Springer Studies in the History of Economic Thought

Arie Arnon
Warren Young
Karine van der Beek *Editors*

Expectations

Theory and Applications from Historical Perspectives

Springer

Springer Studies in the History of Economic Thought

Series Editors

Harald Hagemann, University of Hohenheim, Stuttgart, Germany

Muriel Dal Pont Legrand, Université Côté d'Azur, Sophia Antipolis Cedex, France

Robert W. Dimand, Brock University, St Catharines, ON, Canada

Hans-Michael Trautwein, Carl von Ossietzky University Oldenburg, Oldenburg, Germany

Advisory Editors

Arie Arnon, Ben-Gurion University of the Negev, Beer Sheva, Israel

Tony Aspromourgos, University of Sydney, Sydney, Australia

Michaël Assous, Lumière University Lyon 2, Lyon, France

Vladimir Avtonomov, Higher School of Economics, Moscow, Russia

Katia Caldari, University of Padova, Padova, Italy

José Luís Cardoso, University of Lisbon, Lisboa, Portugal

Alexandre Mendes Cunha, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Ariane Dupont-Kieffer, Université Paris 1 Panthéon-Sorbonne, Paris, France

Evelyn Forget, University of Manitoba, Winnipeg, Canada

Yukihiro Ikeda, Keio University, Tokyo, Japan

Marianne Johnson, University of Wisconsin Oshkosh, Oshkosh, WI, USA

Heinz Kurz, University of Graz, Graz, Austria

Jean-Sébastien Lenfant, Université de Lille, Lille, France

Qunyi Liu, Peking University, Beijing, China

Maria Cristina Marcuzzo, Sapienza Università di Roma, Rome, Italy

Sylvie Rivot, Université de Haute-Alsace, Mulhouse, France

Margaret Schabas, University of British Columbia, Vancouver, BC, Canada

Bertram Schefold, Johann Wolfgang Goethe-Universität, Frankfurt, Germany

Richard van den Berg, Kingston University, Surrey, UK

Isabella Maria Weber, University of Massachusetts Amherst, Amherst, MA, USA

Carlo Zappia, Università di Siena, Siena, Italy

This series offers an outlet for research in the history of economic thought. It features scholarly studies on important theoretical developments and great economic thinkers that have contributed to the evolution of the economic discipline.

The series welcomes proposals for research monographs, edited volumes and handbooks from a variety of disciplines that seek to study the history of economic thinking and help to arrive at a better understanding of modern economics. Relevant topics include, but are not limited to, various schools of thought, important pioneers and thinkers, ancient and medieval economic thought, mercantilism, cameralism and physiocracy, classical and neoclassical economics, historical, institutional and evolutionary economics, socialism and Marxism, Keynesian, Sraffian and Austrian economics, econometrics and mathematical studies as well as economic methodology and the link between economic history and history of economic thought.

More information about this series at http://www.springer.com/series/16457

Arie Arnon · Warren Young ·
Karine van der Beek
Editors

Expectations

Theory and Applications from Historical Perspectives

Springer

Editors
Arie Arnon
Department of Economics
Ben-Gurion University of the Negev
Beersheba, Israel

Warren Young
Department of Economics
Bar-Ilan University
Ramat Gan, Israel

Karine van der Beek
Department of Economics
Ben-Gurion University of the Negev
Beersheba, Israel

ISSN 2662-6098 ISSN 2662-6101 (electronic)
Springer Studies in the History of Economic Thought
ISBN 978-3-030-41356-9 ISBN 978-3-030-41357-6 (eBook)
https://doi.org/10.1007/978-3-030-41357-6

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

The Thomas Guggenheim Program in the History of Economic Thought has been created in 2007 in the Department of Economics at Ben-Gurion University of the Negev thanks to the generous support of Professor Thomas Guggenheim from Geneva, Switzerland.

The **Program** promotes scholarship in the **History of Economic Thought** and **Economic History**, and it organizes international conferences and funds seminars and fellowships through the **Thomas Guggenheim Fellowships in Economics**.

At each conference, the Program chooses a recipient of the **Thomas Guggenheim Prize for Outstanding and Original Research in the History of Economic Thought**. The prize is awarded to a distinguished scholar for his\her life's work. The prize is awarded in a public lecture delivered by the winner at the Program's conference.

The first conference in 2009 in Beersheba, Israel, was on: "**Perspectives on Keynesian Economics**."

The second conference in 2011 in Beersheba, Israel, was on: "**David Hume and the Scottish Enlightenment: Economic and Philosophical Studies**"—**Celebrating 300 Years of Hume's Birth.**

The third conference in 2015, in Geneva, Switzerland, was on: "**Economic Research and Policy Making at the Federal Reserve in Historical and International Perspective**."

The conference was organized by the Thomas Guggenheim Foundation (Geneva), the Graduate Institute (University of Geneva) and the Guggenheim Program for the History of Economic Thought, Ben-Gurion University (Israel).

The fourth conference in 2017 in Beersheba, Israel, was on: "**Expectations: Theory and Applications in Historical Perspectives**."

As may be seen from the above list of subjects, an attempt is made to choose topics that are related either to actual events or to the modern economic literature. This is clearly the case with the first conference on Keynesian Economics. It was not a coincidence that it took place shortly after the beginning of the 2008 global financial crisis, when Keynesian Economics turned to be more relevant and issues of fiscal and monetary policies became of crucial importance.

The topic of the 2011 conference on David Hume was selected to celebrate his 300th birthday (1711–1776). It turned out to be not only an intellectual event on Hume's philosophy, but a discussion of Hume's genuine ideas on international trade balance and money supply in a time of continuous disruption.

The third conference on Central Banking in 2015 was designated to celebrate the centennial of the Federal Reserve (1913) and the 60-year anniversary of the Bank of Israel (1954).

The conference in 2015 on Expectations in Historical Perspective covered the literature on this extremely important element of economic theory from times of old history to the most recent literature in macroeconomics and its micro-foundations.

The present volume contains the public lecture delivered by the winner at the Program's conference, Prof. Duncan K. Foley, and eleven papers presented in the conference.

Recipients of the Thomas Guggenheim Prize were:

2009—Prof. Bertram Schefold, Goethe-Universität Frankfurt
2011—Prof. Sam Hollander, University of Toronto and BGU
2015—Prof. David Laidler, University of Western Ontario
2017—Prof. Duncan K. Foley, New School for Social Research

The **Thomas Guggenheim Program in the History of Economic Thought** operates under an International Advisory Committee, comprising:

Prof. Thomas Guggenheim (Emeritus, University of Geneva)
Prof. Arie Arnon (Emeritus, Ben-Gurion University of the Negev)
Prof. Maria Cristina Marcuzzo (Sapienza Universita Di Roma)
Prof. Joel Mokyr (Northwestern University)
Prof. Jacques Silber (Bar-Ilan University)
Prof. Karine van der Beek (Ben-Gurion University of the Neger)
Prof. Warren Young (Bar-Ilan University)
Prof. Jimmy Weinblatt (Emeritus, Ben-Gurion University of the Negev)
Prof. Amos Witztum (London School of Economics)

This committee acts as a steering and prize committee and thus is also responsible for the selection of the winner of the Guggenheim Prize.

Jimmy Weinblatt
Ben-Gurion University of the Negev
Beersheba, Israel

Introduction

Expectations—their nature, origins and formation, evolution and dissemination, and fulfillment or disappointment—have always played a prominent place in human activities. Whether dealing with the duality of Homo economicus of Smithian, along with Homo Exspectatione of Aristotelian vintage and whether expressed in mathematical form or as the outcome of background beliefs or "animal spirits," the problem of expectations has come—to paraphrase Marshall—to be one of the "chief difficulties" in analysis of economic problems and processes. These "difficulties" perhaps emanate from the *cognitive dissonance* demarcating expectation and belief that impacts the economic agent or institution. According to most conventional wisdom, an *expectation* is something the economic agent or institution *predicts* will happen. Agents or institutions' *belief* is, on the other hand, something *surmised* to be *true*, *albeit lacking definitive evidence*. Philosophers still disagree as to whether expectations are synonymous with beliefs or if beliefs are prior to expectations, in the sense that the same experience affects our expectations of the future in different ways, depending on which background beliefs we hold. This volume, based on papers presented at the 2017 Thomas Guggenheim Conference, brings together the work of economists, historians of economics and economic historians, on issues and events related to expectations in economics and economic history.

The volume opens with the Thomas Guggenheim Prize Lecture by Duncan Foley. In this, Foley reviews the development of expectations in modern economic thought in the context of his own personal intellectual history. He especially focuses on what has come to be known as the Foley–Sidrauski model and its relation to variant types of expectations. After discussing the development of the rational expectations (RE) macroeconomics, he turns to newer approaches such as nonlinearity, limit cycles, chaos, complexity, catastrophe and bounded expectations, in the context of the role of expectations in complex systems. In this context, he presents a quantal response statistical equilibrium (QRSE) that, in his view, can explain the limitations of the RE approach. In his retrospective conclusion, he delimits the efficacy of expectations theory.

The next set of chapters in Part I focus on the utilization of expectations in the "*ancient*" and "*meso*" periods of *high theory*, that is, from Smithian to Keynesian approaches. Some chapters deal with more "modern" applications of expectations in both "Tobinesque–Phillips" and "Harrodian–Solowian" contexts, and the debate between Friedmanite and Keynesian approaches to expectations formation.

The final set of chapters in Part II are essays on the role of economic expectations in historical events and contexts, ranging from the early twentieth century, through WWII, and then to the application of expectations theory to hyperinflation and stabilization, looking at Israel as a case study.

Witzum's chapter sets out the taxonomy of expectations manifest in *classical* as against *modern* economics. The first case is that dealing with change in the value of economic parameters; the second relates to expectations held by others; the third is agent's assessment of benefit valuation anticipations of outcomes. He concludes that the *modern* approach overlooks both expectations with regard to the behavior of other agents and what agents take to be the "expected consequences" of the expectations of others, as most modern mainstream economists assume *a priori* that agents always behave rationally.

Marcuzzo focuses on the approaches to expectations of three "Cambridge economists"—Marshall, Kahn and Keynes. She notes that Keynes was a student of Marshall and Kahn of Keynes. And, whereas they had a very similar view regarding what she calls the "appropriate" method of theorizing, she makes the cogent point that as regards their views on expectations, the linkage between the three is "less straightforward." She first maintains that their method of inquiry into expectations was *not based* on probability, and then presents what she asserts *was* the basis for their respective approaches. In her view, there are at least *three* approaches to expectations formation. The first is based on expectations of future prices; the second on probabilistic or deterministic models; and the third on surveys and experiments. She concludes that based on the evidence she presents, Marshall, Kahn and Keynes opposed dealing with expectations based on a probabilistic approach. Rather, they viewed expectations and their formation as emanating from *customs, conventions and beliefs.*

Boianovsky focuses on how expectations and their impact on employment were modeled by Champernowne in the 1930s and this in the context of the Pigou–Keynes debate over the nature of employment determination, making some important points in this regard. First, he describes Champernowne's adaptive approach to the relationship between price expectations, whether inflationary or deflationary, and how it impacts on the possible reaction of the central bank, so as to converge to the equilibrium unemployment rate of the economy. Second, Champernowne asserted that this process would *not occur* if uncertainty prevailed in the business and industrial sectors. In other words, in Boianovsky's view, *all those* before Champernowne dealt with role of expectations in asset markets *only*, while Champernowne examined expectations in labor *and* asset markets. According to Boianovsky, while the model of Keynes *general theory* set out by Champernowne in his early 1936 RES paper paralleled many of the characteristics of the later approaches of Harrod, Meade and Hicks, in it, expectations formation

was *endogenous*. Keynes talked about *exogenous* "animal spirits" as driving expectations formation, while his later "interpreters" did not deal with expectations *per se*.

