appendix B). They have mainly found support for a positive relationship between globalization and FLFP but are mostly based on individual country-studies or simple pooled cross-country OLS regressions. We therefore introduce our data and methodology in Sect. 3 and improve on previous studies in a number of ways. First,

we estimate separate coefficients by region to allow for regional heterogeneity. Second, we accommodate the sectoral structure of the investigated economies by allowing for interactions of FDI with industrial/agricultural value added and considering overall trade, overall exports and exports in services separately. Third, we investigate heterogeneity of effects across age cohorts. Finally, we reduce potential parameter biases due to unobserved cross-country heterogeneity by basing identification exclusively on over-time variation. We present our results in Sect. 4 and show in Sect. 5 that even under alternative specifications and when replicating previous results in the literature with slight modifications, no evidence of a positive impact of globalization on FLFP can be found. We discuss the results and conclude the paper in Sect. 6.

2 Review of the Empirical Literature[2]

Based on rather descriptive and anecdotal evidence, early case studies such as Cho and Koo (1983), Hein (1984), ILO (1985), or, later on, by Kabeer and Mahmud (2004) suggest that aspects related to globalization, such as export-led industrialization, export processing zones and higher employment in multinational firms have had a positive impact on FLFP. According to Kabeer and Mahmud (2004), the percentage of females in manufacturing in Bangladesh increased from about 4 % in 1974 to 55 % in 1985–86, and urban female labor force participation rates from 12 % in 1983–84 to 20.5 % in 1995–96. Similarly, Çağaty and Berik (1990) state that export orientation led to an increase in urban female labor force participation in Turkey from 11.2 % to 16.9 % between 1982 and 1988. Using a fairly simple OLS regression for 3-digit SIC Turkish manufacturing industries in 1966 and 1982, Çağaty and Berik (1990) show that export orientation had a statistically significant positive impact on the female share of wage workers in private sector manufacturing, however, not public sector manufacturing. Their results also indicate that the switch to export orientation does not lead to a feminization of the labor force in large scale manufacturing. Export orientation in Turkey was accompanied by structural adjustment policies which included lowering of labor standards, wages and relaxation of union activity. A similar empirical strategy is applied to Indian industry data from the late 1990s and early 2000s by Pradhan (2006), who finds that trade has led to an increase in employment of women and unskilled workers while it has had a neutral effect on contract and regular workers and that exports have a significant and positive (though economically small) impact on the female/male working-days ratio. FDI is found to have a negative effect on contract and unskilled workers.

Özler (2000) builds upon this strand of the literature by using plant-level data for the period 1983–1985 from the Turkish manufacturing sector and shows

[2]We also provide an overview of the related empirical literature in Table 10 in Appendix B.

that the female share of employment in a plant increases with the export to total output ratio of its sector. In line with the arguments above, she notes that women are often employed in low-skill and low-paid jobs and especially among those establishments where investment in machinery and equipment leads to a decline in the female employment share, thus pointing to dynamic long-run effects disadvantageous to a feminization of the labor force (in this context, see also Wood 1998, and Seguino 2000). This suggests, globalization may first lead to an expansion of female-intensive sectors which then rationalize production by investment and technological progress. However, while the plant-level perspective of the study has certain advantages, it fails to convincingly resolve the problem of an unobserved heterogeneity bias and cannot reveal any spill-over effects on non-manufacturing sectors. Such spill-over effects are documented in Gaddis and Pieters (in press), who argue that trade liberalization reforms in Brazil in the early 1990s were associated with a decrease in male and female labor force participation but still contributed to a narrowing of the percentage point gender gap in labor force participation and employment rates.

Tying in with the above-mentioned literature on the feminization-U, Çağaty and Özler (1995) use another approach by using pooled data from 1985 and 1990 for 165 countries to investigate the impact of long-term development on the female share of the labor force. They argue that structural adjustment policies have led to an increase in feminization of the labor force via worsening income distribution and increased openness.

Gray et al. (2006) use data for 180 countries at 5-year intervals between 1975 and 2000 to estimate the impact of trade (measured as the log of total imports plus total exports to GDP), FDI (as a percentage of the gross fixed capital formation) and other globalization-related variables on the female percentage share of the workforce and other female-specific outcome variables. We think, their finding (p. 319ff) that none of the two former variables has a significant impact on (relative) FLFP may be due to the fact that they exert a converse impact in developing versus industrialized countries; a heterogeneity that results in overall insignificant estimates.

Similarly, Bussmann (2009) addresses the wider research question whether economic globalization (in particular, trade/GDP) improves certain aspects of women's welfare (especially health and education). Using fixed effects and generalized methods of moments (GMM) techniques for annual panel data in the period 1970–2000, she finds that trade/GDP increases overall FLFP in non-OECD countries.

While there are some opposite arguments highlighting that FDI in developing economies benefits male engineers or computer programmers more than female ones because they are likely to be better educated (Oostendorp 2009), or pointing to occupational gender segregation (Greenhalgh 1985; Anker 1998, Anker et al.

2003),[3] the large majority of empirical studies seems to suggest that globalization has raised FLFP in developing countries.

In our view, however, these supposed "stylized facts" suffer from certain methodological shortcomings that give rise to our empirical re-assessment. First, we find it risky to generalize from country-case studies to an overarching tale of globalization, feminization and development. On the other hand, most cross-country studies so far have suffered from the problem of potentially biased estimates due to unobserved heterogeneity across countries. Finally, rather short time dimensions have imposed certain restrictions on the equilibrium dynamics of the relationship between openness and FLFP. By using a comprehensive panel of 80 developing countries over almost three decades and applying a fixed-effects methodology, we can deal with all of these potential problems and show that this leads to quite contrary results than the ones obtained in the mainstream literature.

3 Data and Methodology

3.1 Data

We use data on FLFP from the 5th revision of the ILO's Estimates and Projections of the Economically Active Population (EAPEP) database (ILO 2009). The EAPEP contains data on the male and female economically active population based on country reports and ILO staff estimates for 191 countries, which includes both industrialized and developing countries. The 5th revision data cover the period 1980–2008; the data thus have a high overlap with our FDI and trade data. In line with Gray et al. (2006) and Gaddis and Klasen (2014) and in order to minimize problems associated with serial correlation and to focus more on long-run effects, we consider the observations for every fifth year over the period 1980–2005 for estimation.[4] The FLFPR is defined as the number of economically active women divided by the total female population (FPOP) of the relevant age group j in country i at time t:

$$FLFPR_{ijt} = \frac{FLFP_{ijt}}{FPOP_{ijt}}. \tag{1}$$

[3]Note that the effect of occupational gender segregation on female labor force participation in the context of globalization is not clear a priori and depends on the elasticity of substitution between female and male labor, the pattern of trade liberalization, and associated relative demand shifts.

[4]This should generally be similar to using 5-year averages. However, much data is only available for every 5th year (e.g. the Barro and Lee 2010, dataset), or values between these observation points are interpolated (e.g. for certain values in the EAPEP database) so that the argument for using 5-year averages is rather weak.

The ILO definition of economic activity (effective until 2013) captures all persons (employed or unemployed) that supply labor for activities included in the United Nations System of National Accounts (SNA; cf. ILO 1990).[5] This includes self-employment for the production of marketed goods and services as well as the production of goods consumed within the household. It does, however, not include the production of non-marketed services (domestic tasks, nursing of own children), since they are not included in the SNA. This distinction is important to remember, as many women outside of the labor force are employed in producing such non-marketed services. It should also be noted that the EAPEP data only provide information on economic activity rates, but not on total hours worked. Hence, the data allows us to investigate changes in labor supply at the extensive margin (participation decision) but not at the intensive margin (hours worked).

One of our most relevant explanatory variables is the stock of inward foreign direct investment (FDI) relative to GDP, taken from UNCTAD, which serves a proxies for the activity of multinational firms in the economy under investigation. Financial stock data, as opposed to operational data (such as multinationals' sales, number of employees) reflects the effective share of foreign ownership in host country firms and is available for a large group of countries and years (cf. Wacker 2013, for a discussion of measuring FDI and multinational firms). Furthermore, we use trade, imports and exports relative to GDP as other relevant measures for globalization. These data include trade in goods *and* services and come from the World Bank World Development Indicators (WDI). WDI also provide most of our control variables such as GDP per capita in constant 2005 international $ purchasing power parities, the total fertility rate (births per woman), and the shares of agriculture and industry value added in GDP. From WDI, we also construct the percentage growth rate of real GDP p.c. (in constant local currency). Since we use fixed effect models, the fixed effect takes out the long-run average growth so that this variable should be interpreted as the cyclical component of the model. For years of schooling we use the female measures of the corresponding cohorts provided by Barro and Lee (2010).[6] An overview over the variables and their summary statistics are provided in Table 7 in the appendix.

Since we focus on developing countries, we follow the convention in the literature to consider countries classified as "low income" or "lower middle income" by the World Bank (for 1987, the first year available). This gives us a sample of 80 developing countries in total.

[5] In 2013, the international definition of employment was revised to work performed for pay or profit (ILO 2013). There is hence no longer a direct link between employment and the SNA production boundary.

[6] If we aggregate their data over various cohorts, we use the ILO female population data as weights. Linear interpolation is used to obtain data points between the 5-year survey intervals. This is necessary since most explanatory variables are lagged by 1 year.

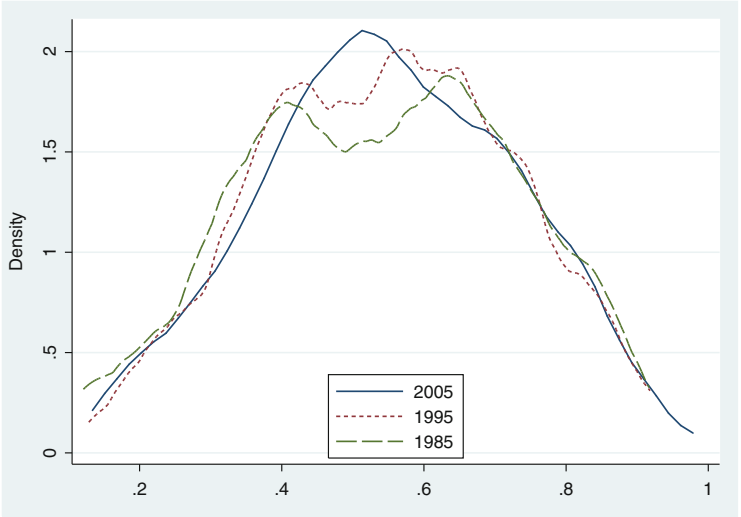

Fig. 1 Distribution of Female Labor Force Participation Rate (weighted aggregate over cohorts). Epanechnikov kernel density estimation for different decades

3.2 Descriptive Analysis

Figure 1 plots the distribution of the FLFPR across the developing world in 1985, in 1995 and in 2005.[7] As one can see, the distribution gets smoother in the center in 2005 when compared to the decades before, which is also reflected in a decreasing standard error in Table 1. The steadily increasing mean of the distribution in Table 1 also shows that FLFPR indeed increased over the period usually referred to as "globalization."[8]

Figure 2 depicts the development of our two variables measuring globalization, FDI stock/GDP and trade/GDP, for the same years. As one can see, trade to GDP increased relatively steadily throughout the three decades while FDI/GDP experienced its main surge only in the last decade.

3.3 Econometric Model

Following the literature of determinants of FLFP (Bloom et al. 2009; Çağaty and Özler 1995; Gaddis and Klasen 2014; Mammen and Paxson 2000), we estimate a linear model, where the dependent variable is the female labor force participation

[7] In order to make the data in and between Figs. 1 and 2 comparable, we only used countries which have no missing observations for FLFPR, FDI/GDP and trade/GDP in 1985, 1995 and 2005 for both graphs. We end up with 77 (developing) countries.

[8] For summary statistics of all other variables, see Table 7 in the appendix.

Table 1 Descriptive statistics of main variables

Variable	Statistic	1985	1995	2005
FLFPR	Mean	0.507	0.524	0.549
	Std. Dev.	0.199	0.185	0.173
	Min	0.121	0.129	0.199
	Max	0.917	0.918	0.913
FDI stock/GDP	Mean	0.211	0.236	0.365
	Std. Dev.	0.339	0.275	0.355
	Min	8×10^{-6}	0.001	0.002
	Max	1.650	1.399	1.606
Trade/GDP	Mean	0.648	0.760	0.844
	Std. Dev.	0.376	0.408	0.401
	Min	0.130	0.025	0.003
	Max	1.517	2.133	2.121

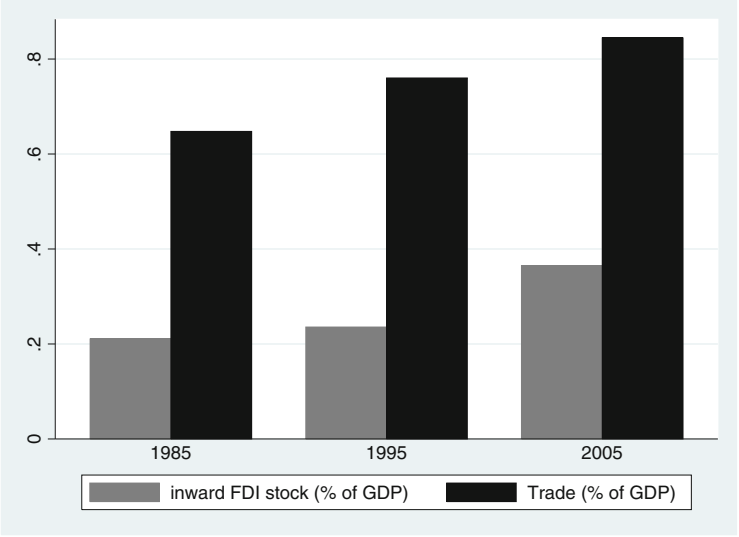

Fig. 2 Development of main explanatory variables. Average over 77 developing countries, trade is the sum of exports and imports

rate (FLFPR) in levels and is explained by a number of covariates x:

$$\frac{FLFP}{FPOP} = FLFPR = x_1\theta_1 + \ldots + x_k\theta_k + u, \tag{2}$$

where u is an error term discussed below.

Our dataset thus has two levels of cross-sections: countries $i = 1, \ldots, N$ and age cohorts[9] $j = 1, \ldots, 10$. In our model, which hence can be considered as "hierarchical," we use country-specific cohort fixed effects, i.e. fixed effects for every cohort which are allowed to vary by country. The reason is, first, that unobserved heterogeneity across countries is likely and the same holds for age cohorts. For example, the age cohort 15–19 years is less likely to join the labor force than the age cohort 35–39 if the former has a higher probability of being in education. Furthermore, we assume that these cohort-fixed effects are country-specific due to different educational systems and differing conceptions of life across countries. Note that not controlling for this unobserved heterogeneity will result in biased and inconsistent results if the heterogeneity is correlated with some right-hand side variables. This is a clear advantage over previous cross-section studies in the field. In our sample of 80 countries with 10 age cohorts, this leads to $80 \times 10 = 800$ cross-section fixed effects.

Furthermore, we control for time-fixed effects. This is motivated by the consideration that there may be global effects influencing FLFP which are correlated with our covariables. This may lead to both, biased results and cross-sectional dependence in the structure of the error term. Formally,

$$\frac{FLFP_{ijt}}{FPOP_{ijt}} = Z_{ijt}\theta + X_{it}\beta + u_{ijt} \tag{3}$$

where Z is a $10\ N \cdot T \times k$ matrix collecting the k country *and* cohort-specific covariables (in our case education), X is a $10\ N \cdot T \times m$ matrix collecting the m country-specific covariables and u has the structure

$$u_{ijt} = \mu_{ij} + \gamma_t + \varepsilon_{ijt}, \tag{4}$$

where μ and γ are the country-cohort and time fixed effects, respectively, which are estimated and ε is an i.i.d. error term.[10] Note that we only take every fifth

[9]The age cohorts are 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, and 60–64. We excluded the cohort of 65+ years from our analysis because labor force participation in this cohort is driven by factors that might be very different from other cohorts.

[10]A potential shortfall of the FE estimator is the fact that the process we explore is likely to have a complex dynamic structure while FE can be seen as a 'short-run' estimator. An alternative dynamic estimator, however, is difficult to specify depending on the complexity of the dynamic process and will potentially suffer severely from parameter heterogeneity (cf. e.g. Pesaran and Smith 1995; Phillips and Sul 2003) which is in fact present as we show in later parts of this study. The FE estimator, in our view, has the advantage that its properties are studied extensively and well-known. Furthermore, our main explanatory variables, FDI stocks and trade (or, exports) relative to GDP are very persistent variables. Under such circumstances the static fixed-effects estimator can be biased from a (consistent) short-run estimator towards the long-run impact. More explanations and evidence are given in Baltagi and Griffin (1984), Egger and Pfaffermayr (2005), and Wacker (2013) but the main intuition is the fact that in the presence of an omitted lag structure, the high correlation between the included variable and its own lags causes an omitted variable bias by incorporating

observation in time, i.e. $t = 1980, 1985, \ldots, 2005$ and that the only cohort-specific covariable in Z is the educational data (hence, $k = 1$). In each of the columns of X, there will be ten identical entries.

In summary, our identification strategy exclusively uses the data variation *within the country-specific cohorts* as a response of FLFPR to data variation *within the country-specific cohorts* (education) or *within the country-specific* variables (all other covariables) over 5-year intervals, accounting for global shocks at every point in time. We generally assume that the response to changes in the country-specific variables is homogenous across cohorts but also allow for cohort-heterogeneity in response to country-wide changes in globalization.

We consider the errors of our model to be correlated at the country level across the time and cohort dimension for reasons explained in Appendix C, which also explains how we econometrically tackle this issue. In short, we simply cluster the errors at the country level (instead of the country-cohort level for which the fixed effects are estimated).