Rivot focuses on the problem of expectations formation during the period of the Great Depression. According to her, many of the early explanations of the Great Depression—such as those of Hayek, Robbins, Fisher and Schumpeter—either ignored or underplayed the impact of expectations. More recently, in her view, what she terms "new classical" analysts such as Lucas and Rapping, and Kydland and Prescott did not place emphasis on the expectation of the Great Depression. Her chapter addresses the Great Depression in terms of what she sees as the Friedman–Keynes divide on expectations formation and its impact over the period. In her view, what characterizes Friedman's approach was the disequilibrium in short-run nominal expectations, while Keynes emphasized dysfunctional long-run expectations.

In their chapter, Assous and del Pont Legrand focus on the *problematic non-inclusion* of expectations in the neoclassical growth and growth accounting framework of what they term the "canonical neoclassical" approach, the Solow–Swan–Meade research program. They present interesting and important evidence that other growth theorists, such as Sen, among others, pointed out the shortcomings of excluding expectations from growth models of Solowian vintage.

Dimand focuses on Tobin's contributions in the context of his approach to expectations, which Dimand maintains flow from *both* Fisherian *and* Keynesian headwaters. According to Dimand, Tobin's treatment of expectations was related to Keynes long-term expectations formation approach combined with Tobin's own approach to market value and replacement cost of equity and capital ("Tobin's q") on the one hand and, on the other hand, to Fisher's treatment of the relationship between nominal and real interests, inflation expectations, along with Fisher's approach to the value of expected earnings and its relation to asset values. Dimand shows that while expectations were at the center of Tobin's work, a degree of "tension" existed between Tobin's approaches to financial market efficiency as against overall systemic stability, reflecting the Fisherian *dissonance* between his view of financial market coordination and his debt deflation theory.

The nexus between expectations formation, prices, wages and inflation in the form of the Phillips–Friedman–Phelps approaches is the focus on Hagemann's chapter. He presents a *typology* of expectations, relating it to inflation in *historical* perspective. He then proceeds to survey extensions, including learning, sticky information and inattention, as applied to inflationary expectations, and then deals with the debates over the efficacy of the Phillips and Friedman–Phelps analytical frameworks.

As noted, Part II of the volume deals with expectations formation and their outcome in times of war, crises, economic dysfunctions and attempts at stabilization.

In their chapter, Oosterlink and Van Gansbeke deal with the impact of expectations regarding the outcome of war on bond prices. They focus on the case study of South African bonds issued during the Second Boer War. Despite the "public

belief" as they put it that the Boers would be defeated, and the military outcome was as expected, bond prices *did not fall, contrary to expectations*, as the British government eventually took over the South African debt.

The chapter by McGregor and Young examines a previously overlooked episode in the history of the Federal Reserve involving the relationship between expectations at the outset of WWII regarding the possibility of German victory in Europe and its economic implications for the US and Western Hemisphere. This is seen in the economic studies conducted by economists such as Hansen at the Board of Governors in the contextual of the Fed's own effort at assessment, and within the framework of that undertaken under the aegis of the interagency "National Defense Program." In the chapter, the changing nature of expectations regarding the war in Europe over the period 1939–1941, that is to say, ***before*** US entry, and its impact on the assessments and economic directions as pointed out by economists at the Fed, shows how expectations of politico-economic events and their outcomes, and economic policy assessments, interconnected.

Razin's chapter focuses on the role inflation expectations played in constraining policy makers trying to regulate inflationary expectations in Israel and the outcome of stabilization. In his chapter, he surveys the transition of Israel from a low-income, high-inflation developing economy to an advanced economy based on its global links.

The final chapter in volume is both a seminal treatment of expectations and a memorial tribute to Allan Meltzer. This chapter, originally an unpublished manuscript by Cukierman and Meltzer dating from 1982, deals with both theoretical and empirical problems. The first issue relates to what they termed the permanent–transitory "confusion" and the case where adaptive expectations may be rational, as applied to market efficiency tests. The second focuses on the application of adaptive expectations to learning and the evolution of expectations during the periods of shock stabilization and, later on, inflation targeting in the Israeli case.

BGU, Israel — Arie Arnon
Bar-Ilan University, Israel — Warren Young
BGU, Israel — Karine van der Beek

Contents

List of Participants

Duncan K. Foley, New School for Social Research, email: duncan.foley@gmail.com
Amos Witztum, HUJI and LSE, email: amos.witztum@gmail.com
Maria Cristina Marcuzzo, Sapienza University of Rome, email: cristina.marcuzzo@uniroma1.it
Mauro Boianovsky, Universidade de Brasilia, email: mboianovsky@gmail.com
Sylvie Rivot, Université de Haute-Alsace, email: rivot@unistra.fr
Robert W. Dimand, Brock University, Canada, rdimand@brocku.ca
Michaël Assous, Université Lyon 2, CNRS, Triangle, email: Michael.Assous@univ-lyon2.fr
Muriel Dal Pont Legrand, Université Nice Sophia Antipolis, email: muriel.dal-pont-legrand@univ-cotedazur.fr
Harald Hagemann, University of Hohenheim, Stuttgart, email: hagemann@uni-hohenheim.de
Kim Oosterlinck, Université Libre de Bruxelles, email: koosterl@ulb.ac.be
Marie Van Gansbeke, Université Libre de Bruxelles, email: marie.vangansbeke@gmail.com
Rob McGregor, University of North Carolina at Charlotte, email: rrmcgreg@uncc.edu
Warren Young, Bar-Ilan University, email: youngwprof@yahoo.com
Assaf Razin, TAU—Tel Aviv University, email: razin@post.tau.ac.il
Alex Cukierman, TAU and Interdisciplinary Center, email: alexcuk@tauex.tau.ac.il
Thomas Lustenberger, University of Basel and Swiss National Bank, email: thomas.lustenberger@unibas.ch
Allan Meltzer, Carnegie Mellon University; Allan Meltzer passed away on May 8, 2017

Part I
Guggenheim Prize Lecture

Unfulfilled Expectations: One Economist's History

Duncan K. Foley

1 What Did You and I Expect?

At the risk of presuming on your generous invitation, I will organize this discussion of expectations in economics around my personal intellectual history.

As an undergraduate at Swarthmore College in the early 1960s under the tutelage of William Brown and Joseph Conard, I read enough of Hicks' *Value and Capital* to understand that the translation of Walrasian general equilibrium to an inter-temporal setting required some kind of assumption about agents' views of future prices. Much of what passed for my thinking in those days was centered on macroeconomic analysis, which tended to treat expectations casually as a secondary aspect of dynamic analysis. "Static" expectations were the default assumption, and the notion of "adaptive" expectations was a somewhat exotic and arcane innovation.

In the fall of 1964, I found myself in the very different, intellectually demanding, and exciting world of Herb Scarf's course in Mathematical Economics at Yale (Foley 1999). Scarf's main interest at the time was the computation of general equilibrium, he included Debreu's *Theory of Value* in our reading list, and we discussed the interpretation of general equilibrium in inter-temporal terms through the re-labeling of commodities to include time dating. This was of secondary interest to Scarf, since it had little to do with the mathematical structure of the general equilibrium concept.

Right from this beginning, it struck me that the key point was that in the real world there were not futures markets in all commodities, for one thing because the commodity space is constantly changing. Arrow's attempt (1964) to economize on markets through "securities" seemed to be going in the right direction, but I knew there still just were not enough markets in the real world to provide the necessary price information to evaluate arbitrary inter-temporal investment or consumption plans.

D. K. Foley (✉)
Department of Economics, New School for Social Research,
79 Fifth Avenue, New York, NY 10003, USA
e-mail: foleyd@newschool.edu

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_1

One implication of this that struck me was that the lack of futures markets left a large hole in the overall general equilibrium picture of consumer sovereignty. In full equilibrium with complete markets, one could see that household preferences and endowments ultimately determined resource allocation, on the assumption that firms were passive price-takers. But with incomplete markets, the investment decisions of firms would have to depend on their subjective valuation of alternate investment plans.

The missing markets seemed to me to be an insurmountable obstacle to deriving a usable macroeconomic theory from general equilibrium foundations. Much of my research effort during the late 1960s and early 1970s pursued this problem. An example is the paper *Economic equilibrium with costly marketing* (Foley 1970), which shows why markets in long-distant future or low-probability contingency commodities are economically unviable.

2 M.I.T. and Miguel Sidrauski

The problem of expectations became a major preoccupation during my years at M.I.T. The dominant figure in my first three years at M.I.T. was Miguel Sidrauski, whom M.I.T. had hired from Chicago in the same wave of junior faculty with me. Sidrauski and I hit it off personally and intellectually and began a collaboration that eventually produced a book, *Monetary and Fiscal Policy in a Growing Economy* (Foley and Sidrauski 1971).

Sidrauski was the kind of person who changed people's lives and also could beat them (or at least me) regularly at chess. He was a few years older than I, an Argentinian of Jewish descent, married to an equally remarkable Argentinian woman, Martha Sidrauski. I realized how passionate was Miguel's identification with Israel during the 1967 "Six Day War," during which he gave the very highest priority to following bulletins from the battlefield. Sidrauski's tragic early death from an untreatable cancer was an irreplaceable loss to me in personal and intellectual terms and, I think, a shaping event in the history of macroeconomic theory (Fig. 1).

Sidrauski and I set out to synthesize a macroeconomic growth model incorporating key ideas we each brought from our graduate educations. Our research assistant for much of this work was Stanley Fischer. One central theme of this work was to clarify stock-flow relationships, which were treated rather casually and inconsistently in much of the macroeconomic literature of the 1960s. Another main focus was to follow up Keynes' notion that interest rates and financial asset prices were determined by "liquidity preference," which we interpreted as stock demands and supplies of financial and real assets.

The model assumed full employment of labor and capital in a two-sector production system that produced a production–possibilities frontier between consumption goods and investment goods strictly convex to the origin. Instantaneous equilibrium

Fig. 1 Miguel Sidrauski

in stock markets for real and financial assets determined a "price of capital" in terms of consumption goods, which in turn led to a profit-maximizing division of productive resources between investment and consumption. The housing sector exemplifies this conception.

This setup has much in common with Jim Tobin and Bill Brainard's analysis of what has come to be known as "Tobin's q" (Brainard and Tobin 1968). Sidrauski and I became aware of this through an anonymous referee report from the *American Economic Review* requesting more adequate citation of Tobin's work in the summary paper we wrote describing the model.

In the Foley–Sidrauski model, the "fiscal-monetary mix" determines the allocation of social resources between consumption and investment, and hence, the rate of capital deepening in the economy. The framework has the formal advantage of making stock-flow relationships transparent and consistent.

Students appreciated the clarity this framework brought to macroeconomics, and the book served as a key text in a few graduate departments until the "rational expectations revolution" overtook the profession. Not everyone was enthusiastic, however. Luigi Pasinetti took a dim view of treating investment as a passive profit-maximizing response of the suppliers of capital goods, as well as of the implicit full-employment assumptions of the model. Peter Diamond viewed the whole project as insufficiently grounded in basic theoretical ideas and a jejune attempt at synthesis through juxtaposition of inconsistent concepts. I found both of these critiques quite persuasive myself.

Because expectations of changes in the price level and asset prices affect relative rates of return of financial and real assets, Sidrauski and I had to make expectations assumptions to close the model. In the simpler pedagogically oriented versions of the model, we adopted the assumption of adaptive expectations, which have the same structure as the stock-flow dynamics in which the rest of the system were analyzed. We were aware of the rather awkward implication of adaptive expectations that expectations could be systematically and indefinitely wrong on steady-state growth paths.

One effort we made to address these problems was to study the model under the assumption of perfect foresight expectations (greatly helped by Karl Shell, who had studied this type of modeling under the guidance of Hirofumi Uzawa) (Foley et al. 1969). In this effort, we encountered some key issues that played a central role in the later development of macroeconomic models. In particular, we discovered that the dynamic equilibria of perfect foresight models were saddle points.

While we could (especially with Shell on hand) work through the mathematics of perfect foresight assumptions, we were stumped to make economic sense of the resulting analysis. In order to avoid economically impossible paths involving negative prices or quantities, it was necessary to assume that in response to a shock the economy would jump instantaneously to the stable manifold of the post-shock system. We could see that this was mathematically consistent, but not what plausible behavior the mathematics might represent, even in the abstract.

Sidrauski saw in his earlier work the relevance of perfect foresight version to the issue of the long-run neutrality of money that he had encountered in his studies at Chicago. His two classic papers curiously embody two contrary, or complementary, views.

Inflation and growth Sidrauski (1967b) models the relation of money to the real economy on lines close to those developed by Tobin and the Foley–Sidrauski model. Money and capital are substitutes in portfolios, and a higher rate of inflation, by lowering the real rate of return to money, raises the stock demand for capital and leads to more rapid investment. In this world, money is not neutral in the long run.

Rational choice and patterns of growth in a monetary economy (Sidrauski 1967a) in contrast sets up a Ramsey-type inter-temporal representative agent model, in which the services of real money balances enter the utility function. In this model, the representative agent "sees through" changes in the rate of inflation, and money is neutral both in the short and long run. This paper establishes Sidrauski as a precursor, and plausibly as an originator, of what later came to be known as real business cycle theory.

Working through the Foley–Sidrauski model emphasized to me the close connection between stock-flow equilibrium and fulfilled expectations. *On two specifications of asset equilibrium in macroeconomic models* (Foley 1975, 1977) summarized these relationships in terms of "beginning of period" (stock) and "end of period" (flow) equilibria. When asset markets meet both stock and flow equilibrium conditions, expectations have to be fulfilled over the period.

I, somewhat naively in retrospect, thought that these observations would lead to a serious methodological discussion of the primacy of stock or flow equilibrium concepts. As it turned out, the mainstream of the profession chose the modeling option of assuming fulfilled expectations (under the somewhat misleading terminology of "rational" expectations). This decision puts macroeconomics back into the box of attained equilibrium theory from which Keynes had sought to liberate it.

3 Rational Expectations

My reading of the rational expectations "revolution" in the 1970s, as it unfolded, differed from that of most economists interested in macroeconomic problems. To begin with, I interpreted the formalism of rational expectations as a variation on general equilibrium with complete futures and contingency markets, rather than as a substantive hypothesis on the formation of agent expectations in real economies.

In principle, there is no reason to reject the abstract picture of inter-temporal general equilibrium with complete markets for commodities contingent on states of nature independent of human action out of hand, though its relevance to real-world economies is highly questionable. The limitation of contingency to states of nature independent of human decisions rules out all of the sources of endogenous uncertainty, related to fallacies of composition and strategic interaction, on which Keynes' thinking rested.

In this rather limited setting, one could imagine an economy of Savage-like Bayesians, ready to lay bets against each other on questions such as the weather, the distribution of mineral deposits, and the discovery of new scientific principles. A general equilibrium system of prices for these contingencies could emerge, which would allow all the agents to evaluate investment and consumption plans. New information could alter the equilibrium prices.

These prices would be consistent in the usual Walrasian sense that the decisions of price-taking agents based on them would be market clearing. But if the Bayesian agents had different priors or information, these equilibrium prices would not in general represent a single consistent "objective" probability distribution describing the actual evolution of the "state of the world." From an informational point of view, there would be no omniscient "representative" agent.

Furthermore, even in the narrow sense of predicting the weather, the distribution of mineral resources, and other states of nature independent of human action, complete markets prices are not necessarily self-fulfilling. Things might happen that would constitute new information for the Bayesian decision-makers and lead them to change their excess demands for contingent future commodities.

These considerations led me to conclude that the assumption of rational expectations was unlikely to lead to theories that would throw much light on the core problems of macroeconomics. My priors led me to think of macroeconomic problems like involuntary unemployment and output fluctuations as reflecting coordination problems characteristic of monetary economies and the limitations of finance as a coordinating mechanism.

The claimed political–economic implications of rational expectations have little to do with individual behavior in real-world economies. The strong implications of the doctrine (e.g., policy-ineffectiveness, efficiency of equilibrium paths) flow from its assumptions of Walrasian price-taking equilibrium and that expectations must be self-fulfilling, which is as much an assumption of equilibrium paths as an assumption about individual behavior.

4 Dynamics and Complexity

The next phase of my thinking about economic theory was centered on the problems of dynamics. My original motives for exploring these questions arose from my work on classical and Marxist political economy, which concerned the behavior of firms. Nonetheless, it turned out that these investigations led to issues highly relevant to the problem of expectations formation.

My first idea was to disaggregate Marx's concept of the circuit of capital to analyze the interaction of an economy composed of a large number of capitalist firms. The simplest formal representation of Marx's circuit of capital leads to a linear dynamical model composed of integral equations (rather than the more familiar differential equations used widely in economic dynamics). "Interesting" dynamics in this framework arise with the introduction of nonlinear interactions.

In the 1980s, interest in pursuing dynamic questions pioneered by Richard Goodwin revived among a number of researchers from various backgrounds in mathematics and economics, including Richard Day and Jess Benhabib, Alfredo Medio and Marji Lines, Peter Flaschel, Willi Semmler and Reiner Franke, and Bill Brock, to name only a representative few. Even "small" nonlinear perturbations of economic models can lead to qualitatively new behavior, such as limit cycles, chaos, and adaptive, self-organized complex systems far from equilibrium.

My own thinking on these questions became intertwined with work with Peter Albin (Fig. 2), who, beginning with research on technology, productive labor organization and distribution (Albin 1978), became a pioneer in the application of Stephen Wolfram's version of cellular automata to economic modeling (Albin 1975). Albin was particularly interested in the implications of complex systems for expectation formation as the essays collected in *Barriers and Bounds to Rationality* (Albin 1998) emphasize.

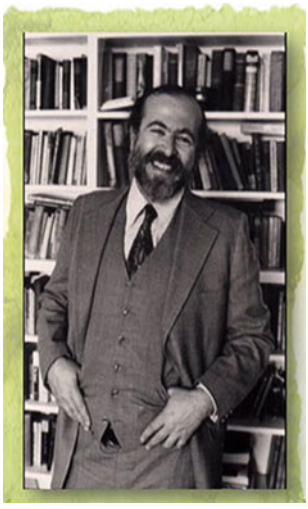

Fig. 2 Peter Albin

Albin was an avid games player and would beat me not only in chess but also in scrabble and any other competition he could inveigle me to try. This one-sided strategic relationship continued after he suffered a disabling stroke in 1991 at the age of 53. (His disappointment at my inability even to learn the principles of Go was touching.) Albin was also very much a New Yorker and particularly prized the rich variety of food and drink the city offers. In our shared research, Albin tended to push the general and abstract results of complexity theory, while I leaned toward applying them to classic problems of political economy, a line of thought that has had long-lasting repercussions for me.

The introduction of nonlinearities in dynamical systems leads to three broad classes of dynamic outcomes: limit cycles and other regular periodic motions; chaos; and self-organized complex adaptive behavior. Limit cycles are suggestive, as Richard Goodwin masterfully showed, of the dynamics behind real-world economic fluctuations, but are too regular to be plausible representations of real-world fluctuations.

Chaotic motions of deterministic dynamical systems exhibit statistical regularities and, like stochastic processes, can in principle provide information from which expectations in the sense of forecasts about the future can be extracted. The main difference between deterministic chaos and stochastic randomness is that chaotic trajectories are typically confined to a low-dimension set within a system state space, while stochastic processes "fill" the whole space.

Macroeconomic time series data is too sparse and too noisy to allow us to distinguish reliably whether it arises from deterministic chaos. Financial data, on the other hand, is very abundant and does provide information of this type, which has been exploited by sophisticated financial traders in profitable trading strategies.

Albin was particularly interested in complex systems in the sense of adaptive, self-organized systems far from equilibrium. These systems have strong elements of both stabilizing negative feedback, which are strong enough to prevent them from being completely chaotic and destabilizing positive feedbacks, which eventually disrupt quasi-stable patterns of behavior. Complex systems are sometimes described as being "on the edge of chaos."

Complex systems produce dynamics in which sustained periods of statistical regularity are disrupted and replaced endogenously by possibly qualitatively different types of statistical regularity. Albin was particularly interested in *cellular automata*, a class of models, including, for example, the "Game of Life" in which dynamics are generated by simple rules. It is possible to show that some cellular automata, including the "Game of Life" are *computationally irreducible*, which means that it is impossible to forecast their behavior except by computing their trajectory, or the trajectory of a system of the same complexity.

As Albin realized, these abstract mathematical results have sobering consequences for the theory of expectations in real-world economies. Real-world capitalist economies, like evolutionary biological systems and the brain, have the characteristic properties of complex systems far from equilibrium. But we know that it is impossible to form detailed expectations about the dynamic behavior of such systems.

As a result of my work with Albin, I came to understand the history of economics, and particularly the methods and conclusions of the classical political economists in a new light (Foley 2003). Smith, Ricardo, Malthus, and Marx conceive of the capitalist economy in terms that are at least consistent with the complex systems approach, as a turbulent, far from equilibrium system. The long-period method they adopted is aimed at understanding the self-organization of the capitalist economy, through the emergence of phenomena like an average rate of profit, natural prices of commodities, and the like.

The complex systems perspective also puts the question of expectation in a new perspective. For example, we might have much more confidence that there will be a bond market and an interest rate fifty years from now than in any particular forecast of the level of the interest rate. The resistance of complex systems to detailed quantitative forecasts of their trajectories does not preclude our understanding as important qualitative features of their self-organizing behavior.

The conceptual world of complex adaptive systems far from equilibrium raises issues at a completely different level from the linear forecasting models through which Milton Friedman, Richard Muth, and their successors approached the problem of expectations. Complex systems analysis suggests that the problem of forecasting concrete historical trajectories for a capitalist economy is more like the problem of forecasting the weather or the behavior of organisms with highly developed brains.