4 Empirical Results

We start with a very simple model specification including only our main globalization variables (and fixed effects for country-cohorts and years) as explanatory variables. Although omitted variables such as GDP p.c. might bias the precise results, we think it is interesting to see how much scope there is for globalization to affect FLFPR via different channels (such as income effects). The results, presented in the first four columns of Table 2, show that the impact is negative throughout and statistically significant[11] only in two specifications for trade and exports. Note that trade and FDI are highly correlated,[12] so multicollinearity inflates standard errors (while parameter estimates are still consistent) and we therefore report specifications with both variables together and separately. The negative impact of trade is driven by exports, so we focus on exports for the remainder of the analysis. The most striking fact besides from lacking statistical significance and the negative sign of the estimated coefficient is the notably small economic relevance of both effects. The highest parameter is −0.064 for exports in column (4), implying that a 10 percentage points increase in the exports/GDP ratio, roughly the increase observed over the

the impacts of deeper lags. We hence think that our FE estimates come at a relatively low risk, especially as we are using only every 5th observation year (hence looking at longer time periods), and will give a good intuition about the underlying economic forces at work. We discuss potential extensions for future research in the concluding section of this paper.

[11] Unless stated otherwise we refer to statistical significance as significance at the 5 % level and call significance at the 10 % and 1 % level as weakly and strongly statistically significant, respectively.

[12] Regressing FDI stock/GDP on the other covariables of model (7) using the same subsample and each 5th yearly observation leads to a highly significant estimator of 0.267 for trade/GDP (t-statistic 2.58).

20 years 1985–2005, leads to a 0.64 percentage points decrease of FLFPR. Considering that the actual increase in FLFPR during the 20 years between 1985 and 2005 was 4.2 percentage points, this is a rather small magnitude.

The remaining models in Table 2 include our seven control variables. While there is some change in the levels of statistical significance, our overall result remains rather stable: There is no evidence so far, that globalization had an economically relevant impact on female labor market participation. With the control variables included, it is the FDI stock that seems to be more robust statistically, however, the magnitude is negligible since the estimated parameter, −0.0116 in the "best" case, implies that a 10 percentage points increase in FDI stock/GDP leads to a 0.12 percentage points decrease of FLFPR. Exports are only statistically significant when FDI stock is excluded (though standard errors are reasonable in model (5) as well), the economic relevance is barely higher than in the unconditional model (4), however.

These first results do not necessarily mean that our measures for globalization have no impact on women in their decision to join the labor force—they are aggregate effects and capture a wide range of different activities. In the remainder of this section, we therefore allow the effect to differ across regions, to depend on the industrial structure of the economy and to vary by age cohort.

We start with allowing for heterogeneity of the impacts across regions, i.e. model (5) from Table 2 above is re-estimated for six regions into which the World Bank classifies (developing) countries. This should reflect that certain effects on FLFP which we explicitly capture in our covariables, interact with certain norms, cultures, and regulations that are assumed to be largely homogeneous within these regions. Furthermore, if one thinks within a standard trade framework, countries will develop those sectors of their economy after trade-liberalization, where they have a comparative advantage. For least developed countries these are lower-skilled labor intensive industries. While women may have a "natural" advantage in some of these industries (for example, certain tasks in the textile sector), most other tasks may benefit from physical strength and hence primarily demand male labor (cf. World Bank 2012: 33). Accordingly, we would expect that the impact depends on the country's comparative advantage and this would suggest that the impact should generally be different between regions and depend on the country's competitive advantage. These considerations are supported by a view at Table 3. For example, we find significant negative impacts of exports on female labor force participation in South and East Asia, and of FDI in Sub-Saharan Africa. Conversely we find a positive effect of FDI on women's economic activity in Eastern Europe/Central Asia. Generally, the table shows a considerable degree of variety between the different regions. It is also noteworthy that the primary sector exhibits a strong positive (and statistically highly significant) impact on FLFP in the MENA countries, whereas industry value added implies a negative and relevant (highly significant) impact in this region. This probably reflects the high share of mining (particularly from oil exploitation) in industrial value added in the region, a sector which traditionally employs few women (see also Gaddis and Klasen,

Table 2 Main regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR
Trade/GDP	−0.0215*		−0.0277**				−0.0142	
(−1)	(0.0127)		(0.0131)				(0.0169)	
FDI stock/GDP	−0.00424	−0.00120			−0.0116*	−0.00573**	−0.0115*	
(−1)	(0.00601)	(0.000950)			(0.00669)	(0.00236)	(0.00637)	
Exports/GDP				−0.0641**	−0.0370			−0.0698**
(−1)				(0.0260)	(0.0316)			(0.0325)
ln(GDP p.c. PPP)					−0.120	−0.154*	−0.127	−0.0433
(−1)					(0.0900)	(0.0918)	(0.0904)	(0.0901)
ln(GDP p.c. PPP)^2					0.00901	0.0115	0.00939	0.00506
(−1)					(0.00684)	(0.00701)	(0.00688)	(0.00675)
Fertility rate					−0.00508	−0.00707	−0.00499	−0.00247
					(0.00751)	(0.00753)	(0.00746)	(0.00709)
Years of schooling					0.00612	0.00290	0.00558	0.00612
					(0.00751)	(0.00780)	(0.00747)	(0.00737)
Agricultural value added					0.0530	0.0448	0.0511	0.0869*
					(0.0523)	(0.0537)	(0.0525)	(0.0512)
Industry value added					−0.0320	−0.0349	−0.0401	0.0152
					(0.0435)	(0.0446)	(0.0431)	(0.0413)
GDP growth rate					−0.0510	−0.0461	−0.0521	−0.0630*
					(0.0406)	(0.0305)	(0.0428)	(0.0378)
Constant	0.551***	0.536***	0.560***	0.560***	0.954***	1.084***	0.987***	0.600*
	(0.0100)	(0.00244)	(0.00992)	(0.00847)	(0.328)	(0.332)	(0.329)	(0.323)

(continued)

Table 2 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4860	5240	5020	5020	3470	3530	3470	3580
R-squared	0.190	0.141	0.171	0.176	0.221	0.200	0.219	0.199
Number of cross-sections	1120	1150	1120	1120	800	800	800	800

Fixed effects regression taking every 5th year. Cluster-robust standard errors in parentheses
***, **, * denote statistical significance at the 1 %, 5 %, and 10 % level, respectively

Table 3 Effects by region

	(1)	(2)	(3)	(4)	(5)	(6)
Region	E. Asia	EE & CA	Latin Am.	MENA	S. Asia	SSA
Variables	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR
Exports/GDP	**−0.156****	**−0.131***	**0.0272**	**−0.0181**	**−0.629****	**0.00111**
(−1)	**(0.0591)**	**(0.0650)**	**(0.0394)**	**(0.0273)**	**(0.160)**	**(0.0328)**
FDI stock/GP	**−0.00584**	**0.0949****	**0.0365**	**−0.0295**	**0.851**	**−0.0118****
(−1)	**(0.0670)**	**(0.0410)**	**(0.0358)**	**(0.0182)**	**(0.480)**	**(0.00445)**
ln(GDP p.c. PPP)	0.170	−0.137	−0.172	−1.209***	0.535	−0.153*
(−1)	(0.225)	(0.0989)	(0.488)	(0.241)	(0.463)	(0.0763)
$\ln(\text{GDP p.c. PPP})^2$	−0.0188	0.0217**	0.0131	0.0839***	−0.0729*	0.0120**
(−1)	(0.0199)	(0.00660)	(0.0316)	(0.0160)	(0.0360)	(0.00573)
Fertility rate	0.0167	−0.0361*	0.0186	−0.0846***	0.0582*	−0.0156*
	(0.0132)	(0.0184)	(0.0181)	(0.00745)	(0.0252)	(0.00873)
Years of schooling	0.0371	0.00421	−0.0229**	−0.00210	−0.0349	0.00459
	(0.0242)	(0.00873)	(0.00878)	(0.00399)	(0.0225)	(0.00849)
Agricultural value Added	−0.0329	−0.178*	0.129	0.359***	−0.422	0.0439
	(0.121)	(0.0833)	(0.125)	(0.0703)	(0.356)	(0.0550)
Industry value added	−0.170	−0.264**	−0.0337	−0.133***	0.624**	−0.0327
	(0.132)	(0.0949)	(0.0713)	(0.0203)	(0.243)	(0.0444)
GDP growth rate	0.0606	0.363***	−0.0447	−0.147***	0.130	−0.0387
	(0.110)	(0.0759)	(0.0396)	(0.0358)	(0.321)	(0.0506)
Constant	0.161	0.744*	0.958	4.866***	0.00351	1.188***
	(0.701)	(0.375)	(1.837)	(0.912)	(1.718)	(0.282)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	550	290	680	270	270	1410
R-squared	0.290	0.238	0.602	0.512	0.485	0.300
Number of cross-sections	120	90	170	60	60	300

Fixed effects regression taking every 5th year. Cluster-robust standard errors in parentheses
EE & CA Eastern Europe and Central Asia, *MENA* Middle East and North Africa, *SSA* Sub-Saharan Africa
***, **, * denote statistical significance at the 1 %, 5 %, and 10 % level, respectively

2014). Sectoral movements also seem to play an important role in the Eastern European/Central Asian countries.

It should be noted that this region-specific estimation dramatically reduces the number of observations, thereby negatively affecting statistical levels of significance. We hence also follow a different approach to capture heterogeneity across countries by allowing our globalization variables to interact with the industrial structure of the economy for reasons similar to those discussed above. Due to unavailability of comprehensive sectoral FDI data, we interact FDI with the value added in the industry sector and the primary sector, respectively, while trade data is indeed available for the different sectors of the economy.

The results, presented in Table 4 (first column) show that FDI stock is highly significant and negative with a similar magnitude as in model (5) of Table 2, while

Table 4 Interaction with industrial structure

	(1)	(2)	(3)
Variables	FLFPR	FLFPR	FLFPR
ln(GDP p.c. PPP)	−0.132	−0.133	−0.144
(−1)	(0.0893)	(0.0885)	(0.0991)
ln(GDP p.c. PPP)^2	0.00987	0.00990	0.0100
(−1)	(0.00681)	(0.00676)	(0.00763)
Fertility rate	−0.00500	−0.00498	−0.00302
	(0.00746)	(0.00756)	(0.00749)
Years of schooling	0.00529	0.00494	0.00604
	(0.00746)	(0.00755)	(0.00814)
Agricultural value	0.0484	0.0537	0.0523
Added	(0.0529)	(0.0542)	(0.0565)
Industry value added	−0.0674	−0.0487	−0.00549
	(0.0494)	(0.0444)	(0.0411)
GDP growth rate	−0.0423	−0.0451	−0.113**
	(0.0425)	(0.0421)	(0.0488)
Trade/GDP	**−0.0217**	**−0.0215**	
(−1)	**(0.0162)**	**(0.0163)**	
FDI stock/GDP	**−0.0179*****	**0.0157**	
(−1)	**(0.00507)**	**(0.0246)**	
Industry v.a. × FDI	**0.0642**		
(−1)	**(0.0480)**		
Agricultural v.a. × FDI		**−0.0376**	
(−1)		**(0.0294)**	
Trade in Services/GDP			**−0.0473*****
(−1)			**(0.0160)**
Exports/GDP			**−0.0251**
(−1)			**(0.0338)**
Constant	1.012***	1.013***	1.007***
	(0.322)	(0.319)	(0.349)
Year dummies	Yes	Yes	Yes
Observations	3450	3450	3280
R-squared	0.216	0.217	0.244
Number of cross-sections	800	800	790

Fixed effects regression taking every 5th year. Cluster-robust standard errors in parentheses
***, **, * denote statistical significance at the 1 %, 5 %, and 10 % level, respectively

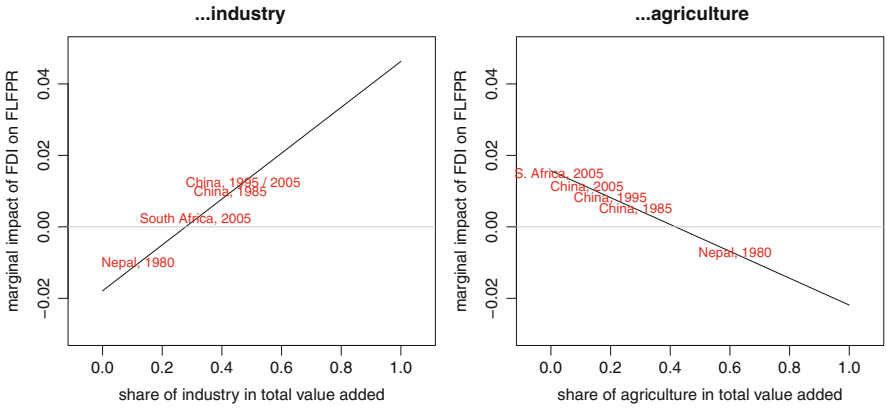

Fig. 3 Impact of FDI depending on the sectoral structure of the economy

the interaction is about 3.5 times the size of the mere FDI parameter and positive.[13] This means that once the industrial sector is developed, more FDI will have a positive impact on the FLFPR. More precisely, once the industrial sector accounts for at least 28 % of value added of the developing country's economy, additional FDI will have a positive impact.[14] The magnitude is still low: Assuming that the whole economy is producing half or all of its output in the industrial goods sector, a 10 percentage point increase in FDI stock/GDP will cause a 0.14 or 0.46 percentage point increase in FLFPR, respectively. This relationship is depicted in the left panel of Fig. 3. It shows that the higher the share of industry value added, the more favorable the marginal impact of FDI on FLFPR. The right panel does the same with agriculture, which basically is the mirrored image of the left panel. In order to get an impression for the economic magnitudes, we added some country examples to the graph. We included China in 1985, 1995 and 2005 because it serves as an example of a developing country that has gone through an enormous structural change over the last decades and is well-known to the profession. From the right panel we see that the share of agriculture in China's value added has decreased from 1985 to 2005. This led to expansion of the industrial sector in the first decade and of the service sector in the second decade (the data point in 1995 and 2005 is almost identical in the left panel). This change has brought China into a more favorable/positive condition concerning the impact of FDI on FLFPR: Our model predicts that the effect of FDI on FLFPR was more positive in later years than in 1985.

[13]The parameter itself is not statistically significant (t-statistic 1.34). The relevant test statistic, however, is an F-test for joint significance of FDI and the interaction term. Here, we can reject that they jointly have no impact on FLFPR on the 1 % level of statistical significance.

[14]A 10 percentage point increase in FDI will have a $0.10 \times (-0.0179) + 0.10 \times 0.28 \times 0.0642 = 0.0000076$ percentage points impact in an economy where industry accounts for exactly 28 % of value added.

Nepal in 1980 serves as an example of a very agrarian economy, the impact of FDI is accordingly negative. Finally, South Africa in 2005 was a fairly modern economy; the model would hence suggest a positive impact of FDI on FLFPR.

These results imply that the factor demand of multinational firms does not necessarily have a (conditional) anti-female bias since the above mentioned negative impact of FDI appears to be mainly driven by changes in the industry structure. This is supported by column (2) in Table 4 where we allow FDI to interact with the primary sector. The negative impact of FDI now vanishes; it becomes insignificant and positive while the interaction with the primary sector is negative and insignificant.[15]

Similarly, for trade in column (3) of Table 4 we find that its negative impact is, somewhat surprisingly, driven by trade in services—including trade in services into the model turns the overall export parameter estimate statistically insignificant, whereas trade in services is negative and highly significant but of small magnitude. This result is rather surprising on a first view because one would expect that women are very likely to work in the service sector. However, especially in the tradable service sector, the skill-premium might be high, hence inducing young women to invest more into education and therefore stay off the labor market in younger cohorts. Furthermore, these results are in line with the findings of Oostendorp (2009) that globalization may benefit male engineers or computer programmers more than female ones because they are likely to be better educated, and with an aspect of the results of Bussmann (2009: 1035), that globalization is related to a lower percentage of women employed in the service sector in non-OECD countries.

Finally, we also show the impact of our globalization measures on different cohorts in Table 5 and Fig. 4. (Note that the vertical axis is differently scaled for the two panels in Fig. 4.) This means that we allow the parameter for the impact of our globalization variables to vary between age cohorts. The overall picture that emerges shows that the impact is stronger for younger cohorts. This corresponds to the rationale that more labor market variability in going on at younger age levels and that the income effect might be particularly strong at these cohorts when compared to the substitution effect: A potential rise in wages due to a globalization boost might increase household income via the father's or spouse's wage bill whereas the substitution effect between staying out of the labor force or joining it may even become negative in the short run since the skill-premia might have risen and this creates supplementary incentives to currently stay out of the labor force and invest in education, especially for young women where the premium pays off over a longer lifetime (Fig. 4).

[15]They are jointly significant on the 1 % level using an F-test but the magnitude of the effect is again small.