Because complex systems theory invokes somewhat arcane notions like computational irreducibility, it may help to illustrate the issues in a context easier to visualize. In Fig. 3, the horizontal axis represents the average choice of some economically relevant variable, such as the bid price for an asset, or the level of spending over the economy. The vertical axis represents the response of typical (one of many identical) agents in choosing the level of the same variable. Equilibrium occurs on the 45° line because the agents are identical.

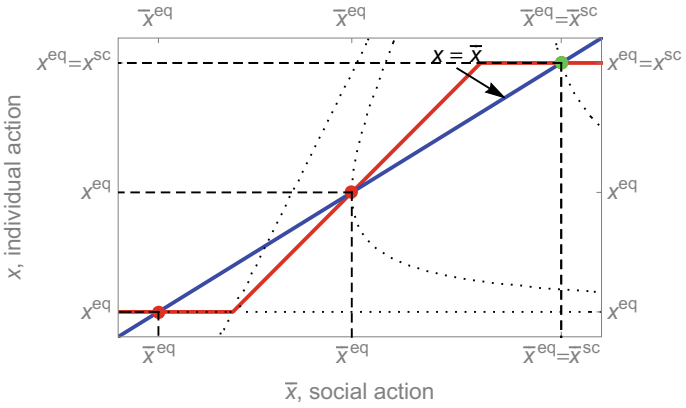

Fig. 3 Multiple equilibria. The basin of attraction of a stable equilibrium shrinks due to drift of the typical agent best response

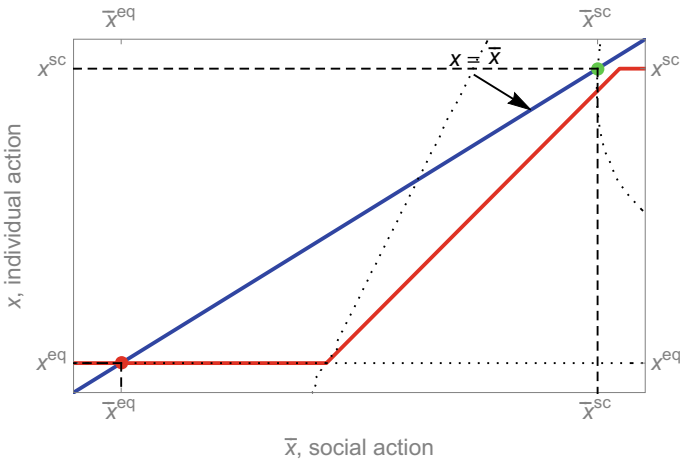

Fig. 4 Cusp catastrophe. After the shift in the best response, the system moves irreversibly and chaotically to the remaining stationary equilibrium

Because of the nonlinearity of the typical agent's response, there may be multiple stable equilibria. A "slow" change in some other variable may lead the typical agent response to drift downward, eventually annihilating the high-level equilibrium. If the high-level equilibrium ceases to exist, the system crashes to the remaining stable low-level equilibrium, as Fig. 4 illustrates.

The recurrence of events of this type makes the idea of "rational expectations" highly suspect for macroeconomic behavior, as Keynes seems to have concluded at an early stage from his investigation of the fundamental principles of statistical inference.

Because macroeconomic outcomes consist of substantial periods in which the system remains close to a stable equilibrium and as a result does not reveal much dynamic structure, punctuated by crises that are hard to track in data and not replicable enough to provide much leverage for statistical inference, the project of reconstructing a "rational expectations" model of macroeconomies statistically faces enormous obstacles.

The empirical investigation of systems of this type faces the additional problem that the response of the typical individual may sometimes drift close to annihilating the high-level equilibrium and then retreat without a catastrophe. It may be very difficult quantitatively to distinguish cases where a crisis is so likely as to be unavoidable from cases where the system "brushes" close to crisis without actually encountering a crisis.

The complex systems vision of the real-world economy suggests elements of self-organization that may result in stability and predictability over substantial periods of time. But slower destabilizing feedbacks are also present in complex systems and lead to complex motions of response curves which can threaten and sometimes annihilate stable temporary equilibria. In this kind of world, it may be more "rational"

to aim to understand the stabilizing forces without presuming to forecast catastrophic punctuating events with any confidence.

It seems likely to me that Keynes had a scenario like this in mind in his evaluation of the likely success of econometric analysis of macroeconomic time series, for example, in his critique of Tinbergen's methodological program.

Faced with these obstacles, Herb Simon's pragmatic problem solvers have some alternatives to trying to predict exactly the future trajectory of prices or quantities. They might, for example, focus on those aspects of economic performance that are robustly self-organized and therefore more predictable. As in the case of weather prediction, the best one can do in economic forecasting is probably some combination of long-run average features of the system modified by whatever short horizon information is available.

5 Two Sorts of Expectations

Economics is a social science and deals with the social life of human beings. Human beings are forward-looking and often act with an eye to the consequences of their actions. *Expectations* can refer either to expectations about the behavior of other participants in social interactions or to expectations about the concrete future outcomes of the interactions.

One property of *equilibrium* concepts, which also play a major role in economics, is that in a state of equilibrium the expectations of participants are fulfilled. If expectations of participants are not fulfilled, they will in general change their behavior and the state of the system. The fulfillment of expectations about the behavior of other participants is the organizing idea of *Cournot-Nash equilibrium* (CNE). The fulfillment of expectations about concrete future outcomes is the organizing idea of *perfect foresight or rational expectations equilibrium* (REE).

The great insight into Cournot-Nash equilibrium is that even when expectations about the behavior of other actors are fulfilled, the resulting outcome may not be *socially coordinated*, that is, an outcome one of the participants would choose given the power to control other participants' actions. Familiar examples are Prisoners' Dilemma games, in which the unique CNE is not socially coordinated, and Assurance games, in which there are multiple CNE but only one is socially coordinated.

Rational expectations, that is, fulfilled expectations of future concrete outcomes, are not sufficient to guarantee socially coordinated outcomes. An example is the model of climate change as an externality analyzed in (Rezai et al. 2011). In this model, identical producers who correctly predict the path of climate damage emit excess greenhouse gases because each believes her emission contributes only a negligible part of the externality and that other producers will continue to emit even if she does not.

6 Boundedly Rational Expectations

Recent developments in the application of information theory to rational choice throw some light on the paradoxes of rational expectations theory. Traditional economic theory, which constituted Milton Friedman's intellectual formation, assumes that decision-makers have complete and transitive preferences (or equivalently a payoff or utility function) over outcomes and choose the action that maximizes their payoff. This turns out not to be true either experimentally or empirically.

When confronted with a decision, for example, whether to buy or sell an asset at a given rate of increase of its price, subjects in experiments, like real-world actors, only approximate the behavior posited in traditional economic theory. Traditional economic theory assumes that the subject has a payoff (say, the difference between the rate of increase in prices in a particular market and her estimate of the "fundamental" rate of price increase of the asset). The maximizing agent will never buy the asset when prices are rising faster and will always buy the asset when prices are rising slower, than her estimate of its fundamental rate of price increase.

Real human beings approximate this behavior in the sense that when the rate of price increase is significantly below their fundamental, they almost always buy (and almost never sell), and when the rate of price increase is significantly above their fundamental they almost never buy (or almost always sell). But for some intermediate range of price increase close to the fundamental, the subject buys and sells with frequencies that depend on the rate of price increase according to the logistic S-curve. This behavior has been studied by psychologists since at least the work of Duncan Luce and his associates (Suppes et al. 1989), is the central focus of Charles Manski and Daniel McFadden's analysis of empirical models of behavior (McFadden 1976) and plays a key role in Richard McKelvey and Thomas Palfrey's theory of quantal response equilibrium in noncooperative games (McKelvey and Palfrey 1995).

From an information theory perspective, quantal responses following the logistic function make sense, because they maximize expected utility given a mixed strategy over the choices (e.g., buy and sell) subject to a lower bound on the informational entropy of the frequency distribution describing the mixed strategy. Given finite information processing capacity, the achievement of the zero-entropy behavior implied by the exact maximization posited by traditional economic decision theory is impossible. But subjects who behave according to the logistic quantal response appear to violate the decision theory axioms of completeness and transitivity of preferences because given the identical choice situations they sometimes do one thing and sometimes the other.

Ellis Scharfenaker and I (2017) have recently developed a model of quantal response statistical equilibrium (QRSE) that can explain the frequency distribution of observed transaction prices in some markets and can explain the limitations of Friedman–Muth concepts of rational expectations. In the QRSE model, a population of identical participants with the same fundamental, μ, and the same quantal behavior temperature, T, buy and sell some asset in a large number of separate sub-markets. The fundamental in this case is the rate of increase in the asset's price. The frequency

with which the typical participant buys the asset depends on the actual rate of price increase in the sub-market, p, according to the logit quantal response function:

$$f[\text{buy}|p] = \frac{\exp[\frac{p-\mu}{T}]}{1+\exp[\frac{p-\mu}{T}]} \tag{1}$$

QRSE further assumes that the individual acts of buying or selling have a nonzero impact on the rate of price increase in the sub-market. Buying raises the price faster and selling slows it down. This impact of actions on outcomes limits the difference in expected outcome (rate of price appreciation) conditional on buying and selling, which constrains equilibrium market distribution frequency distribution. The Lagrange multiplier corresponding to this constraint is β. Without further constraints, the predicted frequency distribution of price increases is symmetrical with mean μ. If, however, the actual mean price, ξ, differs from μ, another constraint with corresponding Lagrange multiplier γ induces the skewness necessary to make the equilibrium distribution fit the actual mean. The maximum entropy statistical equilibrium distribution of price increases, writing $H_{\mu,T}[p]$ for the entropy of the logit quantal response, is as follows:

$$f[p;\mu,\beta,\gamma,T] \propto \exp[H_{\mu,T}[p]]\exp[-\beta\tanh[\frac{p-\mu}{2T}]p]\exp[-\gamma p] \tag{2}$$

The fundamental rate of price increase of the asset μ can be regarded as the expectation of the market participants. It is the rate of price increase at which they buy and sell with equal frequency. When participant expectations are correct, their fundamental will equal the actual mean rate of price increase in the market, $\mu = \xi$ and $\gamma = 0$. In this case, the equilibrium frequency distribution of transaction price increases will be symmetrical, and the transaction rate of price increases, though they will differ in sub-markets, will average out to the fundamental expected by the typical participant as Fig. 5 illustrates. This approximates Friedman–Muth's picture of "rational" expectations, in that the outcome of the market creates no incentives for the typical participant to change her estimate of the fundamental.

When the typical participant's expectations of the fundamental rate of price increase of the asset turn out to be incorrect, $\mu \neq \xi$, the typical participant is buying at "too low" or "too high" a rate of price increase given the actual market outcome, as illustrated in Fig. 6. This is the situation Friedman envisioned as leading to market punishment of incorrect expectations. In the QRSE model, this is exactly what occurs, because when $\mu \neq \xi$, $\gamma \neq 0$ and the equilibrium market price increase distribution is skewed, leading to a long tail that imposes losses on the typical participant relative to the mean market rate of price increase. Where Friedman errs is in presuming that the resulting correction of expectations by market punishment is instantaneous and costless, so the expectations can be identified with actual market outcomes. The information theory-based QRSE model validates the underlying insight into rational expectations theory, but not its overly strong conclusions.

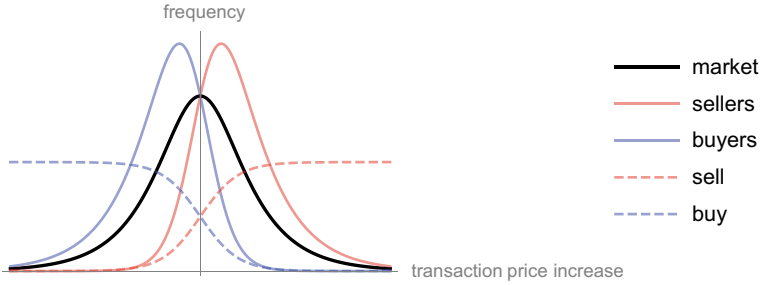

Fig. 5 QRSE market with self-fulfilling expectations. Rates of transaction price increases are on the horizontal axis and frequencies on the vertical axis. The fundamental is marked by the vertical axis. The black curve shows the market distribution of rates of price increase, and the blue and red solid curves the buyers' and sellers' distributions. The dotted blue and red curves visualize the quantal response of buyers and sellers. Buyers on average transact at lower rates of price increase than sellers (the "endowment effect"), but the fundamental coincides with the mean and modal market rate of the price increase

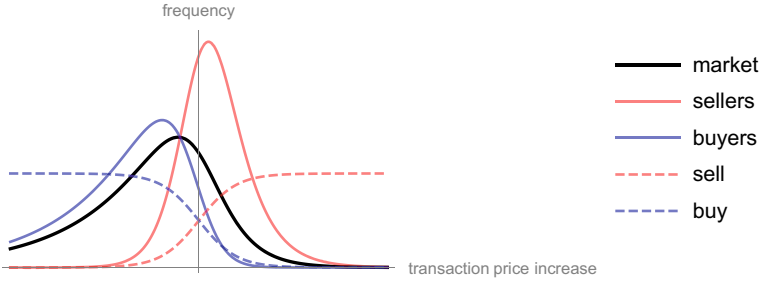

Fig. 6 QRSE market with unfulfilled expectations. The market distribution of transaction price increases is skewed to the left, leading to a lower modal and mean transaction rate of price increase than the fundamental expected by the typical transactors. In order to induce transactions, buyers benefit from much lower rates of transaction price increases. In this way, the market penalizes transactors for basing their behavior on a fundamental that is not confirmed by market outcomes

Prices, like other valuable information, come at a cost. Information theory, by quantifying information in the form of entropy, keeps the real costs of information visible even in equilibrium. Traditional economics accepts as valid conclusions that are non-generic knife-edge cases and are not robust to the realistic introduction of information and information costs.

7 Hindsight: Intellectual Odyssey or Wild Goose Chase?

Perhaps there is not in the end as much to be learned from studying expectation formation as economists hoped. There are aspects of all systems about which we

have limited, incomplete, and noisy information. We have no choice but to treat these aspects of the systems we study as random and at best weakly constrained. The idea that expectations are likely to reflect widely available information puts this type of weak constraint on behavior.

One motivation for studying expectations in macroeconomics was the suspicion that some macroeconomic fluctuations reflect self-fulfilling prophecies. This seems likely to be the case and is one of the endogenous sources of fluctuations in a capitalist economy. Much of this insight can be exploited theoretically in models of social interaction in which the typical agent's behavior depends on the behavior of the other interactors. Contagious expectations in this light are one of a number of interactive factors that can lead to multiple equilibria.

Theoretical focus on expectations formation arises from thinking that the economy as a whole somehow reflects the behavior of a typical or "representative" agent. This idea is strengthened because in many ways the economy as a whole functions analogously to a Robinson Crusoe household economy in allocating resources. It was perhaps inevitable that the investigation of the economy as a whole would start with some construction like Walras', which naively attempts to aggregate individual firm and household decisions through markets.

It is perhaps harder to understand why economic conceptions of the whole have not advanced much beyond this primitive level, particularly given the alternative model of statistical mechanics. What is interesting about the typical agent in a complex economy is precisely that her outcome depends crucially on other agents' behavior over which she has no control, in sharp contrast to Robinson Crusoe. Instead interactions between economic agents are mediated by complex social institutions, including money, finance, and markets.

When Robinson's expectations change (e.g., he sees a storm approaching on the horizon), he can change his behavior accordingly (e.g., to delay work on his house in order to get vulnerable crops to safety). When the expectations of agents in a complex economy change, however, economic institutions may not respond in a linear fashion. One set of agents may speculate against the change in others' behavior, to some degree offsetting their changes in behavior.

Under these circumstances even if we discover quite a lot about the formation of agent expectations, we might not learn very much about the behavior of the economy or economic policy. From this point of view, the theory of expectations is most valuable for exposing the methodological weakness of economic conceptions of the whole and pointing us toward more sophisticated ways of understanding aggregation.

Acknowledgements I'd like to thank Sam Bowles for extensive conversations that clarified my thinking on these issues.

References

Albin, P. S. (1975). *The analysis of complex socioeconomic systems*. Lexington, MA: Lexington Books.

Albin, P. S. (1978). *Progress without poverty: Socially responsible economic growth*. New York: Basic Books.

Albin, P. S. (1998). *Barriers and bounds to rationality: Essays on economic complexity and dynamics in interactive systems*. Princeton: Princeton University Press.

Arrow, K. J. (1964). The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies*, *31*(2), 91.

Brainard, W., & Tobin, J. (1968). Pitfalls in financial model building. *American Economic Review,56*, 99–122.

Foley, D. K. (1999). The ins and outs of late twentieth-century economics. In A. Heertje (Ed.), *Makers of Modern Economics* (pp. 70–118). IV Brookfield, VT: Edward Elgar.

Foley, D. K., & Sidrauski, M. (1971).*Monetary and fiscal policy in a growing economy*. New York: Macmillan.

Foley, D. K. (1970). Economic equilibrium with costly marketing. *Journal of Economic Theory*, *2*(3), 276–291.

Foley, D. K. (1975). On two specifications of asset equilibrium in macroeconomic models. *Journal of Political Economy*, *83*(2), 303–324.

Foley, D. K. (1977). Correction. *Journal of Political Economy*, *85*(2), 401–402.

Foley, D. K. (2003). *Unholy trinity: Labor, capital, and land in the new economy*. London: Routledge.

Foley, D. K., Sidrauski, M., & Shell, K. (1969). Optimal fiscal and monetary policy and economic growth. *Journal of Political Economy*, *77*(4), 698–719.

McFadden, D. L. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior,* 10(1), 6–38.1995.

Rezai, A., Foley, D. K., & Taylor, L. (2011). Global warming and economic externalities. *Economic Theory, 49*(2), 329–351. https://doi.org/10.1007/s00199-010-0592-4. https://doi.org/10.1007/s00199-010-0592-4. ISSN 0938-2259.

Scharfenaker, E., & Foley, D. K. (2017). Quantal response statistical equilibrium in economic interactions: Theory and estimation. *Entropy, 19*(9). https://doi.org/10.3390/e19090444. URL http://www.mdpi.com/1099-4300/19/9/444. ISSN 1099-4300.

Sidrauski, M. (1967a). Rational choice and patterns of growth in a monetary economy. *The American Economic Review*.

Sidrauski, M. (1967b). Inflation and economic growth. *Journal of Political Economy*.

Suppes, P., Krantz, D. M., Duncan Luce, R., & Tversky, A. (1989). *Foundations of measurement. Geometrical, threshold, and probabilistic representations*.

Part II
Expectations and Economic Thought: Classics and Moderns

Expectations and Its Variants: The Nuanced Role of Expectations in Classical Economics

Amos Witztum

1 Introduction

The term expectations, to a modern economist, means the views which decision makers have about the future values of economic variables. These expectations usually suggest a certain understanding of the way in which the economic system determines these values. However, there are two additional, implicit, expectations associated with this which raise doubts about the validity of the understanding upon which the initial expectations (about the values of economic variables) are based. I am referring here to the expectation that other agents will all behave in a particular rational manner and an expectation about the benefits which would be accrued when these values materialise. To distinguish between these three elements which implicitly make up the expectation as is understood in economic analysis, we call the expectation regarding the future value of an economic variable: ***value expectation*** (or, v-expectation); the expectation about the behaviour of other agents: ***procedural expectation*** (or, p-expectation); and the expectation about the benefits: ***consequential expectation*** (or c-expectation).

Thus, by way of example, in a competitive set-up, agents—who presumably know the values of the parameters that determine demand and supply—expect the value of the price to be the equilibrium value assuming that all other agents behave like they do, and, more specifically, have the same expectations about the value of the equilibrium price; as well as expect the benefits which are embedded in the equilibrium distribution of consumer and producer surpluses. Clearly, if agents did not expect to benefit from the expected value, they would behave very differently if they knew that the expected behaviour would lead to its materialisation. This means that the v-expectation is predicated on the p-expectation as well as on the c-expectation. Equally, each other element of expectation is also predicated on the two others.

A. Witztum (✉)
CPNSS, London School of Economics, Houghton Street, London WC2A 2AE, UK
e-mail: amos.witztum@gmail.com

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_2

Leaving aside the more general critique which was raised by Lachmann (1943) regarding the nature and significance of economics if the objects of investigation become increasingly aware of the premises or the mechanisms of the system in which they serve as pawns,[1] all of this sounds somewhat Hegelian in the sense of what he called ***the cunning of reason*** which is the way in which reason extracts universal rules from what appears to be chaotic particular expressions of that rule. Hollis (1987) provides an analysis of the predicament of rationality in general, and rational expectations in particular, as far as economic analysis is concerned. At one point, he provides an insightful allegory to the cunning of reason and the role of expectations in economics. He tells the story of the distinction between a weather forecast for shipping and traffic forecasts. In the case of the former, the expectation is about the weather system which is independent of what the ships—at which it is directed—will do. The probability of the forecast of a gale in an area of the sea will not be affected by whether ships take notice or not of the warning. Traffic forecasts, on the other hand, are not independent of what the motorists will do. If there is a forecast of the build-up of a traffic jam in a certain part of town, this may not materialise if motorists take notice. But in spite of the fact that in both cases, people can avoid the predicament through their own actions (the ships steering away from the affected area and motorists doing the same), in the latter case, their actions would affect the state of the world. The movements of the ships will not affect the probability of a gale developing in the area, but the behaviour of the motorists makes the event disappear altogether. However, the implied expectations here are limited to the event itself (the v-expectation) and not to the behaviour of other agents (which is also predicated on the benefits) but this is not the case in economic analysis when expectations (in particular, rational) are considered. In the case of these expectations, we have something which is akin to Kant's universal test. If motorists expected (p-expectation) that everyone will now use an alternative route this may lead them to conclude that the traffic jam will not materialise and as a result stick to their original plan. In such a case, the traffic jam will indeed materialise. These sensitivities of the outcomes constitute a reason to suspect, as Hollis does, that rational beliefs cannot really be formed in the social world which depends on rational beliefs.