Table 5 Cohort-specific effects

	(1)	(2)
Variables	FLFPR	FLFPR
ln(GDP p.c. PPP)	−0.154*	−0.0433
(−1)	(0.0920)	(0.0902)
ln(GDP p.c. PPP)^2	0.0115	0.00506
(−1)	(0.00701)	(0.00676)
Fertility rate	−0.00707	−0.00247
	(0.00753)	(0.00710)
Years of schooling	0.00290	0.00612
	(0.00781)	(0.00738)
Agricultural value added	0.0448	0.0869*
	(0.0537)	(0.0512)
Industry value added	−0.0349	0.0152
	(0.0447)	(0.0413)
GDP growth rate	−0.0461	−0.0630*
	(0.0305)	(0.0378)
	Effect of FDI . . .	**Effect of Exports** . . .
. . . at age 15–19	−0.00845**	−0.254***
(−1)	(0.00415)	(0.0657)
. . . at age 20–24	−0.00702**	−0.159***
(−1)	(0.00280)	(0.0487)
. . . at age 25–29	−0.00542**	−0.0508
(−1)	(0.00223)	(0.0451)
. . . at age 30–34	−0.00542**	−0.0575
(−1)	(0.00227)	(0.0362)
. . . at age 35–39	−0.00476*	−0.0327
(−1)	(0.00250)	(0.0391)
. . . at age 40–44	−0.00505**	−0.0282
(−1)	(0.00239)	(0.0344)
. . . at age 45–49	−0.00468*	−0.0220
(−1)	(0.00274)	(0.0468)
. . . at age 50–54	−0.00528**	−0.0375
(−1)	(0.00244)	(0.0386)
. . . at age 55–59	−0.00503*	−0.00847
(−1)	(0.00256)	(0.0472)
. . . at age 60–64	−0.00622**	−0.0485*
(−1)	(0.00241)	(0.0259)
Constant	1.084***	0.600*
	(0.333)	(0.323)
Year dummies	Yes	Yes
Observations	3530	3580
R-squared	0.201	0.225
Number of cross-sections	800	800

Fixed effects regression taking every 5th year. Cluster-robust standard errors in parentheses
***, **, * denote statistical significance at the 1 %, 5 %, and 10 % level, respectively

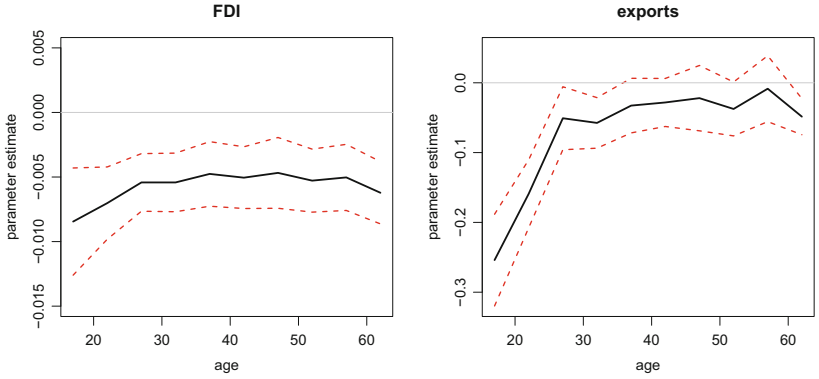

Fig. 4 Impact of globalization variables (±2 standard errors) by cohort

While the impact is still very small for FDI, the impact of exports is now considerable for young females' labor decision: The parameters are 0.254 and 0.159 for the age groups 15–19 and 20–24, respectively. A parameter of 0.2 would imply that a 10 percentage point increase in exports would result in a 2 percentage point decrease in the FLFPR, a non-negligible effect.[16] Note that the estimated impact is negative for all cohorts for both measures of globalization but not for all of them statistically significant in case of exports (the interval of ±2 standard errors roughly approximates a pointwise 95 % confidence interval).

As a robustness check, we investigate to what extent the obtained results change, when specifying another functional form of the model, namely a logarithmic model of the form

$$\log(FLFP)_{ijt} = Z_{ijt}\theta + X_{it}\beta + \alpha \log(FPOP)_{\text{ijt}} + u. \tag{5}$$

In our view, this functional form has the advantage that it is economically more appealing than the standard models in the literature because it allows for interactions of covariables and does not force the response to be linear in the latter. Second, the model in Eq. (5) is more flexible because it does not pose the implicit restriction $\alpha = 1$.[17] Third, the model in Eq. (5) avoids meaningless predictions of the response variable.[18] Finally, the model in Eq. (5) is not necessarily more difficult to interpret

[16]Remember from Table 5 that FLFPR increased by roughly 2 percentage points per decade.

[17]Note that if the restriction $\alpha = 1$ is indeed true, a restricted estimator will be more efficient than the model in Eq. (5). However, in the context of a sample as large as the present one, we find this to be of minor relevance though it may be important for policy making and evaluation on the country level when facing a much smaller sample.

[18]Note that a linear model like in Eq. (2) may lead to predictions of the FLFPR that are smaller than 0 or larger than 100 % which does not make sense economically. Since in the model $\mathbf{E}[\ln(FLFP)] = X\theta$, the predictor for FLFP is $e^{X\theta}$, which is a positive number for any value of $X\theta$, a meaningful prediction of FLFP is ensured.

because changes in any covariable can be interpreted as elasticity of FLFP (if the covariable is itself in logs) or as a percentage change in FLFP if the covariable changes by one unit (if it is not in logs).

The results from this exercise, reported in Table 8 in the appendix, are qualitatively similar to the ones obtained above. Parameter estimates for the interaction of FDI with industry value added and for FDI are statistically significant and highly significant, respectively, and are jointly highly significant (F-statistic of 5.53 with 2 and 79 degrees of freedom). While the prefix of the parameter estimates are the same as in Table 4, the positive impact of the interaction is more dominant over the negative impact of the FDI stock: A positive impact of FDI on FLFP occurs at a level of industrial value added above 16.2 % of GDP. More precisely, for an economy producing half or all of its output in the industrial goods sector, a 10 percentage point increase in FDI will cause a 0.8 or 2.1 % increase in FLFP, respectively. For this functional form, we also find that trade in services absorbs the negative impact of exports.

5 Comparison to Other Results in the Literature

To understand more clearly, why we come to other results in this study than the previous literature, we perform a series of functional re-specifications and replicate the results of Bussmann (2009), which we consider to be the study closest related to our investigation in terms of scope and methodology.

Diverging estimates might result, inter alia, from the fact that we use only every 5th observation year while other studies use annual data and hence rather capture short-run effects, from different panel data estimators, or from the fact that previous studies used FLFP data on the aggregated country level, while we use the country-cohorts as cross section.

The use of annual data leaves our main results qualitatively and quantitatively largely unchanged, although exports are no longer significant at the 10 % level. A change in the sampling period also does not seem to explain the difference, though there is some evidence of a (very small and statistically insignificant) positive impact of trade on FLFPR when we truncate the sample in year 2000.[19]

Furthermore, we added different weights to our cross-sections to rely more on observations representing larger cohorts and aggregated the data at the country level to be more in line with the methodology applied in previous studies. When using female population as analytical weightings and dummy variables for countries

[19]It is well-known that the pre-2000 era of the "Washington Consensus" was a period of considerable big-bang liberalization in many developing countries. It might hence be the case that this led to a big push in input demand in many countries which was satisfied by female labor. However, such a possibility would have to be investigated in more detail (and possibly only holds for a small set of specific countries) and should then rather be seen as a singularity instead of a general relationship between globalization and FLFP.

and cohorts,[20] most of the explanatory variables are insignificant, trade is negative (−0.11**) and significant on the 5 % level, FDI turns positive (0.014) but is very small and far from being significant. More generally, there is no evidence whatsoever that FDI would have a positive impact on FLFPR in different settings; the results are generally of the same magnitude as those reported above and mostly statistically significant (at least at the 10 % level). For trade/exports, the results suggest that the less one accounts for fixed effects at different levels, the more likely is the indication of a positive impact on FLFPR. The export coefficient gets as large as 0.12, still not an overwhelming magnitude, but is statistically insignificant throughout. This generally supports our claim that unobserved cross-country (and/or cross-region) heterogeneity is present in the relationship between globalization and FLFP and failing to account for this fact can lead to misleading inference about the economic causation in this regard.

This is also strongly supported by our intention to replicate the results of Bussmann (2009; see there for a detailed data description). Her dataset is somewhat different from ours, e.g. it ranges from 1960 to 2002, mainly comes from the World Bank WDI, and does not include FDI data. However, this difference in data does also allow us to check the robustness of our finding that the country context matters for the impact of globalization on FLFPR.

We focus on the estimation reported in Table 3, column (5), of her paper, addressing the determinants of FLFPR in non-OECD economies.[21] We were perfectly able to replicate her results which are reported in Table 6, column (1) of this chapter.[22] As can be seen, she finds a positive impact of exports on FLFPR. However, her GMM estimation is run without time fixed effects[23] and, more importantly, without country fixed effects, hence giving space for unobserved heterogeneity as described throughout this chapter.[24] After simply introducing these two-way fixed effects into her specification, we find that the effect of exports turns negative and insignificant, see column (2).

In order to be more in line with our framework, we then estimate the equation using FE with a 1-year lagged export/GDP covariable, first on a yearly basis, and then for every fifth year available in the sample. Again, the results (reported in columns (3) and (4)) show a negative effect of similar magnitude which is

[20] Estimation of country-cohort fixed effects is infeasible in this setting.

[21] For simplicity, we followed Bussmann's classification of countries into OECD and non-OECD countries and assume that the second category captures well what we consider as "developing countries."

[22] There is a minor difference in the constant but this can happen, for example, due to different versions of STATA.

[23] Including those year dummies is important, for example, to capture an underlying time trend in FLFPR that might be correlated with a "globalization trend" and to mitigate the simplest form of cross-sectional correlation (i.e. global shocks) that would plague statistical inference.

[24] In fact, a Hausman test on the difference between the FE and RE estimates as reported in columns 3 and 5 of Table 10 clearly rejects that these differences are random (on a 1 % level), providing very strong arguments in favour of including country FE.

Table 6 Replication and modification of Bussmann (2009)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR
LFPR(−1)								0.962***
								(0.00475)
log(exports/GDP)	**2.452****	**−0.467**						**−0.0967****
	(1.056)	**(0.447)**						**(0.0424)**
log(exports/GDP) (−1)			**−0.350**	**−0.477**	**−0.335**	**1.825***	**2.816***	
			(0.341)	**(0.420)**	**(0.360)**	**(1.005)**	**(1.604)**	
log(GDP p.c.)	−61.36***	−3.612	−3.870	−6.451	−4.420	−56.81***	−38.99***	−0.969
	(13.63)	(6.376)	(6.491)	(7.114)	(6.845)	(13.37)	(13.18)	(0.627)
log(GDP p.c.)^2	3.348***	0.174	0.189	0.347	0.229	3.118***	1.792**	0.0486
	(0.866)	(0.422)	(0.430)	(0.469)	(0.440)	(0.846)	(0.808)	(0.0386)
Political regime	0.0341	0.00630	0.00687	0.00471	0.0146	−0.00107	0.153	0.00160
Type	(0.0973)	(0.0313)	(0.0317)	(0.0334)	(0.0337)	(0.106)	(0.134)	(0.00328)
Fertility	−2.116***	−0.875***	−0.866***	−1.054***	−0.634*	−1.740**	−3.289***	−0.0729***
	(0.718)	(0.307)	(0.312)	(0.331)	(0.339)	(0.806)	(0.899)	(0.0276)
Female secondary	−0.0167	0.0302	0.0304	0.0302*	0.0399*	−0.0301	−0.00779	0.00186
Schooling	(0.0511)	(0.0190)	(0.0192)	(0.0158)	(0.0208)	(0.0514)	(0.0556)	(0.00164)
log(population)	−0.348	−11.08***	−11.03***	−10.21***	−3.820***	−0.460	−0.200	−0.0411**
	(0.557)	(3.126)	(3.155)	(3.292)	(1.285)	(0.528)	(0.601)	(0.0208)
Constant	316.5***		154.5***	168.6***	96.59***	294.0***	241.4***	7.165***
	(53.34)		(34.84)	(37.56)	(27.86)	(53.44)	(52.93)	(2.666)

(continued)

Table 6 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR	FLFPR
Estimation method	**GMM**	**GMM**	**FE**	**5-year FE**	**RE**	**POLS**	**BE**	**System GMM**
Year dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	**No**	**Yes**	**Yes**	**Yes**	**No**	**No**	**No**	**Yes**
Observations	2,419	2419	2421	527	2421	2421	2421	2443
R-squared	0.377	0.204	0.480	0.494		0.397	0.447	
Number of cross-sections			119	107	119		119	119

Robust standard errors in parentheses
***p < 0.01, **p < 0.05, *p < 0.1

statistically insignificant. However, the magnitude of the effect would imply that a 10 % increase in the export/GDP ratio (e.g. from 30 % to 33 %, this is another difference in the model to our specification) reduces FLFPR by 4 percentage points (considering a parameter of −0.4). This is a rather large magnitude compared to the results we obtained above; however, it is clearly more credible than the original result of the study, implying that the same change in the export/GDP ratio leads to a 24.5 percentage point increase in the FLFPR.

Finally, we also show the results for other panel data estimation techniques in the last columns of the table. Random effects (column (5)) and simple pooled OLS (column (6)) both identify over within-country and cross-country variation, whereas between-effects (column (7)) identifies purely over cross-country variation. In line with previous studies, we also find a positive impact when using simple OLS, similar to the GMM results of Bussmann (2009), which also identifies over both within-country and cross-country variation. Since random effects and pooled OLS can be interpreted as an average of FE and between-effects (cf. Maddala 1971), it is also clear from our table, where the positive effect of trade or exports previously found in the cross-country literature stems from: It is the variation *across* countries which drives this result. This clearly supports our point that omitted country characteristics, that influence both FLFP and globalization measures, drive the results in the previous literature and that if one starts controlling for these unobserved cross-country heterogeneity, one ends up, if anything, with a negative impact of globalization on FLFPR for most developing countries. This is also supported by the results from a dynamic System GMM estimation, reported in column (8). Despite some worrisome aspects of this specification, this framework has some advantages, such as allowing for dynamic effects and more convincing instruments,[25] and the results also indicate a statistically significant (but in the long run unconvincingly large) negative effect of exports on FLFPR.

6 Discussion and Conclusion

We have investigated the influence of globalization as measured by trade and FDI on the FLFPR in a panel of 80 developing economies over the time period 1980–2005. The results suggest that openness generally has a negative but small impact on

[25] Bussmann (2009) uses lags of the *levels* series as instruments which is not convincing if the series is weakly dependent, as is the case for trade/GDP data. This is also indicated by a worrisome Hansen J statistic (neither reported here nor in her paper). Instead, System GMM uses lagged *differences* of the series as instruments for current levels which can be shown to be valid instruments under certain assumptions (Arellano and Bover 1995; Blundell and Bond 1998). We instrumented the lagged dependent variable in a collapsed form and the export/GDP ratio with difference lags 2–4, also in collapsed form. The number of instruments (81) clearly outnumbers the number of cross-sections (119), as necessary; the (robust) Hansen statistic does not allow rejecting the null hypothesis that the whole set of instruments is valid (on the 10 % level). The z-statistics of the AR(1) and AR(2) tests are 0.49 and 0.65, respectively.

the FLFPR—which is in contrast to most previous studies that have generally found a positive effect. The main driver of these diverging results comes from the fact that previous studies failed to account for (potentially unobserved) cross-country heterogeneity which generates a positive correlation between openness and FLFPR *across* countries that does not exist *within* countries. As an additional improvement over the previous literature, we have shown that the effect is stronger for young women. We think this is driven by the flexibility in younger years and by the fact that the potential rise in the skill premium due to globalization creates a particularly strong incentive for younger women to invest in education (and to hence not join the labor force immediately) because the returns will be realized over a longer (expected) remaining lifetime. This has been evidenced for Turkey where the LFPR of illiterate women fell from 25.2 % in 2000 to 12.9 % in 2008 (Yenilmez and Isikli 2010). Both theoretical models and micro-econometric studies might help address this channel in the future.

A main takeaway from our study is that one should be very cautious in generalizing results from country-level studies to an overarching tale about the female labor market effects of globalization. First, we have shown that the effect, though being statistically significant, is negligible in economic terms. Also, the results presented in Table 3 show a large degree of regional heterogeneity. Our finding of a statistically significant positive effect of FDI on the FLFPR in Eastern Europe and Central Asia compared with a significant negative effect in Africa further supports our notion that the potentially increased skill-premium due to globalization/FDI creates incentives to build up human capital before joining the labor force: A high human capital stock (with relatively low gender inequality) was a heritage from the past in former centrally planned economies and would allow women to join the labor force and reap the benefits from an increased skill-premium right away, whereas female educational attainment is much lower in Africa (cf. Barro and Lee 2010: Table 3).

Since we show that the direction of the FDI impact on FLFPR depends on the size of the industrial/primary sector, our results strongly suggest that any economic explanation about the impact of globalization on FLFP has to take into account the industrial structure of the economy under consideration. Potential arguments could be built along the lines of a Lewis (1954) type labor market: In agrarian economies, a large pool of laborers is available. Since comparative advantages of these economies lies in sectors intensively using physical labor and surplus labor keeps wages low, multinational and exporting firms might be more likely to go for the "low hanging fruits" by drawing from the pool of male laborers. By still paying a somewhat higher wage (cf. Lipsey 2002), the income effect on the household level might then have a small negative impact on FLFP and the mainstream argument of a female-intensive comparative-advantage sector does not hold for these countries. The more industrialized a country becomes, the smaller the pool of (male) surplus labor becomes and multinational and exporting firms might hence demand more female labor, especially since the process of industrial development and the division of labor will create linkages with the service sector where female labor is not "physically disadvantaged" and gender wage gaps might

even provide an incentive to employ women, so that the mainstream arguments come into force at this development stage.

Our results can be seen in the context of the finding of Gaddis and Klasen (2014) that different industrial structures of the economy generate different dynamics for female employment. They also do not necessarily conflict with the results of previous case studies since they have been conducted in countries where industrial development was rather high compared with other developing economies which might have driven the results of these country-case studies.[26]

In terms of welfare and policy, our results of a generally negative effect of globalization on FLFPR is not necessarily bad news for women since their decrease in labor force participation might simply be the optimal response to benefit from an increased skill premium or because household income is sufficiently high and allows women to stay home if they want to. Indeed, Gray et al. (2006, pp. 317ff) find that trade (but not FDI) decreases female illiteracy rates for 180 countries (although the elasticity is rather small) and Bussmann (2009, p. 1032) also finds some evidence that women in non-OECD countries get more access to education when trade/GDP is growing, at least in primary and secondary schools. We show in simple regressions of female years of schooling on the globalization measures (reported in Table 9 in the appendix) that increased exports (for which we find a stronger impact on FLFPR than for FDI) are also positively correlated with female educational attainments in our sample. The parameter of 0.78 in the first column of the table would mean that women respond to a 10 percentage point increase in exports/GDP by staying 7.8 years longer in school—years they are generally absent from the labor force.[27]

Therefore, while our aggregate results challenge the viewpoint of a large fraction of the literature arguing that globalization generally has a positive impact on FLFP in developing countries, this does not mean that a negative relationship necessarily exercises an adverse impact on female well-being or empowerment, as we try to suggest with Aristotle's opening quote at the beginning of this chapter. However, problems may arise under bounded rationality, e.g. if women do not enter the labor force because family income is sufficient, but do neither engage in educational programs even though this will decrease their probability of finding a job in the future. If a shock occurs in the future (as is likely to be the case in open developing countries), and household income declines, females will find it more difficult to make up for this income decrease because of forgone job-market experience.