But while these are all formidable questions of grave methodological significance for the relevance of economic theory, we are not going to focus on them. Instead, we will look at a missing element in the above story. In the case of the traffic forecast, what is not taken into consideration is the fact that some agents may choose not to use their car and use instead, public transport in the form of train or trams. This, in turn, may be another reason why the expectation of the traffic jam will not materialise even if some people who expected others to avoid the area will now drive through, but the main point here is that it would mean that those who chose not to use their car at all, have, so to speak, opted out from the framework which was under consideration: the use of roads. Of course, one could say that it is possible to add these considerations into the model, but this is not the point. The issue here is that the relevance of the

[1] Given…. the natural proclivity of every science to become more limited in scope as it grows more conscious of its premises (12).

model (in terms of making correct predictions when based on other people's choices) depends on agents' ability to decide whether to take part in the story as well as on the kind of reasoning which they will follow. It may indeed be the case that they cannot opt-out but in such a case, the social significance of the equilibrium (or the materialised prediction) may well be different than the one normally assumed.

Put differently, for the expectation of a traffic jam to be formed, we must assume that all agents will use cars regardless and that their reasoning is the same. Hence, if half the people, upon hearing about the traffic jam, become concerned about the environmental effect of their driving, will choose not to use their car, then even if the rest think that the others will choose alternative routes and therefore use the one expected to be congested and that route would stay free and the expectation will not materialise.

But there is also the question of what one would expect to find (c-expectation) when the expectations materialise. In our example, to be stuck in the traffic jam would mean that the prediction of the model was correct. Presumably, the benefits accrued when stuck in a jam are some form of disutility associated with the possibility of missing an appointment and the annoyance of being stuck in a car. But if one discovers that being stuck in a traffic jam is an opportunity to flirt or learn languages—namely, the benefits of the system are not necessarily the one anticipated—the agents may change their behaviour and upon hearing about an expected traffic jam will drive straight into it. This, of course, will make the expectation materialise but the benefits from it will be very different indeed from the one initially anticipated. Such behaviour would be rational but for entirely different reasons from the one which made avoiding the jam rational.

What we are trying to emphasise here is that while the dependency of the outcome (the v-expectation)—in the traffic jam case—is clearly dependent on the expectation about the behaviour of others (p-expectation), the effects of the proposed benefits (or cost) (c-expectation) have not been properly explored. Namely, not only is the outcome conditioned on the distribution of behavioural patterns but also the consequences depend on those behaviours which, in turn, may influence the behaviour and the probability of the outcome to materialise.

This story of the traffic jam is clearly the mirror image of the problem which expectations face in economic analysis. However, in the standard economic take on a simple competitive market, there are no explicit expectations. Agents make the best decision for themselves based on their own circumstances (endowments) and known prices. But economics does not tell us much about how these prices come about. It nevertheless promises that if people behaved in this way competitive equilibrium prices will emerge and with them, the benefits embedded in the distribution of surpluses among producers and consumers (c-expectation) would be accrued. So, there are implicit expectations here: if people were to expect other people to behave in the same manner (p-expectation), the equilibrium outcome would emerge (v-expectation) and the benefits will materialise.

But what if people did not believe that the equilibrium price will ever emerge, and therefore the anticipated benefits will never be accrued? Would they continue behaving like competitive rational utility maximisers or will they seek alternative

forms of behaviour? In part, the answer to the question depends on how one understands the economic system. If one believes, somewhat deterministically, that people are naturally rational utility maximiser and that such a behaviour is independent on whether it leads to beneficial outcomes when the economy is organised in a competitive framework, then there is no problem. However, given that the competitive set-up ostensibly depicts a state of natural liberty, such a view would not appear sound neither in the short run (for simple logical reasons) nor in the long run (for evolutionary reasons). In other words, in any adopted form of behaviour, there is an implicit expectation that it would yield that which the individual is seeking to achieve.[2]

Assuming for a moment that people are right in their expectation that competitive equilibrium prices would indeed yield a desired outcome, they are, in principle, ready to stick with the behaviour of rational utility maximisation and expect others to do the same. In such a case, the main problem is whether there exists a mechanism that can support such an expectation given the way in which equilibrium prices are formed.

In general, the economists would have us all believe that competitive equilibrium prices are something worth believing in on the basis of what may be generally termed, stability conditions. To wit, as prices below equilibrium suggest excess demand and those above it, excess supply, it appears evident that in the former case prices would rise towards equilibrium and in the latter, they would fall towards it. However, as can be easily demonstrated, there is nothing further from the truth. A simple cobweb model can show that with more structure on such dynamics (including expectations) this is far from obvious. In other words, the stability of competitive equilibria is far from being a convincing argument. Fisher (2011) laments the state of theory in this area and draws our attention that without a proper proof, the trust we have in the institutions of competition—the implicit expectations which we have in the beneficial outcome of competitive interactions—must be curtailed.

But things are far more complex and less promising when we acknowledge the difference between stability and price formation. While there is clearly a connection between them, stability assumes the existence of an equilibrium price, whereas price formation has to explain whether there exists at all a process which can lead to these prices. Walras himself suggested the tatonnement as the mechanism of equilibrium price creation but even if we overlook the absurdity of his own mechanism and take a more general look (as did Scarf 1960) we will find that it is far from obvious that there is a meaningful process that leads to such prices.

Altogether, therefore, economics is making three crucial implicit assumptions about what agents expect in a competitive system: they expect the equilibrium price to materialise if they, and all others, would behave in a particular rational manner;

[2]Some of these difficulties have been captured in the problem of the Prisoners' dilemma. There, the widely accepted rational form of behaviour seems to produce an outcome which is inconsistent with the expectations of the participants. The experimental and behaviour literature in economics suggested that people would behave differently and achieve the expected outcome. From our perspective, it is as if they would change their behaviour when they realise that the equilibrium of their initial behaviour does not yield the expected benefits.

they also expect this outcome to yield the benefits embedded in these prices. Hence, all three aspects of expectations are rolled into one story where expectations are not even explicit.

What we find interesting is that in classical economics, notably, Adam Smith, there is a similar—though more interesting—interplay between these types of expectations with surprisingly different conclusions. There, expectations about the consequences of competitive interactions lead to an outcome in which their materialisation does not really produce the anticipated benefits which triggered the competitive interaction in the first place. In terms of the traffic jam metaphor (hence with reverse signs), it is as if people tried to avoid the traffic jam but it nevertheless materialised and when it did, they discover that being stuck in a traffic jam is actually a good thing. Would they then change their behaviour?

In a nutshell, Smith describes a world in which people seek social approbation (moral approval). Through a deception by nature they are led to conflate the ease with which individuals empathise with riches with moral approbation. This, in turn, leads them to pursuit the augmentation of material wellbeing—which is nature's own purpose—in their pursuit of social approbation. However, while the outcome of such a system is indeed an augmentation of material wealth, its natural (market) distribution suggests that very few people achieve through it the objective of being socially approved or recognised. If everyone becomes equally rich, it would no longer be wealth which distinguishes a person and make other socially approve of him or her. If, on the other hand, only a few acquire wealth and through it, command the admiration of others, the poor, will fail to achieve that which they expected when entering the game. Will they change their behaviour? Well, not really because there are other elements in Smith's theory which explain why many of them will tend to approve of the system for the wrong reasons. They are at awe with its 'beauty' even though it fails to deliver that which they sought to achieve.[3]

In some ways, there is something in Smith's analysis, which may resemble the idea of rational expectation and which is, in many respects, Kantian in nature: it is the notion of the impartial spectator. This, in the end, leads agent to choose rules of behaviour which could be explained as the outcome of an analysis in which people examine the world had all people chosen to act in a particular manner. It is in this manner that they develop principles of behaviour which are, in essence, moral. Some may argue that there is a clear difference between the way in which people choose to behave morally and the way in which they make rational choices. Indeed, there is a difference between expecting something that ***should*** happen and expecting something that ***would*** happen. However, such a presumption is based on the understanding that people's behaviour can easily be segregated into the moral and the rational based on whether or not it involves an explicit view of the other. Put differently, it is the presumption that self-interested actions are exclusively about the 'is' while moral questions are only about the 'ought to'.

[3]One cannot but see a parallel between the functions of the deceptive power of wealth in Adam Smith and the social function of the American Dream.

Thus, when agents have to choose the profit maximising level of output, they need to know what the price in the market would be and not what it ought to be. When they form an opinion based on what others would do (like our motorists in the traffic forecast example), they will be able to choose the level of output which would indeed maximise their profits. However, their presumption is that the price that will be—the equilibrium price—will inevitably be the one in which all people were driven by the same considerations: profit maximisation. If, for instance, there were some forms of cooperatives or other forms of organisations among those producing the good, their objectives may not be the same as the purely profit maximising agents and this would mean that if they extrapolate from themselves towards the other, this will lead to a very different outcome.

In other words, only in a world where a complete segregation exists between morality and self-interested actions, can the 'is' and 'ought to' be so unrelated. This means that if a self-interested profit maximising agent extrapolates his own considerations onto everyone else, and if they all have the same information, rational choice will lead to an outcome in which they all get that which they originally intended to achieve. But if there is a moral dimension behind human behaviour—even self-interested one—it is not evident that the outcome will necessarily merge the 'is' with the 'ought to'. In such a case, anticipation will be frustrated and this, in principle, should be a reason for a change in behaviour.

The case of Adam Smith is exactly a case like this. In Smith, society precedes economic organisation and the division of labour is not a rational choice. The reason people engage in it in the first place is to seek social approbation. Even when they realise the potential benefit of extended division of labour, they have not lost the search of social approbation which is at the heart of human motivation in Smith's analysis.

As social approbation is, by definition, associated with morality, it subjugates the economic realm to the social/moral one. In such a complex world, the dual purpose of the economic system is both to provide individuals with life's necessity and, most importantly, with the approbation they seek. However, as there exists a tension between that which Smith calls nature (the physical world to which he attributes an almost Darwinian dynamic) and what he calls human nature, the realm of economics is also a place where the struggle between these opposing forces takes place. It is evident that the objectives of nature are the 'is' of this world but the objectives of human nature are always that which 'ought to be'.

Therefore, at the heart of Smith's analysis lies the deception by nature (the physical one) which leads to a behaviour which appears to the agents as consistent with their search for social approbation, but which produces results consistent with nature's objective but not necessarily with human nature. Namely, agents get into the system to achieve one thing and their rational expectation leads them to an outcome where the results are not those which they anticipated but are consistent with the objectives of the nature of things that originated the deception.

2 Expectational Fallacies

When we contemplate expectations, we are, in fact, thinking about two separate levels of analysis. The one, most commonly referred to by economists is a dynamic concept and it pre-supposes that in principle, all three aspects of expectations work in tandem as far as the basic model is concerned. Thus, if people behave in a competitive rational manner and expect others to do the same (p-expectations), the outcome will be the competitive equilibrium price (v-expectation) and it would be, in the end, beneficial (c-expectation). The implication of this is simply that competitive institutions, in general, are worth pursuing. In such a well-organised world, expectations that concern economists are concentrated on the movement of the system rather than its justification. Yes, there may be some transitional anomalies but in the end, it is all about the shift from one good outcome (equilibrium) to another. In this category, we will find all the expectations which come with dynamic expressions of the model like, for instance, those dealing with labour contracts, real interest rates or real exchange rates in a world where there is inflation. These kinds of expectations—which we may title 'dynamic adjustments expectations'—are not the focus of our attention here. Instead, we would like to focus on those expectations which do not necessarily have a dynamic expression but lie behind the presumption about the validity of that model for the shifts of which we created the notion of dynamic adjustment expectations. In other words, we would like to focus not on those expectations that facilitate the movement across equilibria but the expectations which people may implicitly have about those equilibria.