[26]The sampling period of Çağaty and Berik (1990) coincides with the time when Turkey reached the threshold level of industrial development of 28 % that we find in our study. Özler (2000) uses data from the mid-1980s when the size of the industrial sector in Turkey was about 27 % and hence close to our threshold of 28 % and clearly above the threshold of 16 % found in the multiplicative model. The data of Kabeer and Mahmud (2004) come from a 2001 survey when the industrial share made up for 26 % of the Bangladeshi economy. For Pradahn's (2006) study on India around 2000, industrial value added was always over 25 % of GDP (all sector data: WDI).

[27]We also include a Random Effects specification in column (2) to take into account variation *between* countries and hence a longer-run perspective.

An important policy implication stemming from this study is that countries that open up for globalization should tightly monitor developments on their female labor market. Long-term employability of women who leave the labor force because of sufficiently increased household earnings should be ensured. This may include continued education programs or offering more flexible working schedules.

Acknowledgements and Disclaimer We would like to thank Stephan Klasen, Jenny Aker, and other participants of a seminar at the University of Göttingen as well as an anonymous referee and the editors for their helpful comments. Special thanks go to Margit Bussmann for sharing her data and replication files with us. K.M. Wacker thanks the German Academic Research Foundation (DAAD) and the Chinese Scholarship Council (CSC) for financial support. All remaining errors are ours. The views expressed in this paper are those of the authors and do not necessarily represent those of any of their affiliations.

Appendix A

Countries Included:

Albania, Armenia, Bangladesh, Belize, Benin, Bolivia, Botswana, Burundi, Cambodia, Cameroon, Central African Republic, Chile, China, Colombia, Congo, Rep., Costa Rica, Cote d'Ivoire, Cuba, Dominican Republic, Ecuador, Egypt, Arab Rep., El Salvador, Fiji, Gambia, Ghana, Guatemala, Guyana, Honduras, India, Indonesia, Jamaica, Jordan, Kazakhstan, Kenya, Kyrgyz Republic, Lao PDR, Lesotho, Liberia, Malawi, Malaysia, Maldives, Mali, Mauritania, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Namibia, Nepal, Nicaragua, Niger, Pakistan, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Rwanda, Senegal, Sierra Leone, Slovak Republic, South Africa, Sri Lanka, Sudan, Swaziland, Syrian Arab Republic, Tajikistan, Tanzania, Thailand, Togo, Tonga, Tunisia, Turkey, Uganda, Ukraine, Vietnam, Yemen, Rep., Zambia, Zimbabwe.

Table 7 Summary statistics based on those observations included in model (5) of Table 6.

Name	Obs	Mean	Std. Dev.	Min	Max
FLFPR	3470	0.54	0.23	0.01	0.98
ln(GDP p.c. PPP)	3470	6.67	1.02	4.69	8.82
Fertility rate	3470	4.21	1.61	1.10	7.813
Years of schooling	3470	4.79	2.85	0.26	11.53
Agricultural value added	3470	0.24	0.14	0.02	0.72
Industry value added	3470	0.29	0.10	0.10	0.72
GDP growth rate	3470	0.02	0.05	−0.14	0.37
FDI stock/GDP	3470	0.26	0.52	$8.09e-06$	6.91
Trade/GDP	3470	0.77	0.39	0.11	2.20
Exports/GDP	3470	0.34	0.20	0.03	1.12
Trade in services/GDP	3220	0.18	0.15	0.02	2.06

Trade and Exports include goods *and* services

Table 8 Multiplicative model

	(1)	(2)	(3)
Variables	ln(FLFP)	ln(FLFP)	ln(FLFP)
Trade/GDP	**−0.00833**	**−0.0417**	
(−1)	**(0.0472)**	**(0.0444)**	
FDI stock/GDP	**−0.0164**	**−0.0402*****	
(−1)	**(0.0183)**	**(0.0124)**	
Industry v.a. × FDI		**0.248***	
(−1)		**(0.131)**	
Trade in services/GDP			**−0.0757***
(−1)			**(0.0436)**
Exports/GDP			**−0.000931**
(−1)			**(0.0961)**
ln(GDP p.c. PPP)	−0.343	−0.363	−0.363
(−1)	(0.251)	(0.245)	(0.273)
ln(GDP p.c. PPP)^2	0.0271	0.0289	0.0281
(−1)	(0.0192)	(0.0186)	(0.0209)
Fertility rate	−0.0142	−0.00740	−0.0123
	(0.0266)	(0.0248)	(0.0265)
Years of schooling	0.0114	0.0121	0.00674
	(0.0237)	(0.0234)	(0.0263)
Agricultural value	0.152	0.108	0.170
Added	(0.147)	(0.136)	(0.160)
Industry value added	−0.0672	−0.192	−0.0667
	(0.104)	(0.122)	(0.104)
GDP growth rate	−0.0799	−0.0367	−0.225*
	(0.110)	(0.107)	(0.128)
ln(FPOP)	1.208***	1.210***	1.232***
(−1)	(0.0494)	(0.0491)	(0.0535)
Constant	−2.276**	−2.223**	−2.474**
	(0.905)	(0.887)	(0.962)
Year dummies	Yes	Yes	Yes
Observations	3470	3450	3280
R-squared	0.864	0.864	0.867
Number of cross-sections	800	800	790

Fixed effects regression taking every 5th year. Cluster-robust standard errors in parentheses
***, **, * denote statistical significance at the 1 %, 5 %, and 10 % level, respectively

Table 9 Correlation between globalization and female education

	(1)	(2)
Variables	Years of schooling	Years of schooling
Exports/GDP	0.779**	0.918**
(−1)	(0.375)	(0.363)
FDI stock/GDP	−0.111*	−0.119*
(−1)	(0.0638)	(0.0625)
Constant	3.925***	0
	(0.121)	(0)
Time dummies	Yes	Yes
Estimation	Fixed effects	Random effects
Observations	3750	3750
Number of cross-sections	830	830

Cluster-robust standard errors in parentheses
***, **, * denote statistical significance at the 1 %, 5 %, and 10 % level, respectively

Appendix B

Table 10 Overview of related literature

Study	Data and coverage	Dependent variable	Globalization-related variables	Methodology	Disaggregations	Impact
Aguayo-Tellez et al. (2010)	Mexico, census, household and establishment surveys data (manufacturing), 1990–2000	Female employment rate, gender wage gap and female wage bill share (industry level)	Effective tariff rates and trade flows (industry-level), exports and FDI (plant-level)	Decomposition (between and within industry shifts)	–	Trade liberalization under NAFTA and FDI deregulation led to rising female employment
Baslevent and Onaran (2004)	Turkey, labor force survey data, 1988–1994	Women's labor force participation and employment decision (individual and plant level)	Overall and female-intensive export-orientation (share of (female) export-oriented sectors in manufacturing) (province-level)	Probit (with lagged macro-economic variables as regressors)	Short- vs. long-term, single vs. married women	Positive effect of export orientation on female labor force participation in the long-run (esp. young/single women), effect vanishes if one controls for GDP
Bussmann (2009)	134 countries (high income and developing), 1970–2000	FLFP and female employment by sector	Trade/GDP, Export/GDP, Import penetration (country-level)	GMM	Sector	Positive effect of trade on FLFP in developing countries (particularly via employment in agriculture and industry) but negative effect in OECD countries
Çağaty and Berik (1990)	Turkey, plant-level data (manufacturing), 1966–1982/85	Female share of employment (industry level)	Export-orientation, skill-intensity, labor-intensity (plant-level)	Pooled OLS	Economic policy regime	Export orientation increases female employment

(continued)

Table 10 (continued)

Study	Data and coverage	Dependent variable	Globalization-related variables	Methodology	Disaggregations	Impact
Çağaty and Özler (1995)	96 countries, 1985–1990	Female share of the labor force	Exports/GDP, Participation in structural adjustment programs (country-level)	Pooled OLS	–	Exports have a negative effect on FLFP (but sometimes insignificant), interaction between structural adjustment and exports has a positive effect on FLFP
Chamarbagwala (2006)	India, household survey data, 1983/94–1999/2000	Employment rate (at the level of demographic groups)	Net imports/Output (industry-level)	Decomposition (between and within industry shifts)	Sector and education	Trade liberalization increased the demand for skilled labor; trade in manufacturing has a negative impact on demand for female labor, but trade in services generated demand for female college graduates
Dell (2005)	Mexico, employment survey, 1987–1999	FLFP (state level)	Imports, Exports, FDI (industry-level)	Difference-in-difference	–	Trade liberalization increased FLFP in Central Mexico, no separate effect of FDI (but difficult to disentangle)
Ederington et al. (2010)	Colombia, plant-level data (manufacturing), 1984–1991	Female share of employment (plant-level)	Tariffs (industry-level)	OLS, logit (with tariff reductions as regressors)	Plant characteristics	Trade liberalization increased female employment

Gaddis and Pieters (in press)	Brazil, 1991 and 2000 census data, 1990–1998 tariff data	Women's and men's labor force participation, employment and unemployment	Tariffs (industry-level)	Difference-in-difference	Education, sector	Trade liberalization reduced female and male labor force participation and narrowed the percentage point gender gap in labor force participation and employment
Gray et al. (2006)	180 countries (high income and developing), 1975–2000	Female share of the labor force	Trade/GDP, FDI/GFCF[a] (country-level)	FE	–	Trade and FDI come out insignificant
Hyder and Behrman (2011)	Pakistan, historical census data and labor force survey data, 1951–2008	LFP gap (f-m)	Trade/GDP		–	Trade openness reduced the gap between male and female LFP
Meyer (2006)	120 countries, 1971–1995	FLFP	Trade/GDP, Exports/GDP, Trade volatility (in TOT), Commodity concentration, and a trade openness index (based on factor analysis), FDI/GDP	OLS (static and dynamic)	Income level and region	Positive effects of trade openness on FLFP in the static model and negative effects in the dynamic model, results differ by region and income-level (pos. effect in MICS), FDI is insignificant in the static model but has a positive effect in the dynamic model

(continued)

Table 10 (continued)

Study	Data and coverage	Dependent variable	Globalization-related variables	Methodology	Disaggregations	Impact
Özler (2000)	Turkey, plant-level data, 1983–1985	Decision to employ females and female share of employment (plant level)	Export-orientation, skill-intensity (plant-level)	OLS (on averages)	Plant characteristics	Export-orientation increases the likelihood to employ females and the female share of employment
Pradhan (2006)	India, plant-level data (manufacturing), 1999/2000–2001/2002	Employment gap (f-m)	Imports, Exports, In-house R&D, Foreign technology imports, Capital-intensity, FDI (plant-level)	Pooled OLS	–	Trade (via exports) increases female employment, technology upgrades are linked to lower female employment, FDI has an insignificant effect
Siddiqui (2009)	Pakistan, 1990	FLFP	Average tariffs (industry-level)	CGE Model	Skilled vs. unskilled	Trade liberalization leads to higher FLFP (mainly unskilled women)
Siegmann (2007)	Indonesia, household and plant survey data, 1999–2002	Female employment share	Foreign capital in a firm's total capital stock	OLS and qual. focus group discussions	Sector	Qualitative interviews show positive effects of FDI on female employment, quantitative analyses show mixed results (negative effects in manufacturing/hotels)

Study	Data and coverage	Dependent variable	Globalization-related variables	Methodology	Disaggregations	Impact
Terra et al. (2007)	Uruguay, 2000	Female employment	Tariffs (by sector)	CGE Model	Skilled vs. unskilled	Trade liberalization has a positive effect on female employment (skilled women faring better)
Wood (1991)	52 countries (high income and developing), 1960–1985	Female share in manufacturing employment	Manufacturing export ration, Import penetration	Descriptive statistics, scatter plots (first differences)	High income vs. developing	North-South trade has increased the demand for female labor in the manufacturing sector in developing countries

[a]*GFCF* Gross fixed capital formation

Appendix C: Error Structure of the Model

A concern of our model is the correlation structure of the idiosyncratic error ε in Eq. (4). Despite using a 5-year interval, autocorrelation is one potential issue. Together with potential heteroscedasticity, this can easily be accommodated by using the heteroscedasticity and autocorrelation (HAC) robust approach of Huber (1967) and White (1980) to estimate the variance-covariance (VCV) matrix. However, the hierarchical structure of our model (cf. Wooldridge 2003 and 2010: ch. 20 for an introductory treatment to such models) poses additional problems since, for example, the error ε_{ijt} is likely to be correlated with the error $\varepsilon_{i,j+1,t+1}$ because the individuals in cohort j in period t will be in cohort $j+1$ in period $t+1$. Furthermore, there might be correlation between all errors $\varepsilon_{.jt}$ within country i if there is a systematic measurement error on the country level. All these potential problems with standard inference in linear models point to different forms of error correlation *within* countries. In line with the conventional panel data literature and given the dimension of our data set, we can assume that $N \to \infty$ and hence the number of countries, which are considered to be the "clusters," is large while the size of these clusters (i.e. the cohorts by country) is small. As discussed in Wooldridge (2003: 134, see also 2010: 864ff) a robust estimate for the VCV matrix is obtained by clustering the errors on the country level. Assuming that the matrix W_i contains all fixed effects and explanatory variables, classified as X and Z above, for country i and that the corresponding parameter vector δ contains β, θ, μ, and γ, a robust VCV estimator for δ is given by

$$\widehat{VCV}\left(\hat{\delta}\right) = \left(\sum_{i=1}^{N} W_i' W_i\right)^{-1} \left(\sum_{i=1}^{N} W_i' \hat{\varepsilon}_i \hat{\varepsilon}_i' W_i\right) \left(\sum_{i=1}^{N} W_i' W_i\right)^{-1}, \qquad \text{(A1)}$$

where $\hat{\varepsilon}_i$ is the $10\,T \times 1$ vector of residuals for country (i.e. cluster) i.[28] Using time-fixed effects is important in this context because it prevents the most likely form of cross-section, i.e. contemporaneous, correlation of the error term. We want to emphasize that clustering the errors at the country level has a tremendous impact on inference, as one would expect (cf. Wooldridge, 2010: 865). If one would (wrongly) cluster on the country-specific cohort level instead, which would be the standard option in most econometric packages, standard errors would be severely underestimated (cf. Table 8 in the appendix to the Working Paper version of this chapter).

[28] An alternative approach would be using some feasible generalized least squares (FGLS) model. Depending on the assumptions, this might provide statistically more efficient results; it is, however, computationally less efficient. We hence prefer our approach because we find the assumptions less demanding and in the worst case, our framework will provide conservative inference compared with potentially efficient FGLS results.

References

Aguayo-Tellez, E., Airola, J., & Juhn, C. (2010). Did trade liberalization help women? The Case of Mexico in the 1990s, NBER Working Paper No. 16195, Cambridge, MA.

Anderson, E. (2005). Openness and inequality in developing countries: A review of theory and recent evidence. *World Development, 33*, 1045–1063.

Anker, R. (1998). *Gender and jobs sex segregation of occupations in the world*. Geneva: International Labour Office.

Anker, R., Melkas, H., Korten, A. (2003). *Gender based occupational segregation in the 1990s*. International Labour Office Working Paper No. 16.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error components models. *Journal of Econometrics, 68*(1), 29–51.

Baltagi, B. H., & Griffin, J. M. (1984). Short and long run effects in pooled models. *International Economic Review, 25*, 631–645.

Barro, R., & Lee, J.-W. (2010): A new data set of educational attainment in the world. NBER Working Paper No. 15902, Cambridge, MA.

Baslevent, C., & Onaran, O. (2004). The effect of export-oriented growth on female labor market outcomes in Turkey. *World Development, 32*(8), 1375–1393.

Berik, G. (2000). Mature export-led growth and gender wage inequality in Taiwan. *Feminist Economics, 6*, 1–26.

Blau, F. D., & Kahn, L. M. (2007). Changes in the labour supply behaviour of married women: 1980-2000. *Journal of Labour Economics, 25*, 393–438.

Bloom, D., Canning, D., Fink, G., & Finlay, J. E. (2009). Fertility, female labour force participation, and the demographic dividend. *Journal of Economic Growth, 14*, 79–101.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–143.

Blundell, R., & MaCurdy, T. (1999). Labor supply: A review of alternative approaches. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (1st ed., Vol. 3, Chap. 27, pp. 1559–1695). Amsterdam: Elsevier Science.

Boserup, E. (1970). *Woman's role in economic development*. New York: St. Martin's Press.

Bussmann, M. (2009). The effect of trade openness on women's welfare and work life. *World Development, 37*(6), 1027–1038.

Çağaty, N., & Berik, G. (1990). Transition to export led growth in Turkey: Is there a feminization of employment? *Review of Radical Political Economics, 22*, 115–134.

Çağaty, N., & Özler, S. (1995). Feminization of the labour force: The effects of long-term development and structural adjustment. *World Development, 23*, 1883–1894.

Chamarbagwala, R. (2006). Economic liberalization and wage inequality in India. *World Development, 34*(12), 1997–2015.

Cho, U., & Koo, H. (1983). Economic development and women's work in a newly industrialising country: The case of Korea. *Development and Change, 14*(4), 515–531.

Dell, M. (2005). Widening the border: The impact of NAFTA on female labor force participation in mexico. In *Presentation at the 2006 IZA/World Bank Conference 'Employment and Development'*.

Ederington, J., Minier, J., & Troske, K. (2010). *Where the girls are: Trade and labor market segregation in Colombia*. Lexington: University of Kentucky.

Egger, P., & Pfaffermayr, M. (2005). Estimating long and short run effects in static panel models. *Econometric Reviews, 23*, 199–214.

Gaddis, I., & Klasen, S. (2014). Economic development, structural change and women's labour force participation: A Re-examination of the feminization U Hypothesis. *Journal of Population Economics, 27*, 639–681.

Gaddis, I., & Janneke, P. (in press). The gendered labor market impacts of trade liberalization: Evidence from Brazil. *Journal of Human Resources*.

Goldin, C. (1990). *Understanding the gender gap. An economic history of American women*. New York: Oxford University Press.

Goldin, C. (1995). The U-shaped female labor force function in economic development and economic history. In T. P. Schultz (Ed.), *Investment in women's human capital and economic development* (pp. 61–90). Chicago: University of Chicago Press.

Gray, M., Kittilson, M., & Sandholtz, W. (2006). Women and globalization: A study of 180 countries, 1975–2000. *International Organization, 60*(2), 293–333.

Greenhalgh, S. (1985). Sexual stratification: The other side of 'Growth with Equity' in East Asia. *Population and Development Review, 11*, 265–314.

Hein, C. (1984). Jobs for the girls: Export manufacturing in Mauritius. *International Labour Review, 123*(2), 251–265.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 221–233). Berkeley, CA: University of California Press.

Hyder, A., & Behrman, J. R. (2011). Schooling is associated not only with long-run wages, but also with wage risks and disability risks: The Pakistani experience. *Pakistan Development Review, 50*(4). Papers and Proceedings Parts I and II: The 27th Annual General Meeting and Conference of the Pakistan Society of Development Economists Islamabad (pp. 555–572).

ILO. (1985). *Women workers in multinational enterprises in developing countries*. Geneva: International Labor Organization.

ILO. (1990). *Surveys of economically active population, employment, unemployment and under-employment. An ILO Manual on Concepts and Methods*. Geneva: International Labour Organization.

ILO. (2009). *ILO estimates and projections of the economically active population: 1980-2020* (5th ed.). Geneva: International Labour Organization.

ILO (2013). Resolution I. *Resolution concerning statistics of work, employment and labour underutilization*. Geneva: International Labour Organization.

Kabeer, N., & Mahmud, S. (2004). Globalization, gender and poverty: Bangladeshi women workers in export and local markets. *Journal of International Development, 16*, 93–109.

Killingsworth, M. R., & Heckman, J. J. (1987). Female labor supply: A survey. In O. Ashenfelter & R. Layard (Eds.), *Handbook of labor economics* (1 st ed, Vol. 1, Chap. 2, pp. 103–204), Amsterdam: Elsevier Science.

Klasen, S., & Lamanna, F. (2009). The impact of gender inequality in education and employment on economic growth: New evidence for a panel of countries. *Feminist Economics, 15*(3), 91–132.

Lele, U., & Mellor, J. W. (1981). Technological change, distributive bias and labor transfer in a two sector economy. *Oxford Economic Papers, 33*(3), 426–441.

Lewis, A. (1954). Economic development with unlimited supplies of labour. *Manchester School, 22*(2), 139–191.

Lipsey, R. E. (2002). *Home and host country effects of FDI*. NBER Working Paper No. 9293, Cambridge, MA.

Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica, 39*(2), 341–358.

Mammen, K., & Paxson, C. (2000). Women's work and economic development. *Journal of Economic Perspectives, 14*(4), 141–164.

Mehra, R., & Gammage, S. (1999). Trends, countertrends and gaps in women's employment. *World Development, 27*, 533–550.

Mellor, J. W., & Lele, U. (1973). Growth linkages of the new foodgrain technologies. *Indian Journal of Agricultural Economics, 23*(1), 35–55.

Mellor, J. W., & Lele, U. (1975). Interaction of growth strategy, agriculture and foreign trade-the case of India. In G. S. Tolley & P. Zadrozny (Eds.), *Trade, agriculture, and development*. Cambridge, MA: Ballinger.

Meyer, L. B. (2006). Trade liberalization and women's integration into national labor markets: A cross-country analysis. *Social Indicators Research, 75*(1), 83–121.

Oostendorp, R. H. (2009). Globalization and the gender wage gap. *World Bank Economic Review, 23*, 141–161.

Özler, S. (2000). Export orientation and female share of employment: Evidence from Turkey. *World Development, 28*, 1239–1248.

Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics, 68*, 79–113.

Phillips, P., & Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal, 6*(1), 217–259.

Pradhan, J. P. (2006). How do trade, foreign investment and technology affect employment patterns in organized Indian manufacturing? *Indian Journal of Labour Economics, 49*(2), 249–272.

Sauré, P., & Zoabi, H. (2014). International trade, the gender wage gap and female labor force participation. *Journal of Development Economics, 111*, 17–33.

Seguino, S. (2000). The effects of structural change and economic liberalisation on gender wage differentials in South Korea and Taiwan. *Cambridge Journal of Economics, 24*, 437–459.

Sen, G. (1999). *Gendered labour markets and globalisation in Asia*. UNCTAD/UNDP Occasional Paper No. 133.

Siddiqui, R. (2009). Modeling gender effects of Pakistan's trade liberalization. *Feminist Economics, 15*(3), 287–321.

Siegmann, K. A. (2007). Globalisation, gender, and equity: Effects of foreign direct investment on labour markets in rural Indonesia. In *Proceedings of the German Development Economics Conference*. Research Committee Development Economics, No. 30. Göttingen: Verein für Socialpolitik.

Standing, G. (1989). Global feminization through flexible labor. *World Development, 17*(7), 1077–1095.

Terra, M., Bucheli, M., & Estrades, C. (2007). Trade openness and gender in Uruguay: A CGE analysis. Poverty and Economic Policy Research Network Working Paper No. 2008–16.

Wacker, K. M. (2013). *On the measurement of foreign direct investment and its relationship to activities of multinational corporations*. ECB Working Paper 1614, European Central Bank.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica, 48*, 817–830.

Wood, A. (1991). North-South trade and female labour in manufacturing: An asymmetry. *Journal of Development Studies, 27*, 168–189.

Wood, A. (1998). Globalisation and the rise of labour market inequalities. *Economic Journal, 108*, 1463–1482.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review, 93*(2), 133–138.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.

World Bank. (2011). *World development report 2012. Gender equality and development*. Washington, DC: World Bank.

World Bank. (2012). *World Development Report 2013. Jobs*. Washington, DC: World Bank.

Yenilmez, F., & Isikli, B. (2010). The comparison of labour force participation rate of women in Turkey with the world country groups. *Anadolu University Journal of Social Sciences, 10*, 77–92.

Part VII
Globalization: What We Know and What We Would Like to Know

Business on Globalization—A Panel

Jørgen Vig Knudstorp, with Keith Maskus, David Teece, Bent Jesper Christensen (Moderator)

Abstract This chapter is based on a transcription of the panel "Business on Globalization," moderated by Bent Jesper Christensen, Aarhus University, on November 10, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark. Introductory remarks by Jørgen Vig Knudstorp are in Sect. 1, remarks on globalization, trade, power, culture, digitalization, demography, and planet issues in Sect. 2, on free trade, complexity, cycles, and inequality in Sect. 3, and on operating models, location decisions, and marketing in Sect. 4. Questions by panelists Keith Maskus and David Teece are in Sect. 5, along with replies by Jørgen Vig Knudstorp.

1 Introduction

It is a big honor to be here. There are numerous people here that I have quoted in the various papers that I wrote through my student life and also my Ph.D., so it is a big honor to be here and to be able to speak to you. Globalization is also a topic that I am extremely engaged in, so I am happy to share my thoughts, in the humble hope that I will get some reactions. Otherwise, I will not learn anything from this, as you know.

http://econ.au.dk/research/other-research-entities/cap-cycles-adjustment-and-policy/past-events/conference-on-globalization-strategies-and-effects/.

J.V. Knudstorp
CEO, LEGO Group, Billund, Denmark

K. Maskus
University of Colorado, Boulder, CO, USA

D. Teece
University of California, Berkeley, CA, USA

B.J. Christensen (Moderator) (✉)
Aarhus University, Aarhus C, Denmark
e-mail: bjchristensen@econ.au.dk

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_23

But I think it is a very important topic, maybe one of the most important topics for business at all. I will of course try to give sort of a general business perspective, but I will probably give very much a LEGO Group perspective, because that is my world.

So let me just say a few words for those of you who are not familiar with the company. It is a family-owned business, through 80 years, and if you have been a family-owned business for 80 years, I think there is less than a 1 % chance of survival. If you have also never made an acquisition, or been sold, I think it is less than a 0.01 % chance of survival, so we are very far out in the tail of possible survival. We have also had to experiment a little with whether it is worthwhile surviving, a couple of times.

It is a family-owned business. I am the fourth CEO in 80 years, and I am not a family member. I am the first one to run the company from the outside. It is currently held by the third generation owner, my predecessor. The main product of the company, the only product of the company, is of course the LEGO brick, which I hope you know. We make 65,000 LEGO bricks every minute, 24 h a day, in a few manufacturing facilities around the world. Those of you who are very astute on your calculus will realize that we are into the billions, actually, well above 50 billion pieces that we produce every year. We largely make them in three facilities. We ship them in around 200 million product units. They are sold in 130 countries all over the world. Many of them have to be delivered within a time frame of a few hours in a specific distribution center, or to a customer that can be located anywhere all over the world, although 80 % are located in Western markets.

If there ever was any level planning of economics Stalinistic arrogance, it is the people in LEGO who deliver product supply, because they basically make the invisible hand extremely visible, by deciding what elements to produce the next few hours. The secret to solving this economic problem of allocation of bricks? There are 7000 variants of bricks in any given year's assortment. Many of those bricks enter many of the same boxes, and of course there are economies of scale in keeping a mold, that is, the tool that produces the plastic brick, running on a machine without interruption, so you have to optimize the number of interruptions and change-overs on the molding machines. You have to schedule maintenance, and you also have to decide when you pack the bricks. You pre-pack them in plastic bags, and then you final-pack them from plastic bags into boxes. Boxes go into cardboard outers, cardboard outers go into distribution centers, they go from distribution centers to local distribution centers, and into customer distribution centers, and out to customer shelves.

To optimize that value chain, we use the SAP system. It is a German enterprise economic system. Once you have bought into it, you will never get out of it again. Fortunately, the company is now run by a Dane. That makes it a little bit easier to have a dialogue with them, but it is a German company, and I do not offend anybody when I say that a lot of people in the finance community call SAP Hitler's revenge. Sorry.

But that makes LEGO an extremely interesting business. It is asset-intensive, but not overly so. Those who are close to my work, or were, when I did my Ph.D.,

of course claim that everything I do daily on my job is what was in my Ph.D. dissertation, which is sort of true. But it also has to do with globalization. So with that introduction to the company, maybe we should just say it is also about a 3 billion Euro business, 10,000 employees, and we are probably the leading brand in our business. We are definitely the most profitable, and some people say we are the Apple of the toy industry, but we are still only number three globally, and with a market share of no more than 7 %.

2 Globalization

2.1 Globalization, Trade, Power, and Culture

In terms of globalization, the first thing I want to say is that there is nothing new about globalization and global trade, also not for LEGO. LEGO does less than 1 % of its business in its home country of Denmark, but if you read Roman history, or the history of the British Empire, you realize that they were all about trading, and the fact that the navy was strong allowed Britain to trade. What is important to us in that aspect is that trade and trade-supremacy always lead to power, and power always leads to cultural dominance.

So what is the world's leading innovator in toys today? It is Japan, because Japan of course since long ago established itself as a leader of the industry, whereas today 80 % of the global manufacturing of toys is in China, of course. If you go 50 years back in time, that was in Nüremburg, Germany. So in that sense, for us, it is of course of relevance to understand trade patterns, because they basically drive the power structures and the cultural influences. So that is one thing.

2.2 Globalization and Digitalization

I would not argue that there is nothing new about globalization. I think that when we talk about something new about globalization, there are two topics that make globalization different than at the time of the British Empire, if you like, and those are two very different things. One is digitalization. I think digitalization probably really started in 1978, that is, when I got my first microcomputer, at the age of 10, and I think what has happened since then is of course that the world has moved from many alphabets to one. It is the alphabet of binary language, and that basically means that the ability to collaborate and communicate has increased dramatically, and also the transaction cost has declined very rapidly, and the death of distance has occurred, to some extent. I will come back to that specific aspect.

The other aspect of digitalization is of course SAP. That is why I mentioned it. You can say that a lot of the global competitive advantage of LEGO is not to

mold one brick. That is very easy to copy. But to produce 7000 different pieces in 40–50 billion prints a year, in a completely consistent quality, and then to distribute them optimally, so that you optimize your logistics costs, your total manufacturing costs, and of course are still able to deliver superior service to your retail customers all over the world, is a product of the digital revolution. And it is a massive cost optimization opportunity for any company.

Then digitalization is changing the way we communicate with people around us. There has always been a big community of LEGO fans who are adults. We speak about maybe a million people. They have consumption. You know, the difference between men and boys is the cost of their toys. These people were loosely organized before the internet arrived, but about 11 years before Facebook emerged, there was something called the LUGNET, the LEGO User Groups on the net, and those were people who were LEGO fans, who got together and shared ideas and so on, and they are still extremely active.

The third aspect of digitalization is of course products, and LEGO is today the world's second largest producer of video games for families with children. The other one is Nintendo. We do believe that is a massive impact, because children increasingly spend their time online, instead of in front of a TV. So that is one thing that I really think is changing globalization dramatically, and of course the problem with digital products, as you economists are well aware of, is that the marginal cost is zero, so it is very, very hard to charge for digital products and services, because everybody knows that the equilibrium is zero. So you have to charge in some pretty strange pricing models, like when we all go onto Google, and we say: "Wow, it's free." Well, it is owned by one of the world's most profitable companies, so it is probably not so free, which has to do with that this Christmas, when you go on your mobile phone and say "LEGO Star Wars," for your children or your grandchildren, I will have paid Google to make sure that you land on my page.

So that is the digital revolution. And I feel that sometimes it is dramatically overlooked how important digitalization is, and how it is completely changing society, how it is changing industries. Just look at music, movie distribution. DVD sales are falling. We are talking about such sweeping changes that I think it is hard to imagine. I was told the other day in Boston that the only thing that will parallel digitalization the next 100 years is innovations in life sciences, but I do not know anything about that, so I will not say a lot about it.

2.3 Globalization and Demography

What I will say is that the second major thing that makes globalization very different today are the demographics, and here I am happy to report that Demography was my first subject when I joined the Aarhus University Department of Economics and Business Economics. Of course, like everybody else, I did not understand why Demography was so important. Today I am surrounded by people who also still do not get it. But there is something very important about demographics, and that is

to understand the difference between stocks and flows. If there is a very slow exit from the stock, then the total population will be growing. Of course, you know that we passed 7 billion people. We will be 9 billion before I die, I hope, and that means there will be 8 billion more than there were 1000 years ago, which is a pretty fast span of time in human history.

That is one thing that is driving globalization. The other aspect of it, which is sometimes forgotten, is that in the countries where we have slow growth, we have aging, or, that the population is getting really old, so we are seriously now talking about developing products for people who are retired and want to look for a technical hobby, so we are now marketing to dads and granddads in Germany who build really complicated LEGO sets, and it is actually working.

Why are we talking so much about globalization? I actually think that the most important thing is digitalization. If there were not digitalization, China would not be able to enjoy the success it has the past 25 years, and then there are the demographics, of course.

2.4 Globalization and Planet Issues

This leads to a third important trend of globalization, which is of course extremely important to business, and it is the host of planet issue we face. I just heard on the radio yesterday that greenhouse gas emission increased by 6 % last year, maybe because we were in a crisis. The climate issue is a huge one, of course, of globalization, and for anyone who travels to one of these rapidly growing economies, that is pretty evident. I still have not seen the clear sun in Beijing in all the times I have been there. So it is a huge issue, and a huge externality, of course, as we would call it, and today companies are not paying for it, really, for a number of reasons. I think that in the future we will have to pay for it.

I was just sharing with David Teece that I am changing an organization right now, and the three reasons for changing an organization that has tripled in size in 5 years, and is enjoying 30 % growth also this year, is simply that we are not ready for these three major events: Digitalization, globalization, and the planet issues. We are not ready, and if we do not adapt our company to that, I will probably not be sitting here in 10 years.

3 Globalization and Effects

3.1 Globalization and Free Trade

A little bit about, then, second chapter, if you like, the impact of globalization on our business. A number of things come to my mind. I will just mention the most

important ones. We have a lot of discussions about trade agreements. Is NAFTA a good idea? Was the EU a bad idea or not? I will just say that from our perspective it is a fantastic idea. You will not be surprised to hear that. And we are predominantly manufacturing in Mexico and Eastern Europe, to be able to service our major markets in North America and Europe. The reason is that government is fairly reliable, crime is not bigger than you can manage it, rule of law is established, there is some degree of stability of the overall economy, and these things are actually extremely important to us.

The second aspect, which is very curious, is that while of course there is free arbitrage on finance and such areas in a globalized world, there is of course not free arbitration, if you like, or labor arbitrage is of course still possible, with huge profits. People are moving jobs around all over the world. For instance, in our factory in Hungary, which is literally 50 km from the border of Ukraine and 70 km from the border of Romania, is one of the best IT universities in Central Europe, and we are moving a lot of IT services to that region, because we can frankly recruit people who are better educated than the people we are able to attract here in Jutland, Denmark. Of course, the cost is significantly lower.

But I will say that a lot of work has been moved around the world for the wrong reasons. Companies tend to look at this from an accounting profit point of view. Either they are irrational, or stupid. But more likely they are trying to please their external shareholders, which is a privilege I do not enjoy. I have a family as an owner, so I have of course with my background made sure that we base our decisions on economic profit, economic value, and there is a huge difference between making labor arbitrage from those two angles, because when you use economic profit, one of the most important components is cost of transport, and also cost of missed sales, or opportunity costs. In our case, and this is maybe very much a LEGO angle, but we are not the only ones who are dependent on Christmas, it is so that 50 % of our annual sales happen in 4 weeks. Not for us, but for our customers, and therefore the marginal profit of being able to ship within 2 days instead of 8 weeks is pretty high, as you can imagine, and so if you do that calculation, then you wonder why anybody ever moved to China to service European or North American markets. That is why we are also not there. There are other issues with China that have to do with intellectual property rights, and that the rule of law is even more funny than in Hungary.