This fundamental level of expectation, which may explain why, to begin with, people adhere to a particular form of behaviour, is in itself divided into three different components. The first is the question of whether sticking to a particular pattern of behaviour and expecting others to do the same will yield and outcome which is desirable. The second is the question of whether that pattern of behaviour which yields the material benefit agents crave is also consistent with the norms of their morality. The third question is whether or not there exists any mechanism which can form these promised values. In this section, we will take them in turn.

2.1 Expectations Regarding the Beneficial Nature of Competitive Institutions

Economists would like to think that their competitive conception of the world is depicting that which is indeed natural. Evolutionary speaking—even in the very short term—it is unlikely that people would adopt a pattern of behaviour if they knew all along that it does not serve their interests. These interests may be comprised of both material and non-material benefits where the latter refers to the social context, and morality, of such behaviour and its influence on one's social standing.

As far as material wellbeing is concerned, economic theory suggests that given the natural endowments which people have, they will all be able to improve their material situation if they act rationally (in the sense of the rational utility maximiser) and interact competitively. This means, among other things, that they should all specialise and trade both as individuals and as societies. Allowing things to develop naturally (i.e. without intervention) will not only create material gains but will benefit everyone. One can easily imagine that if this proposition were proven to be wrong, people are not very likely to choose this form of behaviour or, if it were natural to them to behave in this manner, they are not likely to tolerate the non-interfering institutional set-up. Thus, in a competitive set-up, individuals have an implicit expectation that if they specialise and trade competitively, they would gain from such behaviour provided that others would do the same.

Let us examine this basic proposition in some details. Suppose that there are only two goods in the world, x and y and that one needs only labour to produce them. Suppose too that there are only two individuals (or groups of two types of individuals) whom nature endowed in such a manner that by themselves they face the following sets of bundles which they can acquire:

The individual on the left (in Fig. 1) can produce either 6 units of y or 3 units of x or any other linear combination of the two. The individual on the right can produce either 3 units of y or 6 units of x or any linear combination of the two. Given their preferences, without interacting with the other group, each individual, or economy, would choose a certain combination of the two goods. Assuming that the allocation is done through competitive markets, we know that the equilibrium point will be on the production–possibility frontier (where they have extracted the most of what nature has given to them) and the competitive price will reflect the real cost of producing the goods.

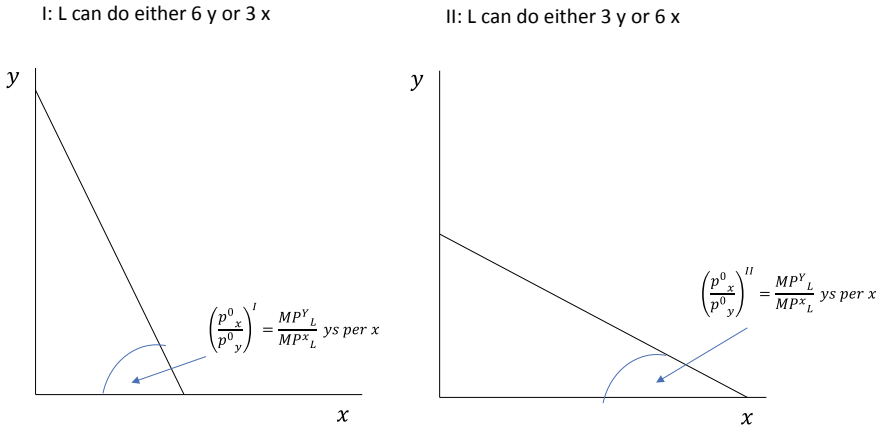

Fig. 1 Simple example: general equilibrium—complete markets

Now, when they encounter the other group of people, rationality and efficiency dictate that each group should specialise in that in which it has a comparative advantage and trade, again, competitively, with the others. However, to know what it is in which we have comparative advantage we must know the real opportunity cost of each good or, technically, we need to know the slope of the production–possibility frontier. However, the frontier is not an observable as it does not really exist. There must therefore be some other means for us to be able to know what the slope of the frontier is at the point where we are currently producing. Here, one of the side benefits of competition comes into operation by promising us that if markets were perfectly competitive, market prices will reveal the slope of the frontier. In other words, when markets are perfectly competitive, we pay for goods exactly what it cost to produce them.

As a result, the two individuals, or the two economies made up of one type of individual each, would specialise and trade according to the realisation—through observing competitive market prices—that economy I has a comparative advantage in y while economy II has a comparative advantage in x. The outcome is given in Fig. 2.

If we think for a moment about two individuals in Fig. 2, then it is clear that each one of them was expecting the equilibrium value to be the agreed price (i.e. to lie between their respective reservation prices); each one of them expected the other to behave in the same manner and each one of them expected to benefit from these actions. So here is the simplest exposition of the idea that certain type of rational behaviour and competitive interactions—and hence, competitive institutions—lead to fulfilled beneficial expectations.

In such a world, when people are convinced that competitive interaction of rational utility maximiser will lead to a beneficial outcome, any dynamics which deals with

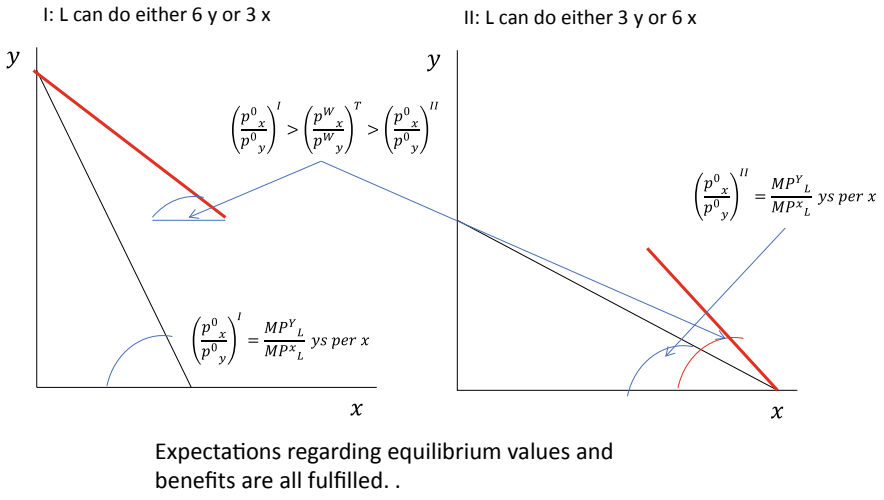

Fig. 2 Simple example: general equilibrium—complete markets

the move from one beneficial equilibrium to another could easily be based on the presumption that agents will have no incentive to change their behaviour.

However, given the pitfalls of the theory we can easily see how fragile is the promise embedded in the competitive paradigm. Suppose for a moment that both economies (one comprised of individuals type I and the other of individuals type II) are perfectly competitive but that there are, as is always the case, missing markets.

Recall that the production–possibility frontiers are not observable and that agents rely on competitive prices to reveal their comparative advantage. As we know, when there are missing markets, competitive prices no longer reflect the slope of the frontier. To make this easy, suppose that the missing markets we are discussing are those embedded in the problem of externalities. Hence, for argument sake, in Fig. 3, we assume that in economy I the production of x generates negative externalities for the production of y and vice versa in economy II. As a result, competitive equilibrium prices will be denoted by the heavy black line in each economy. If we start at point B—which is where the individuals (or the economies) are before they specialise and trade)—we can see that they have made the best with what nature has given them. Will they benefit from following the behaviour edicts of the competitive world?

The answer is clearly negative as can be seen in Fig. 4. Not knowing what the real frontier looks like, they will wrongly conclude from the competitive market prices—which they believe reflect the slope of the frontier—that individual I has a comparative advantage in x while individual II has a comparative advantage in y. When they specialise and trade, the new equilibrium price may be the exact same price they would have reached in the previous case (hence, the same v-expectation as before) but the benefits will not be accrued as they would both become materially worse off. There is an interesting question regarding whether or not they will become aware of the fact that they are worse off as this whole process is notional and in

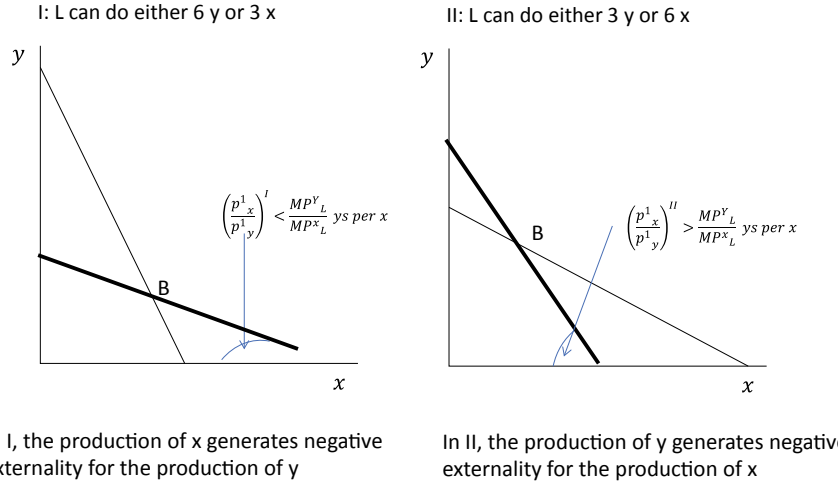

Fig. 3 Simple example: general equilibrium with missing markets

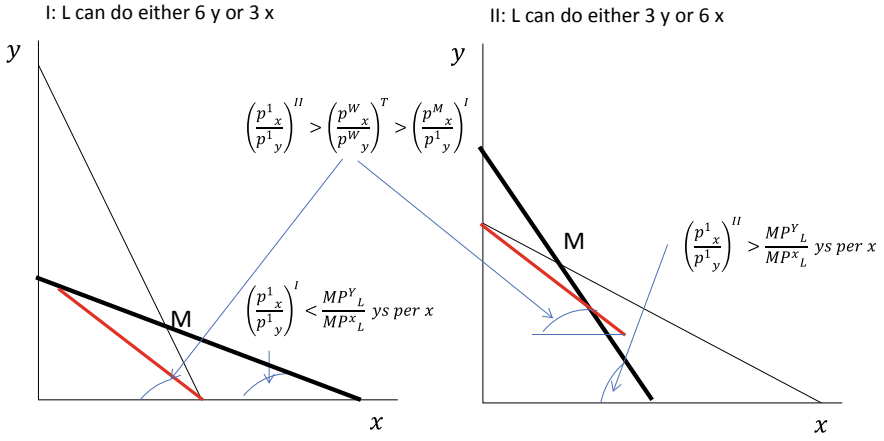

Fig. 4 Simple example: general equilibrium with missing markets

reality, people just specialise and trade believing that in so doing, everyone will gain from trade. But if they are aware that rational competitive behaviour (actions and interactions) would lead to the expected value but that value would not yield the expected benefits, would they continue to behave in the same manner?