3.2 Globalization and Complexity

Then I will say that the other thing that comes with globalization is a huge complexity of the management and allocation tasks. That is also why I mentioned SAP again. That is actually fairly easy, because it is a straightforward technical problem to solve, or, straightforward is maybe a big word, but it is a problem that can be solved. Somebody suggested that a company's complexity is a product of the number of different stock keeping units it produces, the number of people it

employs, the number of legal units it has, and so on. If you make one product in Denmark, and you employ one person, your job is not very complicated, but as you spread out all over the world, complexity increases, your mind is a little bit blowing up, because you are trying to understand what happens in China, but you are hearing that Turkey is now growing really fast, so you need to go to Istanbul, and of course you do not learn anything about Turkey from being in Istanbul, and so you have all of these things, so you need to build some operating models. I will come back and talk about how you can try to solve that complexity challenge in a moment, but it is a big one.

3.3 Globalization and Cycles

Then another topic about globalization's impact that we are discussing a lot is that it seems that maybe the last 30 years were not normal, especially if we look at the amount of leverage by governments and companies. There was a period where people literally went to talks like this, maybe not in front of you, because you would not allow them, but in other audiences, where they would say that basically capital is free, liquidity is abundant, and labor is infinite, it will remain at a low cost for a long time, and raw material prices are declining when adjusted for inflation, or especially when calculated in adjusted currencies, and of course I think now we may look at those 30 years as an anomaly. Actually, we are now facing a period with increasing raw material prices, increasing labor costs, especially out of China, and of course it is very hard to find the amount of leverage that was possible before.

What we are really trying to build scenarios for, and we do not know where it is going, I am curious if you do, is, are we headed for a period of inflation, because of quantitative easing? Of course, even though the US economy is growing, productivity is going nowhere, so all the growth in the US, maybe even for a long time, is all inflation, so maybe the US did not grow any more than Japan, especially if calculated from a Japanese perspective, given the appreciation of the Yen, so we are a little bit unsure about where we are on that one. Or, will we come into a period of deflation or deleveraging, a good old Irving Fisher? But that is a big issue for globalization, because the engine of globalization has been this fact that labor was nearly free, cost of capital was very low, and there was an abundance of raw materials, and that moment seems to have passed. Maybe the next 30 years will be different from the previous 30 years.

The other thing we know about the next 30 years is that 80 % of our markets today will grow 1–2 %, but the world economy should be growing 4 %, and so 80 % of the growth that we can have if we are somewhat correlated to GDP, which we do think we are, that happens in a part of the world where we are very small today. That is probably true for a number of businesses, and a number of businesses are struggling with sort of de-averaging these numbers and figuring out where is the exact opportunity for us, and what is the go-to-market model for that?

3.4 Globalization and Inequality

Then a final statement about the impact of globalization, maybe the one that I am personally the most concerned about, is the growing inequality. I come, of course, from a country that is extremely high in equality, but even here, inequality is growing. The impact it is having is on education, where I think the education system is really struggling in many places to deliver in the more advanced areas. We see it everywhere we go, that we need to train people very much, for them to be able to deliver and carry their wages, and we see then that there is a small group of the labor force that becomes an elite, that is, to use your terms, they carry a lot of rent, because they become very specialized in their skill, or their ability, so they become a relatively unique resource. The trend is basically that companies are paying very large sums of money for that kind of specialist people, and increasingly paying in real terms a flat wage to a lot of other people, and I think that business is driving a trend that will harm society long term, because when people are not able to sustain good living, I think eventually equality and education are going to suffer.

4 Globalization and Strategies

4.1 Globalization and Operating Models

So, those are the most important impacts that we have noted, as for globalization, and then just what are we doing to address it? I will say something about that, in three areas. One is: How do we set up our operating model? One thing we do, of course, is that we build a balanced network. We operate in all regions, so we can follow currency fluctuations, earth quakes, tsunamis, whatever happens. We have the ability to scale the business up and down, depending on where we are. We spend a lot of time speculating about what is the right operating model.

If it is a little unclear to you what we are talking about when we say operating model, let me give some examples. Audi is one of the most successful brands and strengths of the Volkswagen car business. It is very much based on their strength in China. Basically, the Audis sold in China look very similar to the Audis sold in the rest of the world, but if you think about Kentucky Fried Chicken in China, it is owned by the same company that owns them in other countries, but it is a completely different food concept. So their model for succeeding in China is of course completely different, and the question is, when you develop your business on a global scale, do you try to do something like that?

You have probably all flown Ryan Air: They only operate one type of aircraft, train everybody to the same, buy spare parts for the same kind of aircraft, and then turn around their assets a lot faster, because they do three flights a day where other people do two, I don't know, something like that. And so how does that translate?

The key point for business then becomes that the only way to become competitive globally is to focus on what makes you really unique.

I always use the example of the hair dresser in the little town of Billund, where we are based. If you are a good hair dresser in Billund, you are competitive within a range of 10 miles, maybe, and that's it, but that is not good enough for LEGO, because we try to sell in Beijing and Moscow, as well, so to really do that, we have had to focus our business very dramatically, to sustain our advantage. The response of most companies to this in their operating model is to integrate horizontally, so they skip the vertical chain, and focus on one element of the horizontal chain, and then they try to become extremely good at that, like Walmart that focuses on executing the retailer, 1500 stores around the US, building similar types of stores in other places in the world, but sticking to that step of the value chain. A very purposeful company. You can argue whether you like it or not, but when I meet with Walmart and offer them a soft drink, they will put a dollar bill on the table, to make sure that I did not load cost into their supply chain, so that they can keep their prices low. And that, I can tell you, is not the response of a French retailer. If you offer them a dinner, they will ask for a hotel. If you give them a 5 % discount on a product, they will cash it in for themselves, whereas in Walmart a 5 % discount actually goes to the consumers, so it is a very effective execution-driven engine. Another company that is extremely horizontally integrated and very focused would be a consulting company like McKinsey, which is very knowledge-driven, obviously so a very different operating model than Walmart, but same logic.

Our operating model is different, maybe a little bit different than most, because of the vertical integration, and the reason is really in the IT system, where the only way that we can optimize the value chain cost is to control everything, from raw materials and even to some stores and theme parks we run. LEGO brand is to be able to optimize our system, but we would never make an acquisition of a competitor, because that would distort our uniqueness, and the brand position. So that is one of the responses to the complexity that I was addressing.

4.2 Globalization, Location Decisions, and Marketing

Another aspect of this is how to make location decisions, and is it possible to train labor to the same levels of effectiveness? It is often argued that China is cheaper, but then they are not as good as Danes, because Danes are very clever, obviously, we all know that. I will just say that there is absolutely no evidence of that. If you look at a factory, or any operation of your business, it is a product of national culture, it is a product of company culture, and it is a product of professional culture. A month ago, I opened Latin America's biggest plastic-molding facility in Monterey, Mexico, surrounded by armored cars, because there is unfortunately a lot of crime. But it is a fantastic factory. We used to run factories like that in Switzerland, which the Swiss consider similar to factories making watches, and other nice examples of Swiss precision engineering. It now happens in Monterey, Mexico, with completely

local staff. It is a super-clean, super-efficient, very, very productive factory, and the labor cost is about 500 dollars per month per person, so it is a little bit cheaper than Switzerland, and it is two hours from the US border.

I use that example to suggest that for globalizing businesses, this ability to set up and be able to create a culture in a factory is a key skill, but it is also far from impossible. You need to be able to, I think in economics you would call it open the black box, and work with culture. It is very important to understand what motivates people in different countries, and how you can integrate into their sometimes slightly difficult infrastructure, and so on. I think that must be one of the most interesting topics to understand: Why are we locating in Mexico, why is India difficult, why is China better, what is happening in Argentina, what is going on in Brazil? They are all very different stories.

Finally, the other issue is of course sales and marketing. How do you succeed in these markets? The Audi story is a good example of that. But the basic issue is whether you can keep a similar position, similar capability system, all over the world, or do you need to change your business dramatically? My experience would suggest that if you need to change your business model dramatically to succeed in another country, your likelihood of success is extremely small.

So those would be our perspectives and impacts on how we are trying to deal with these many issues.

5 Panel Discussion[1]

[1]Transcription of the discussion at the panel "Business on Globalization," moderated by Bent Jesper Christensen, Aarhus University, on November 10, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark.

David Teece: That was an extremely insightful talk. I would like to explore some of the tensions between what you were talking about early on around complexity and operations, and what you were talking about later, around strategy and focus. And you have come into the company when it was not doing well, and it is now doing better. So the question is, what counts in terms of long term survival? Is it possible to say what is more important? If you get the operations right, but the strategy is wrong, how far can you go? Or if you get the strategy right, and operations are not so strong, which one is going to kill you first, I suppose? And, secondly, you explained how you build first class operations. How do you build first class strategy and achieve the agility that you talked about as needing for the global market place?

Jørgen Vig Knudstorp: Yes, it is a classic. One is that operations are far more important, and that is because in many companies strategy is just something people like to say. It is great to do planning, then you do not have to do real work, and people abuse the word "strategy" a lot. Strategy can mean anything from an action plan to a grand vision. I think strategy is about how you gain advantage, what commitments you make that are very hard to change your mind about later, and that is the essence of strategy making. So, making a factory in Mexico is a strategic decision, obviously, since it is a little bit hard to move that one. But it is also clear that you cannot optimize your operational system, your operating model, unless you have a clear view on your strategy overall for the business, and you have to have a clear answer to the question, "Why do you exist?"

In our case, it is the building system. Incidentally, there is still no other physical material in the world that allows you to build anything out of something like LEGO, and then take it apart. Normally, you would have to glue it, and then you have to cut it, or do something like that. And we stick to that, our unique advantage, and then we build our capability system around that. We know best how to product develop on it, we know best how to merchandise it, we know best how to manufacture it in a consistent quality, and then we really invest in building those operating capabilities. What was wrong with LEGO when I arrived was that the company was trying to be a lot of different things, and so had moved from basically being a plastic manufacturing company that meant creativity, to wanting to do everything you could hardly argue was creative, like books, for instance. And then you move into businesses where you have no advantage, and you are in a very different territory.

That is why, basically, in economics we would call it economic rents, you must understand what carries economic rents. I am sorry, but if you do not know, I will tell you a very hard truth, which is of course that within your own business and subsystem and so on, you strive for efficient markets, and you try to set prices and transparencies, so people can optimize and minimize waste, but, unfortunately, when you go outside the boundaries of the firm, you want to look for market inefficiencies all the time. You want to be the only one to offer this product, you want to be faster than the other guys, you want to get the best space in the store, you want to be on the front page of the catalog, you are basically looking for all sorts of opportunities to drive rents, and that is what they in business schools call strategy, but it is really about rent-seeking behavior. So, it is very important to have an economist's view on that, because otherwise "strategy" becomes a buzzword.

Keith Maskus: I will just ask a simple question. I was really taken by your figures, especially that you have over 7000 brick types, and I assume that is the innovation, coming up with these new building materials for the customer. So is it really right, that is the innovation? And is that really important, that you continue so, or do you think that you are coming to some sort of equilibrium?

Jørgen Vig Knudstorp: The Group story is of course that basically we have not made any real innovation since 1958, when the basic brick was invented, and then every year we re-package them. But we do make new elements. Actually, making a new element in LEGO is like a religious event, because what you have just done is that you have added to the alphabet. And you should be careful doing that, because you increase complexity whenever you make new elements. A lot of the 7000 are variations, where about 1000 are new every year. It is called variations, not shapes, but of course sometimes we introduce new shapes, like the hair for a specific LEGO mini-figure, because it is part of Harry Potter, Alien Conquest, or whatever it is called, Ninjago, and so on.

There are several aspects to innovation. One is innovation in the business model and operations, so, how you charge for it, how you work with the retailers, to make sure you optimize the business. A trivial example would be understanding buying occasions. When are moms looking for a birthday gift? We actually videotape consumers in stores and share it with retailers. We study how consumers spend on their credit card. In Target, in the US, which is a little bit of an up-market, low-cost retailer, the average buyer of a LEGO set earns 70,000 dollars a year, and sometimes they go into a Target store to buy a LEGO set, and then what else they buy is because they are in Target. That is valuable innovative information. Another completely different aspect of innovation is working with our families, making the products they ask for, so we work a lot with the community, invite them in.

But the most important innovation I think we do is that we have ethnographers who live in families' homes and write diaries and basically live with the family and the children, to understand the needs and behaviors, and what is cool, and what is it that really drives kids. And then we have about 80 kids from Germany, the US, China, and the Nordic countries visiting our product development area every week, to test out product concepts. There are many aspects to it, also because there are many customers. You have to be acceptable to the grandparents, you have to be preferred by the parents, and you have to be cool to the kids, and sometimes you work on different angles to achieve that.

David Teece: To what extent is the analogy between what you build as a product the same as the way you build the company? A kid with a toy starts a creative period, then there is enactment, which is building it, then they tear it apart, which is transformation. The nice thing about it is that you get a click when you actually sock them together. In building the company, there is obviously a creative piece, there is enactment, and then there is transformation, and you have described all three. What is the limit to the analogy between the product that you build, and the company that you have?

Jørgen Vig Knudstorp: Yes, it may sound contrived, but there is an analogy. Here is how it happens. One of the things I was searching for when I took over was, what

is the deeper purpose of this company? Because I could tell that a lot of people had an affinity for it. And we have since the beginning of the 1980s been one of the largest sponsors of the Media Lab at MIT. I do not know if you are familiar with it. It was founded by Nicholas Negroponte, who wrote the book Negroponte (1995) titled "Being Digital" many years before many of us knew what digital was. We have a LEGO professor at MIT, and they run something called The Lifelong Kindergarten of learning, and what we believe happens when children play with LEGO is basically that we bridge the left and the right side of the brain. This is quite an important skill in school. I do not know if you have tried this test of asking a child to do this really fast: Opening one hand while closing the other, then switching and repeating? If they are unable to do it, it is because they have attention deficit, at least it is one of the crude tests of that. A lot of kids struggle with this, especially the boys. That is why it is hard to teach them anything as they grow up, until they reach fifth grade in school. Frankly, a lot of people think we should not even try to teach them anything before that, because they are not ready.

I think it is Kahneman, who studied stock traders with the psychologist Tversky and won the Nobel Prize in Economic Sciences (it is funny that an economist would make major innovations in cognitive psychology), and what they studied was people who were able to control their emotions, which is why they were winning on the stock exchange. If you do not have the connection between the left and the right side of your brain, you are unable to control your emotional reactions. That is what Kahneman (2011) calls system one and system two in behavioral control.

With LEGO, you learn basically how logic works, because LEGO is a binary system of zeroes and ones, and the child intuitively learns how logic is created, how things fit together, how the system of plants, the Linnaean taxonomy, or classification, any sort of scientific logic, is intuitively understood. Of course, they have no idea, they are just having fun. That is why many architects, software engineers, and other people who are really into logic, mathematics, and so on, are also big LEGO fans. On the other hand, they also learn that with LEGO you can build anything you can imagine, and even stuff you did not know. You imagine things as you design, and so a lot of designers are also LEGO fans.

But we think it is very important to give children both systems: Logic and creativity. For any one of you studying network economics, LEGO is a fantastic example of network economics, in its whole business model, but, on the other hand, they learn creativity and design. And I personally think that one of the most important things we can do in the education system is to teach our children creative problem solving, because that is what it is.

So what we do at the company is to train all our employees in a concept we call systematic creativity. Most people say those are two opposing ideas, and that is the point, it is to be able to reason systematically, think creatively, and work collaboratively. And so our leadership training program at LEGO is called "Opposable Thinking." What we want to teach people is that there is never one hypothesis that explains what happens in real life, that is unfortunately the case, there are multiple competing hypotheses all the time, and your value as a senior

leader in our company comes not from having one hypothesis, but from containing two, or maybe five competing hypotheses.

I think Scott Fitzgerald is quoted for saying "It is a sign of first rate intelligence to be able to hold two opposing ideas at the same time," which, I think, is a bit of a low bar for intelligence. But it is interesting, if you take, for example, that I worked for McKinsey for some years as a management consultant, and thought that I was going to be their strategy genius because I had a Ph.D. But then I quickly found out that what you advised at very high rates every day to customers was that if they spent more marketing money, they would also be able to gain more market share. That is a fantastic causality. I did not need statistics from Aarhus University to calculate that. There is a huge value in management to be able to understand that, and it is very hard to train people to do that. But I actually think it is grounded in our product, and so it is what we try to make our culture, as well. So there is a link between product and culture. That was a very long answer, but it is very important to me.

References

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
Negroponte, N. (1995). *Being digital*. New York: Knopf.

Globalization: Some Reflections and the Road Ahead—A Panel

Avinash Dixit, Ronald W. Jones, Dale T. Mortensen, with Carsten Kowalczyk (Moderator)

Abstract This chapter is based on a transcription of the panel "Globalization: Some Reflections and the Road Ahead," moderated by Carsten Kowalczyk, Tufts University, on November 11, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark. Remarks by Avinash Dixit on the collapse of trade after the financial crisis, bad governance as a trade barrier, and the interaction between internet use and globalization are in Sect. 1, remarks by Ronald W. Jones on globalization and the "sclerotic economy" are in Sect. 2, and remarks by Dale T. Mortensen on unemployment, reallocation, and frictions are in Sect. 3. A transcription of the panel discussion is in Sect. 4.

1 Globalization: Trade, Crises, Governance, and Technology. Remarks by Avinash Dixit[1]

I will focus on three topics that I judge to be timely and important, and could benefit from more research. Near the end of a conference with so many great papers on so many topics, it may seem too negative an attitude to dwell on things we don't know

Sadly, Dale T. Mortensen passed away on January 9, 2014.
http://econ.au.dk/research/other-research-entities/cap-cycles-adjustment-and-policy/past-events/conference-on-globalization-strategies-and-effects/.

[1]Remarks by Avinash Dixit for the panel "Globalization: Some Reflections and the Road Ahead," moderated by Carsten Kowalczyk, Tufts University, on November 11, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark.

A. Dixit
Princeton University, Princeton, NJ, USA

R.W. Jones
University of Rochester, Rochester, NY, USA

D.T. Mortensen
Northwestern University, Evanston, IL, USA

C. Kowalczyk (Moderator)
Tufts University, Medford, MA, USA
e-mail: carsten.kowalczyk@tufts.edu

B.J. Christensen, C. Kowalczyk (eds.), *Globalization*,
DOI 10.1007/978-3-662-49502-5_24

instead of celebrating all that we do know, but I take a positive perspective and hope it will spur thinking and action, and perhaps help generate some papers for the next such conference!

A precautionary statement at the outset: For some years now I have been away from the mainstream of research in international trade, and may not know of some related work. If so, I apologize to the authors. But looking at it more positively, such work would be supporting evidence for the importance of the issue.

1.1 Collapse of Trade after the Financial Crisis: Competing Explanations

The financial crisis and its aftermath have shown that globalization is not a one-way process of expansion; it can suffer big setbacks. It is important to know the causes and mechanisms that can lead to such reversals. The magnitude of the collapse observed in 2008–9 was significantly larger than expected based on previous estimates of the relationship between world trade flows and GDP. The Houthakker and Magee (1969) finding that the elasticity of trade with respect to GDP was around 2–3 seemed, if anything, an overestimate in business cycles of the previous 10–20 years, but even at its most extreme it can account for only 70 % or so of the collapse in trade in 2008–2009.

The Great Recession differed from the cyclical downturns of the previous half-century, in that its proximate cause was the financial crisis. Therefore it is natural to expect that financial factors, for example drying up of trade credit, should have a prominent role in explaining the abnormally large decline in trade. But research on this topic has yielded quite divergent results.

International trade contracts have longer time lags than domestic ones between writing and fulfillment, and involve additional risk and information asymmetry. Therefore they depend more heavily on credit lines and credit default insurance, which become more costly or less available in a financial crisis. Thus, the mechanism linking finance and trade is quite intuitive, and has been formalized in general equilibrium models (see Ahn 2011; Ahn et al. 2011). Their empirical work finds that exports in industries more heavily dependent on external finance suffered more. However, Levchenko et al. (2010, 2011) find that U.S. imports in sectors with vertical linkages (intermediate inputs) and cyclically sensitive sectors suffered more, but find no significant excess decline in sectors more dependent on trade credit. It seems important to have a clearer resolution of this issue.

1.2 Bad Governance as a Trade Barrier

Theoretical modeling, as well as empirical analysis, suggest that a country with poor protection of property rights and poor enforcement of contracts will have less trade and will attract less foreign direct investment. It would be useful to estimate the height of this barrier, and in particular, do so in a way that facilitates comparison with the heights of more conventional barriers such as tariffs and quotas. In other words, is there an "equivalent tariff" to governance quality (or rather, lack of it)? And when poor governance and tariffs or other barriers coexist, how do they interact—subadditively or superadditively?

Anderson and Marcouiller (2002) and Leeson (2008) examine the effects of contract enforcement on trade. Anderson and Marcouiller find that "inadequate institutions constrain trade as much as tariffs do." For example, if Latin American countries were to have the same quality of governance as the European Union, import volumes of Latin America would rise by 30 %, whereas if their tariffs were lowered to the levels in the United States, their imports would increase by 35 %. Leeson finds that the effect is "significant but modest compared to intuition." He estimates the effect of the New York Convention (under which signatory countries undertake to enforce the awards of international arbitration tribunals without rehearing the whole case), and finds that state enforcement would increase trade by "about 15 to 38 %." The estimates in the two papers are actually not far apart, but the authors' interpretations are far apart: dramatic versus modest. More work of this kind, especially exploring the interaction issue, would be welcome.

Poor contract enforcement has mixed effects on FDI: the overall volume of transactions may be lowered, but the ones that do occur may take the form of FDI with wholly owned subsidiaries, that is, internal governance, instead of modes such as exporting, licensing, or working with local partners that require external contracts. Globerman and Shapiro (2002, 2003), Henisz (2000), Javorcik and Wei (2009) and others report these results. Again, this is an area with promise for continued future work.

A country with weak governance may be an especially good source of outward FDI, for two reasons. First, firms from this country, which have accumulated profit there, may not want to invest more in the insecure environment, but instead invest in other countries with better governance. Second, and more interestingly, the experience these firms have gained in coping with bad governance on their home turf may become an advantage when they invest in other countries with similarly poor governance. This may be one factor that explains the surge in south-to-south FDI. I explore this in a theoretical model (Dixit 2011), and cite in support some empirical evidence from the literatures; for example the finding of Darby et al. (2010) that the negative impact on FDI of bad governance in the destination country is smaller in magnitude if the source country also has bad governance. This seems to be an interesting hypothesis for further research.

1.3 Interaction Between Internet Use and Globalization

Some very interesting research has shown how information technology has revolutionized search for the best price and matching; for example access to smartphones enables fishermen in Kerala to find the best prices for their catch (Jensen 2007). Buyers can similarly find low prices anywhere in the world. This has had an impact on international trade; for example writers of textbooks like me lose from globalization because students in the U.S. can now buy their books far more cheaply from web sites in developing countries and get them delivered to the U.S quite cheaply, which reduces my royalties. But I am suggesting research with broader scope, to look for other avenues by which the spread of internet and smartphone use affects international trade. For example consumers in third world countries become far more quickly informed about new products and global brands; this affects their preferences, and therefore patterns of trade. As some discussions in this conference brought out, information technology influences matching and outcomes in labor markets, and that the effect can differ across occupations and industries. Cross-country differences in technology and policy with regard to internet access can be exploited in empirical work on such questions. All of this offers rich opportunities for future research.

2 International Trade and the Sclerotic Economy. Remarks by Ronald W. Jones[2]

The theory of international trade is founded upon a basic economic principle that was enunciated nearly 200 years ago by David Ricardo: The doctrine of *comparative advantage*, whereby a country can gain by leaving a state of autarky and, instead, engage in international trade in commodities with other countries in world markets. In general, the idea that individuals might mutually benefit from trading with each other was not a surprise, especially if each individual had a decided (or *absolute*) advantage in producing the commodity he or she sells. In a world in which there is little or no mobility of labor or capital between countries the word *comparative* comes into its own. Countries mutually gain by exporting commodities that are *relatively* (or comparatively) inexpensive at home, even if the importing country could produce such commodities with less labor and capital than the exporting country—higher wage rates or returns to capital than in the exporting country would be the reward for such absolute advantage.

Such reasoning has, since Ricardian times, served as the basis for arguing that open international trade, relatively free of artificial barriers to trade, would lead

[2]Remarks by Ronald W. Jones for the panel "Globalization: Some Reflections and the Road Ahead," moderated by Carsten Kowalczyk, Tufts University, on November 11, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark.

to a great *concentration* in production of commodities that a country will export or at least produce to augment its imports from other countries. Such relative concentration in production leads to an important corollary if a country enjoys growth and/or is subject to price changes in world markets: The commodities in which a country possesses a comparative advantage may *change* as a consequence of growth or a disturbance of world prices, and this leads at the micro level to a *churning* of productive activity. A country may enjoy, say, a relatively smooth rate of growth of around 4 % at an aggregate macro level, but at the micro level such smooth growth may mask growth in some new sectors of, say, 25 % annually accompanied by a 20 % fall in production in output in sectors in which the country previously had a comparative advantage. Smooth growth may well be associated with relatively *balanced* expansion in the country's *consumption* amounts, but decidedly not in its *production* pattern. International trade is by definition unbalanced, and a country's pattern of production may well change with growth or with alterations in world prices.

In recent decades, the nature of international trade has been undergoing a significant alteration: The fastest growing type of trade is trade in parts and components. This is often referred to as *outsourcing* at an international level and is a natural outcome of the greater ease with which production processes can be fragmented. In brief, as outlined in Jones and Kierzkowski (1990), most production processes consist of a series of rather separate *production blocks* which are linked together by various *service link* activities consisting of expenditures on communication, insurance, transportation, and other activities that serve to co-ordinate the overall production process. In recent decades the costs of such service link activities have been significantly lowered (especially those involving communication, but also such costs as transportation and gathering of information). As argued by Jones and Kierzkowski, service link activities are *not* typically of the "iceberg" type suggested by Samuelson (1954) (and earlier by von Thünen 1826), with constant returns to the scale of production. Instead, service link activities result in expenditures that are either independent of the *scale* of production or at least increase less than proportionally with output. (For example think of the costs of communication). As scale expands, firms may now find it expedient to have some production blocks produced in other countries, where there may be a better fit between factor prices and relative intensities of factor inputs. And the importance of the doctrine of comparative advantage suggests that it may pay to outsource, say, a labor-intensive production block to a country even if foreign labor is less productive in the activity abroad if this is more than compensated for by an even greater reduction in wage rates.

The extent of international outsourcing, whether within a firm or undertaken at arm's length with a different firm, has greatly increased, both because overall output levels have increased (as income levels generally rise) and because of significant reductions in service link costs. This leads to a greater division of productive activities and greater gains from trade (consistent with the Ricardian argument as more activities, i.e. production blocks, have their location determined by considerations of comparative advantage). And for some developing countries

new opportunities for trade may open up as these countries may not have a comparative advantage in an entire production process, but may well be able to compete as trade opens up in some individual production blocks.

Such a proliferation of possible global assignments reflecting comparative advantage may also pose a threat to overall gains from international trade. Most governments are concerned, in part, with encouraging certain productive activities *within* the country. A question often raised is whether a country's government (or public sector) is better at predicting or forecasting local productive winners than is the private sector. (My own guess is that it is typically better done by the private sector, with more "skin" in the game). There is, perhaps, a more important attribute exhibited by decisions made by the public sector: It is often difficult for the public sector or government to *cease* protection provided to a group in the national private sector after it has originally been granted. Government is not very good at terminating protection granted to local activities, especially if there is a threat from *foreign* competitors. The issue of *us* versus *them* usually interferes with such decisions. The aforementioned feature of international trade based on concentration of productive activity into those areas in which a country possesses a comparative advantage is a setting in which there is much churning from year to year as new winners overtake previous winners. The very act of trading in world markets in which only a few local firms or industries can successfully compete implies that over time there will be many local activities that can no longer maintain a presence in global markets *unless* they receive some support from government, say in the form of subsidies or tariffs imposed on foreign suppliers. Often the phrase used is that the role of government is to create "a *level playing field*." Put differently, the role of government is often one of interfering with the forces of comparative advantage in order to favor "us", i.e. home industries, over "them", i.e. foreign challengers.

The combination of a trend towards relatively greater government activity, some of which is aimed at supporting through various means those domestic productive activities that in earlier times were primary participants on world markets, with the increased role of international trade, supported in large part by an increased role of outsourcing of many *production blocks* to overseas locations, serve to question the ability of international trade to provide the level of gains to real incomes suggested by the doctrine of comparative advantage. It may not be easy for a country to cut back on activities that in earlier times did indeed succeed in exhibiting strong comparative advantage so that more recently emerging activities with stronger comparative advantage can obtain the resources required to compete successfully in world markets. However, it is in the very nature of trade and growth for open economies that at the micro level some industries will be falling back with lower sales as new industries emerge at rates of growth far exceeding the national average. International trade in a setting of growth does not call for balanced growth of production, and attempts to prop up previous winners threaten to lower growth rates and may lead to sclerosis of the economy.

3 Globalization: Unemployment, Reallocation, and Frictions. Remarks by Dale T. Mortensen[3]

3.1 Financial Crisis Effects on Unemployment

The Great Recession was clearly related to the global financial crisis. As it is the first time that this has been the case in the U.S. since the 1930s, it may not be surprising that labor and macro economists underestimated the size of its impact on employment. There is now a frantic research effort among these same economists to catch up. I see this effort going on across the board. Proponents of both DSGE and neo-Keynesian models are running races in their attempts to integrate a meaningful financial sector into their models. The same is going on among search theorists. The common approach is to adopt and adapt the existing partial equilibrium models of Kiyotaki and Moore (1997) and Bernanke and Gertler (1989), which emphasize moral hazard in the credit relationship between banks and employers, for the purpose. I, too, am working with a current student on this topic.

There are other related efforts going on regarding how imperfect competition in the product markets and the dependence of consumption on wealth shocks particular to the value of real estate may also propagate these effects. Although I am not aware of current research on the issue, it is obvious that the impact on and from international trade, that Professor Dixit raised, may well be another important propagation channel.

3.2 Reallocation and Employment: Bad Governance and/or Friction

The realization that large firm productivity differences exist, coming out of recent studies of micro data on firms, has had a huge impact on trade theory and is now affecting both labor and industrial economics. This observation has now been integrated with the long-known huge cross-country total factor productivities (TFP) differences. The already classic Quarterly Journal of Economics paper by Hsieh and Klenow (2009), who look across countries at productivity dispersion across firms, has played a key role in the discussion. In the paper, they document that the 90/10 TFP percentile ratio in the US is 3.3, while the same ratio in China and India is about 5. They do a real back of the envelope naïve calculation. If all of these differences

[3]Remarks by Dale T. Mortensen for the panel "Globalization: Some Reflections and the Road Ahead," moderated by Carsten Kowalczyk, Tufts University, on November 11, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark. Sections 3.1 through 3.3 consist of prepared remarks submitted by Dale T. Mortensen. Section 3.4 contains additional remarks from a transcription of the actual panel discussion. Sadly, Dale T. Mortensen passed away on January 9, 2014.

are due to some intrinsic differences in productivity across firms, then they show that you can increase GDP in both China and India by 100 % if the productivity differences were fully equalized through resource reallocation, and by 50 % if they were reduced to the level of the US. Obviously, doing so would have a huge impact on world well-being.

There is, however, an obvious question: What is the cause of the so-called allocative "wedges" that now prevent equalization? Why all of this dispersion? Hsieh and Klenow, as well as Prescott and his co-authors (e.g., Parente and Prescott 2000), and other researchers, blame bad governance. Specifically, there is unequal taxation across industries, differential treatment of firms, protection of interests, as well as bribes and outright corruption. That certainly explains some of it, particularly in developing countries. But what if we take a country like Denmark? The argument does not address the fact that the 90/10 ratio is 3.3 in the US, or the 2.3 number for Denmark that my colleagues and I (Bagger et al. 2011) have found in a country that is reputed to be among the best governed in the world. There must be some other reasons going on.

This has been a focus of the research group here at Aarhus University. There are frictions that prevent equalization. And trying to understand what those frictions are is important. Documenting the dispersion for Danish firms and attempting to find evidence for the reasons for the differences have been the major research goals for my Danish colleagues and me for the past 10 years. Bent Jesper Christensen, Jesper Bagger and I (Bagger et al. 2011) are estimating a model that attempts to answer the following question: Are these differences in firm productivity intrinsic, or can they be explained by the fact that more productive firms employ more productive workers? We find that in Danish manufacturing, differences in TFP are an important part of the explanation. In our paper in Econometrica (2008), Rasmus Lentz and I argued that the dynamics of firm entry and exit, together with the fact that new firms are born small and face growth friction (the Penrose effect), are possible reasons. We estimate a version of the Klette and Kortum (2004) model of firm entry, exit, growth, and so forth using Danish data. We are able to explain dispersion in that model by the fact that new firms are born young and there are costs of growth—frictions in the growth process. And new firms eventually are going to have products that are displaced by other new firms and by other competition. The continual dynamics of innovation, given that there are costs of growth, are going to maintain this dispersion. There's going to be a natural distribution of differences in productivity across firms. That is, the best firms are not going to be able to take over the market like they're supposed simply because they can't live long enough. The old firms can survive in that environment.

Now, what that says is that policies you really need to worry about are impediments to firm growth. In fact, we calculate from our estimation that you can attribute up to half the growth in aggregate output in Denmark to these impediments to full reallocation through the process of firm selection. It comes out of the analysis that you certainly don't want to do anything to prevent that process.

3.3 Internet Effects on Job-Worker Matching at Home and Around the World

The internet has become more important than any other formal channel as a medium through which workers and firms find one another. This is not only true in the fully developed world of North America and Europe. A few years ago, one of my students wrote an innovative version of the usual discrimination experiment in which equally qualified but racially different workers apply for the same job. What was so original about her experiment was that she simply submitted via the internet two identical applications from bogus workers with names that reflected different caste membership in India.

I have been asked for years now about the impact of the use of the internet on matching efficiency. In the early 90s, I said in response that it was too early to answer the question, because the use was limited to a few professions, mainly computer programmers, at the time. The first serious research on the issue that I know was by Peter Kuhn and one of his students (Kuhn and Skuterud 2000, 2004). Based on estimates drawn from the 1998/2000 Current Population Survey (CPS) computer use data, they concluded in their 2004 American Economic Review paper that internet search was ineffective. I am happy to report that a new working paper is now available on Peter's web site (Kuhn and Mansour 2014), in which they applied the same methodology on data from the same source 10 years later. For the later period, they find that internet matching reduces unemployment duration by an average of 25 %. As this number translates into roughly that same percentage decrease in unemployment, the magnitude is huge.

This is only the beginning. Already there are sites that match employers that have specific jobs to do with qualified professionals around the world. This is a new form of outsourcing, with volume doubling every few months or so. Be aware, this innovation may represent one of the biggest revolutions in the global labor market, because it is a way to realize some of the gains of international labor reallocation that were only possible through emigration in the past.

3.4 Response to Dixit and Jones

Today, at this panel, in some sense I'm the local, having an association with Aarhus University. And if not the local, I'm certainly the outsider in that my work has not had a lot to do with globalization. But I requested I go last, in part so I could respond to these experts on globalization, Avinash Dixit and Ronald Jones, in a number of different ways.

One of the things that occurred to me is that globalization may be threatening to your subfield. When the wall came down, and the Soviet Union disappeared, all of my friends in comparative systems lost their comparative advantage. And that could happen to you, too, with globalization. But let me say, even if that does happen, we

labor economists have long known about comparative advantage. And I don't think those principles will die with the field.

A couple of topics of future research: Avinash brought up the crisis and a surprise about the trade effects of the crisis. Ditto for labor economists: The impact of the global crisis in any one country, let alone across countries, and the depth of the effect on unemployment, were certainly not expected. What seem now to be the long-term consequences, particularly in the US, which was hardest-hit by the financial crisis, are still to be explained.

For example, I've been running around the world talking about how one might look at the employment effects of the Great Recession with the Diamond-Mortensen-Pissarides model, not, of course, revealing that when I shift my curves around to explain the drop in demand that accounts for the increase in unemployment, we have no quantitative basis for the extent to which I'm shifting those curves. We really don't have a good sense of what the channels and propagation mechanisms are that affected employment to such a large extent. And I think this a real challenge for macroeconomists. And it really has to do with extending our thoughts beyond just a single market—just focusing on the labor market like my own theories do—by linking up the additional propagation that operates through the credit market, through imperfections in the product market, and through trade itself. There's a big challenge for young people.

My second point is going to jump off on Ron's remark that macroeconomists don't see the richness of microeconomics. That may be true in the past. It's no longer true of current research in the present. As an example, many of us in macroeconomics have become Schumpeterians. Sam Kortum is a well-known trade theorist, but he has also had a big influence on my views of how the labor market works in terms of firms' births and deaths, and what that involves. And my own recent research has been on that point.

Let me try to explain this. We teach in microeconomics that output is maximized by equalizing the value of marginal input products across firms. So a lot of us were surprised when we actually started looking at firms—that arbitrage doesn't take place to the extent that one would expect. There are huge differences in productivity across firms. These huge differences have now revolutionized one aspect of modern empirical trade theory, but it's also having a big impact on how we think about macroeconomics more generally. This is discussed above, in relation to the paper of Hsieh and Klenow (Sect. 3.2). Productivity differences across firms may be explained, in part, by bad governance in developing countries, but frictions appear to be part of the explanation, as differences exist in developed countries, too.

There is another issue, too. Of course, this research supposes that there are intrinsic differences across firms. It's certainly true in our measurements that those differences appear, but those may be a mirage, really. Our ability to compare productivity across firms, of course, is highly complicated by the fact that firms differ tremendously in terms of who's employed, for example. You want to compare labor productivity of a firm composed of lawyers—how much value added do they provide per person—with a construction firm. There's a natural apples and oranges problem there. But that's another topic that needs far more research, and it's actually

one that we're working on. How does sorting—what kind of role does it play in these differences? And if it plays a major role, is there really that much room for reallocation to improve output?

Another important labor topic is the effect of trade on employment. We could go on for a long time about that, but it's an old topic, really, and we're all aware of it. More and better research needs to go on. But I'd like to raise in addition information technology and how that relates to labor market issues at the global level. I've been asked, since it's natural for people to ask me, what has the impact of the Internet been on the matching function? Does increased communication really make it easier for workers and employers to get together and reduce the frictions there? That's an interesting question. First I sort of backed away from it in the 1990s by saying that there are very few professions that actually use the Internet for matching. But that's changed. We're now in the twenty-first century, and the Internet, in the United States, has virtually replaced newspaper ads as a source of announcing vacancies and providing information about job availability, so it's time to answer that question.

Again, the first research I was really aware of was that by Peter Kuhn (Sect. 3.3), who used the Current Population Survey special survey on computer use right at the turn of the century. And in that data he could identify people who did use the computer in their job search with those who did not. The conclusion at that time was that the Internet was ineffective—it had no impact on unemployment directions. Well, of course I had a story for that, too. If you reduce that kind of friction, then the natural response of everybody is to raise the bar, that is, to insist that the match be better, because you can have a trade-off between these increases in efficiency. One of the ways you exploit that is to create better matches.

But I've been saved recently because Peter has redone his work with data from 10 years later. And now he's found that, in fact, everyone has learned how to use the Internet, and there's a 25 % reduction in unemployment durations. That's huge. Because the flow into unemployment is more or less insensitive to this, that's roughly equivalent to a 25 % reduction in the unemployment rate, or at least the search unemployment rate.

That's not the only place that the Internet is making inroads. The other is outsourcing. There are new websites that allow employers in Western countries to farm out jobs around the world—to the Philippines, or to India, to appropriately educated people. And that matching business is doubling in volume every couple of months. It's not very big yet, but it's growing very rapidly. So again, we have another effect from globalization.

It occurred to me, thinking about this, that maybe this has a very positive effect, in the sense that it allows us to exploit these differentials in input costs and comparative advantage without migration. And maybe that will solve some of the problems that are implicit in migration. Because we know that has also been a recent topic. We just had a recent article in Journal of Economic Perspectives (Clemens 2011) that calculates that if we opened our borders completely and did international equalization of the value of the marginal product, then that would double the world's

output. So maybe the Internet is going to provide some very important ways of exploiting that.

I've been on the optimistic side, but Ron's already covered the negative part, so I won't have to go into that. Thank you.

4 Panel Discussion[4]

Carsten Kowalczyk: Any questions?

Jeffrey Bergstrand, *Notre Dame:* We embrace comparative advantage as economists. Now put aside that a lot of growth in international trade over the past decades is intra-industry, and we then know that the argument is not as much reallocation between industries. But we also have this booming trade going on between developed economies; for example, China and India are growing. Trading according to comparative advantage requires reallocation of labor. But the other side of comparative advantage is that you can use lump-sum transfers from winners to losers and make everybody better off. Yet we do not really pursue that avenue. I was wondering if all of you can comment on how we, as a profession, are pushing this comparative advantage to policymakers and businesses, but whether we have done enough on discussing the sharing of those gains? For example, some years ago Carl Davidson and Steven Matusz, and others, estimated the gains of comparative advantage to exceed the costs of reallocation by orders of magnitude, like a hundred to one. Can you gentlemen speak to that?

[4]Transcription of the discussion at the panel "Globalization: Some Reflections and the Road Ahead," moderated by Carsten Kowalczyk, Tufts University, on November 11, 2011, at the Aarhus University conference Globalization: Strategies and Effects, Hotel Koldingfjord, Denmark.

Jones: I think there is a very general aspect to attitudes about comparative advantage and what should be done with changes in foreign competition, et cetera. We tend to concentrate on the import side or on our outsourcing. But as Walter Wriston wrote shortly before he died, "has anyone ever heard of insourcing?" And he mentioned that for the U.S. it's around 50 % greater than outsourcing. Often arguments about trade policy, or commercial policy, get pointed at the difficulties of struggling industries. Or to use a phrase that I think does almost creative destruction to the concept of comparative advantage: "we should have a level playing field." That phrase is often used – that there is something unfair about one group of people in some area, or especially in some country, having an advantage over us in something. Usually the critics forget about comparative advantage because they are thinking of absolute advantage or really about the outcome of absolute advantage on markets, where productivity might be much to our advantage but wages, et cetera, are perhaps an even greater source of difference. There is a natural reason for a lot of this: the "us-versus-them" aspect of dealing with foreigners. Government can often do things against firms, making distinctions between one and another. If it is us versus them then they can do so internally. That is also possible if it is one part of the country versus another part, although they have to be a bit more careful about that.

In the United States, people like the TV journalist Lou Dobbs concentrate on how trade hurts U.S. import-competing activities or even U.S. exporters when they are going to get hit. On the other hand, relatively little is said for the other side, about the exporters.

One reason why it is easier to talk about import competition is that if you are talking about, for example, members of the steel industry getting put out of work—we can call them John and Jane and Richard—we know who they are, we can make it very personal. When you come back and say, yes, but if we levy tariffs, we will export less, well what will be involved, who are not getting jobs in the export sector? We do not have their names yet, it is too vague. So people who push for protection have that kind of advantage.

In addition, I think that if there is any way of measuring what are the responsibilities and role of government, it seems to have become increasingly important to take care of those who are going through a hard time—whether it is the unemployed or those affected adversely by international trade. I think the net effect will be the introduction of more sclerosis into the economies, in the sense that things will get more sticky, and it will get harder to let comparative advantage and competition play themselves out. The only group that I know of that likes competition are athletes who wouldn't make money without out it—and the ones who are very good.

Dixit: Let me pick up the issue of compensation. Yes, sure, there are these losers, but, as you, Jeff, pointed out we can organize lump-sum transfers. I think we can do even better than that: lump-sum transfers on a personalized basis are very difficult to organize because they need information which the parties in question have no interest in revealing. But you can even do Pareto improving compensation using Diamond-Mirrlees commodity taxation. In principle that ought to be possible. But I think the issues are much more political on the one hand and philosophical on the

other. The political issues are problems of commitment: if this is being done through a democratic process, the politicians cannot be sure that the people who receive compensation will vote for them. And for the people receiving compensation—since you can't give it to them in kind of one big sum—the politicians will want to give it to them in little bits to keep that incentive to vote intact. The commitment issue will then switch to the other side: namely, the people will not be sure that the politicians will keep their promises. Philosophically, I think establishing the kind of general principle that anybody who is hurt in certain kinds of circumstances is entitled to be compensated is opening up to various dangers. People will then start to be reckless and claim that the compensation principle applies because whatever they are trying recklessly didn't work out. It's only bankers who get away with it. The rest of us don't.

Mortensen: It is an excellent question, because I always felt that as a profession we emphasize the pie is bigger, but then deemphasize how you are going to compensate the losers. We know there are problems of doing it, and that is what you are saying, Avinash. But just because there are problems does not mean we should not be putting in more effort, maybe as a profession, inventing better ways to do it, and you also addressed that, Avinash. We have gone through a period of globalization—and I am not only talking about trade between countries—where there has been a move to making markets more flexible. For example, there has been a big push at the OECD to make labor markets more flexible. But in doing so, you also make it easier for the winners to gain and the losers to lose. In the past, things like employment protection were put there for a reason. It may not have been a very good one, but it was a reason. You take those things off, you improve efficiency but you do not necessarily improve distribution. Now we are facing the 1 % movement, and I think to a large extent we are facing that movement because, of course, at least in the U.S. and in the UK, only the top 10–15 % of the income distribution have benefited by these forms of openness. I think we are going to face a political reaction to our success in improving efficiency. I'd like to get another response on that.

Kowalczyk: There is a political economy dimension to this. We have a number of small countries that have actually voted large trade liberalizations through the years, imposing large reallocation of factors in their economies. To some extent one could argue—and there are folks who have looked at this—that the notion that potential losers may not lose that much because there is a built-in income shifting, through income taxes or whatever, may actually make the median voter less concerned about the possible consequences from these large shocks. So it's a mixed bag: compensation may introduce moral hazard but it may also allow otherwise risk-taking but growth-enhancing voting.

Mortensen: Yea, but you're talking about a couple Nordic countries, right? Not the rest of the world.

Kowalczyk: So there is a potential concern that future liberalizations, or other changes, may not be voted in and the system freeze up. There is not much work on how to introduce such compensation mechanisms, nor on how to phase them out.

Dixit: Actually, Dale's remark about Nordic countries is almost exactly right. I remember a book by Peter Katzenstein, a political scientist at Cornell, called *Small States in World Markets* (1985), in which he looked at case studies of small and culturally homogenous countries, which he said are able to strike and maintain a social bargain that says that they will keep their markets open but have a social safety net that takes care of losers. One can see why both the small part and the homogenous part are necessary for the political process to work in that way. His phrase for this was that these countries have "low-voltage politics," and for countries, particularly like America these days, that have very high-voltage politics, that may not be feasible.

Mortensen: But, I might add, it may become necessary.

Ian Coxhead, Wisconsin: I was struck, Ron Jones, by your response to Bertil Ohlin—to which you referred in your opening remarks—to do nothing on the trade front. As a development economist I want to just ask you to revisit that—the kind of advice you might give someone in a developing economy setting. I think what we are learning in this era of globalization is the remarkable extent to which standard stories about economic development are really contingent on international factor immobility and trade costs, all of which prevented complete specialization and created a very broad domain for structural change. So you could move out of agriculture into other industries because you had capital that was stuck in the economy and because you had high trade costs, which gave you a degree of natural protection. All of that is going away, and I think one of the big negative consequences of that is for resource-rich countries which see themselves now facing a world in which they will be specializing in natural resources and have a hard time escaping the natural consequences of Dutch Disease. What kind of advice should trade policy specialists give to the governments of those countries? Do nothing?

Jones: No, I'd say not. You know, be better informed about things. One thing I wish they would say, or the media might add, is the fact that there is competition and churning activity going on even without considering the international trade aspect of it. Every month the turnover of people in the United States taking new jobs is very high compared with the number of jobs lost to outsourcing. I think that what interferes with the media and others doing this is, in part, the "us-versus-them" aspect of things—that we are more ready to accept that people's tastes may change internally, and that some businesses seem better than others. I think on a longer haul, including the development dimension, one of the things that helps a lot is mortality: That you get an intergenerational change and thereby a movement out of one activity into another. Consider for example the U.S.: if you look at 1900, what fraction of the labor force was in agriculture as compared to fifty years later or a hundred years later? Well, was that because a lot of farmers went out and learned how to make automobiles and things? The answer is no. Rather, the kids realized that following in dad's or mom's footsteps might not be a smart thing to do. And that tends to shift the focus in more general terms to education. I personally note this when I see my five and six year-old grandchildren being able to use their thumbs and play with their iPhones, and I don't know what they are doing, much less why! But they have learned these things. I think one aspect of the development process

generally is that the culture and the set of laws and provisions are dictated by the relatively older part of the population that has wealth, and that they are traditionally threatened, but increasingly so, by two potential competitors: one is foreigners, and the other is the younger generation. And the younger generation has more human capital than physical capital, and I think that gap is increasing. They may have a rough time in recession periods but generally the growth of human capital is more important to them. We are going to see a lot of adjustment taking place that way. It is much easier to blame it on the foreigners, and you get much more sympathy for it.

Dixit: May I come in with two quick remarks? One is actually following up on what Ron has been saying. That's a process that should be very familiar to all of us. It's the kind of Kuhnian process of scientific revolution. Scientific revolutions take place, basically, not because senior professors change their mind, but because graduate students pursue different ways of thinking. On the question specifically on what advice I would give to resource-rich country governments: be like Norway, or be like Chile. Do not renounce the great advantage you have; just use it wisely.

Mortensen: Well, I just wanted to comment a little bit of an ironic note about all these old guys up here, you know, spouting the benefits of the younger generation taking over.

Kowalczyk: Thank you so much.

References

Ahn, J. (2011). *A theory of domestic and international trade finance*. IMF Working Paper 11/262.

Ahn, J., Amiti, M., & Weinstein, D. E. (2011). Trade finance and the great trade collapse. *American Economic Review: Papers & Proceedings, 101*, 298–302.

Anderson, J. E., & Marcouiller, D. (2002). Insecurity and the pattern of trade: an empirical investigation. *Review of Economics and Statistics, 84*, 342–52.

Bagger, J., Christensen, B. J., & Mortensen, D. T. (2011). *Productivity and wage dispersion: Heterogeneity or misallocation?* Working Paper.

Bernanke, B., & Gertler, M. (1989). Agency costs, net worth and business fluctuations. *American Economic Review, 79*, 14–31.

Clemens, M. A. (2011). Economics and emigration: Trillion-dollar bills on the sidewalk? *Journal of Economic Perspectives, 25*, 83–106.

Darby, J., Desbordes, R., & Wooton, I. (2010). *Does public governance always matter? How experience of poor institutional quality influences FDI to the south*. CESifo Working Paper No. 3290.

Dixit, A. (2011). International trade, foreign direct investment, and security. *Annual Reviews of Economics, 3*, 191–213.

Globerman, S., & Shapiro, D. (2002). Global foreign direct investment flows: The role of governance infrastructure. *World Development, 30*, 1899–1919.

Globerman, S., & Shapiro, D. (2003). Governance infrastructure and US foreign direct investment. *Journal of International Business Studies, 34*, 19–39.

Henisz, W. (2000). The institutional environment for multinational investment. *Journal of Law, Economics and Organization, 16*, 334–364.

Houthakker, H., & Magee, S. (1969). Income and price elasticities in World Trade. *Review of Economics and Statistics, 51*, 111–125.

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics, 124*, 1403–1448.

Javorcik, B., & Wei, S.-J. (2009). Corruption and cross-border investment in emerging markets: Firm-level evidence. *Journal of International Money and Finance, 28*, 605–624.

Jensen, R. (2007). The digital provide: information technology, market performance, and welfare in the South Indian Fisheries Sector. *Quarterly Journal of Economics, 122*, 879–924.

Jones, R. W., & Kierzkowski, H. (1990). The role of services in production and international trade: A theoretical framework. In R. W. Jones & A. O. Krueger (Eds.), *The political economy of international trade*. Oxford: Basil-Blackwell. Chapter 3.

Katzenstein, P. J. (1985). *Small states in world markets: Industrial policy in Europe*. Ithaca, NY: Cornell University Press.

Kiyotaki, N., & Moore, J. (1997). Credit cycles. *Journal of Political Economy, 105*, 211–248.

Klette, T. J., & Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy, 112*, 986–1018.

Kuhn, P., & Mansour, H. (2014). Is Internet job search still ineffective? *Economic Journal, 124*, 1213–1233.

Kuhn, P., & Skuterud, M. (2000). Job search methods: Internet versus traditional. *Monthly Labor Review, 123*(10), 3–11.

Kuhn, P., & Skuterud, M. (2004). Internet job search and unemployment durations. *American Economic Review, 94*, 218–232.

Leeson, P. T. (2008). How important is state enforcement for trade? *American Law and Economics Review, 10*, 61–89.

Lentz, R., & Mortensen, D. T. (2008). An empirical model of growth through product innovation. *Econometrica, 76*, 1317–1373.

Levchenko, A., Lewis, L., & Tesar, L. (2010). The collapse of international trade during the 2008–2009 crisis: In search of the smoking gun. *IMF Economic Review, 58*, 214–253.

Levchenko, A., Lewis, L., & Tesar, L. (2011). The role of trade finance in the U.S. trade collapse: A skeptic's view. In J.-P. Chauffour & M. Malouche (Eds.), *Trade finance during the great trade collapse* (pp. 133–147). Washington, DC: The World Bank.

Parente, S. L., & Prescott, E. C. (2000). *Barriers to riches*. Cambridge: MIT Press.

Samuelson, P. A. (1954). The transfer problem and transport costs, II: Analysis of effects of trade impediments. *The Economic Journal, 64*, 264–289.

von Thünen, J. H. (1826). *The isolated state*. Hamburg: Perthes. English translation: Oxford: Pergamon, 1966.