Given that the problem of missing market is endemic and beyond repair, it raises a serious question as to why do people keep behaving in the same manner and expect others to do the same? After all, the expected outcome will not have the benefits they all expect to find in it. It may well be the case that rational behaviour and competitive interactions would produce plenty in terms of material goods but whether or not all agents gain from it is another question altogether. This is somewhat similar to what Adam Smith has referred to as the ***deception by nature*** (we will revisit this in Sect. 3) but modern economics remained oblivious to it.

In terms of the cunning of reason terminology, in the presence of missing markets, the system of competitive interactions—the combined expectations (v-expectation, p-expectation and c-expectations)—will yield an outcome which is inherently inefficient. This, in terms of the traffic jam metaphor, would mean that in spite of the warning about the traffic jam, we ended up in it and we found it to be a good thing (remember that the metaphor is a negative image of the economic one).

2.2 Moral Expectations

It is, of course, evident that if people's expectations regarding their benefits are systematically frustrated, they are not very likely to stick with the pattern of behaviour that yields such outcomes. But could people's expectations with regard to benefits be fulfilled and yet, they may not stick to the pattern of behaviour which generated it?

Mandeville (1714) famously posed a dilemma which is, in principle, about this very question. In it, he implicitly asks the following question: if that which produces plenty of material wellbeing is morally repugnant, should we condemn the pursuit of it, or morally vindicate that which produces it? In other words, it seemed inconceivable in his eyes that a system that produces plenty and is based on morally repugnant principles can be sustained without a resolution of this tension. Now, whether this resolution leads to vindicating what may appear as repugnant behaviour or the scraping of the system is a different, and complicated, question.

Following the idea of praxeology, we accept that individuals act purposefully. They seek, or expect, to achieve something through their actions. A natural, or spontaneous, order exists when the outcome of all human interactions is such that all have achieved that which they had ***reason to expect***. In other words, a situation where the actions of all individuals are compatible is, in essence, a natural order and it is also what is meant by equilibrium in economics.

The dilemma which Mandeville has raised is of a fundamental nature as it is about the relationship between the 'is' and the 'ought to be' in society in general and in human action in particular. Even if human actions are compatible in the sense that all achieve that which they had reason to expect, is it possible that the actions which produced this compatible outcome are morally unacceptable? If fraud, luxury and pride—as cited by Mandeville—are the necessary foundation of that human action which brings us opulence, and if they can be found to be compatible (with one another), would this constitute a natural order?

In some positive sense, the answer should be in the affirmative. If such motivation produces an equilibrium which is efficient, surely it is a form of a natural order. However, if we accept that humans are social beings which means that moral values (which mean their relationship with the others) are part of their human life in as much as the pursuit of their own private affairs, then such an order will also depend on people approving of such behaviour. Naturally, a form of utilitarianism (i.e. consequentialist theory) may, in such a case, suggest that if fraud, luxury and pride are what it takes to produce plenty of material wellbeing then these are morally approved foundations of human actions.

But if morality is not utilitarian in nature and people deem these foundations of human behaviour as morally repugnant, would such an order be sustainable? The answer should be negative. If indeed there is such a view of the behaviour which makes up the compatible system, there will be expectations that people would behave differently and in so doing, alter the nature of the order that would emerge. One must hasten to say that the way in which morality is formed is also a natural process and therefore, in the end, one should expect a convergence of the two. Namely, over time, that which is deemed to be morally approved should become the foundation of behaviour which yields compatible outcomes. In such a world, the expectations of the acting agents would be fulfilled both in terms of their immediate effects as well as in terms of their moral acceptability.

But in modern economics, this is not the case. Individuals solve their problems without much reference to the social or moral significance of their actions. Their expectations, so it seems, are merely focused on obtaining the bundle they prefer

most, irrespective of what it may mean socially or morally. When a coordinated outcome emerges in the form of general equilibrium which is allocative efficient, they can no longer do better for themselves and in a sense, their expectations—those which they have *reason to expect*—are fulfilled. However, while that which they have reason to expect may depend on their initial endowment, their social or moral expectations may not be limited in the same manner. They may, for instance, expect a certain sense of justice from the system within which they expect the solution of their economic problem.

In principle, the way in which individual behaviour develops should not be independent of the way in which their moral values evolve. After all, the moment they become interdependent, the others become part of their daily lives and one way or another, people would form an opinion about the nature of these relationships and the norms which should govern them. This, however, is a complex subject about which modern economics has little to say though classical economics in the shape of Smith's system has plenty with which to enlighten us. Therefore, at this stage, we shall assume that people have a certain natural conception of what constitutes justice in the form of desert. This means that people believe that individuals should receive returns reflecting their relative contributions. Hence, while the expectation is to be able to acquire that which they have reason to expect, this comes with an added anticipation that the way in which these expectations are met is consistent with one's views on what is justice (here, in the sense of fulfilling the principle of proportional remuneration).

Consider for a moment the following situation. A certain quantity of a commodity (say $\bar{y}$) is produced by the labour of two individuals who are equally productive. The individuals do not choose the quantity to be produced (it is a given) but they choose how to divide their labour and how to divide the output between them. To some extent, this captures the economic environment in which we offer our contribution to the production process and expect a share in that which has been produced.

To produce $\bar{y}$ there is a need for $\overline{L}$; ($\bar{y} = f(\overline{L})$). Let θ denotes the share of agent 1 in the labour required (such that $l_1 = \theta \cdot \overline{L}$), hence $1-\theta$ is the share of individual 2 in the labour. Also, individual 1 will get the share of x of the total output while individual 2 will get the share of $1 - x$. Each individual has a utility from consumption and disutility from labour which can be defined in this case over the parameters (θ, x); hence, $u_i = U(\theta, x)$. Clearly, individual 1 prefers a greater x and a smaller θ while individual 2 prefers the exact opposite. Assuming the right properties to these utility functions, it is easy to see that the following economic problem emerges:

The indifference curves of both agents in Fig. 5 are positively sloped as an increase/decrease in the share of good increases their utility is compensated by an increase/decrease in their share of labour. The respective indifference curves are represented by the red and blue lines below. Agent 1's utility is rising upwards to the left while agent 2's utility is rising downwards to the left on these parameters.

The nature of a competitive solution is that it is Pareto efficient. We can clearly see (in Fig. 6) that there exists a combination of shares (θ^*, x^*) which is Pareto efficient:

At point A in Fig. 6, we have the allocation of contributions and distribution set at $\left(\theta^A, x^A\right)$ which is efficient in the sense that we cannot increase the utility of one

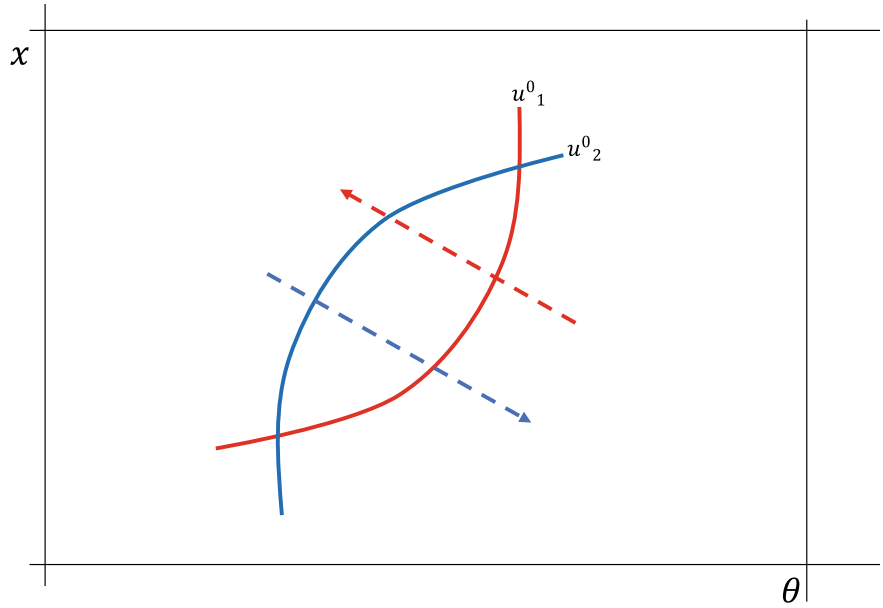

Fig. 5 Distributional expectations

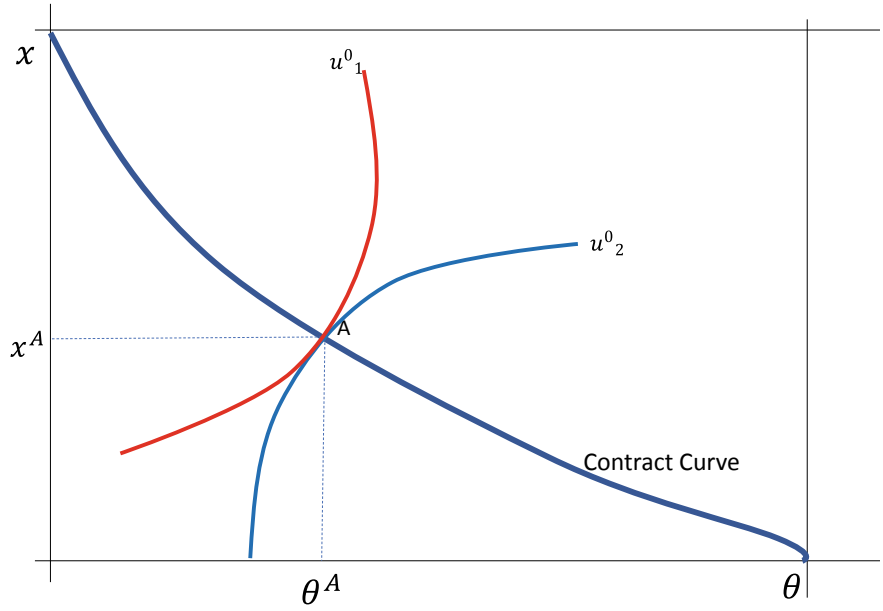

Fig. 6 Distributional expectations—efficiency

agent without decreasing that of the other. Naturally, there are many such allocations and the point to observe is that they are arranged along the contract curve which is downward sloping from left to right. This set of efficient allocations points at all the candidates for allocations that will emerge within the competitive decentralised system.

However, if we ask ourselves how would we arrange the values of (θ, x) if we thought about them in an ethical manner? Clearly, there is a moral connection between them as one of them represents the distribution of efforts, or contribution, and the other, the distribution of reward. In what seems to be a natural way of thinking about this issue it seems obvious that for individuals who only differ in their effort (not ability or any other characteristic), the higher θ is—the more is the output due to the efforts of individual 1 and less to the effort of individual 2—the higher should be the reward to individual 1. Namely, if we were to draw the line connecting these two parameters on moral grounds as we do in Fig. 7, we will find it orthogonal to the one that markets would produce:

Naturally, there will always be one allocation (like A in Fig. 7) in which both efficiency and morality are satisfied but whether or not A emerges as the solution to market interactions depends, in this case, on their taste parameters. One could say, of course, that if their taste parameters lead them to an allocation on the contract curve which does not coincide with the moral contract curve the latter cannot be a meaningful construct as people prefers to be at a different point. However, the

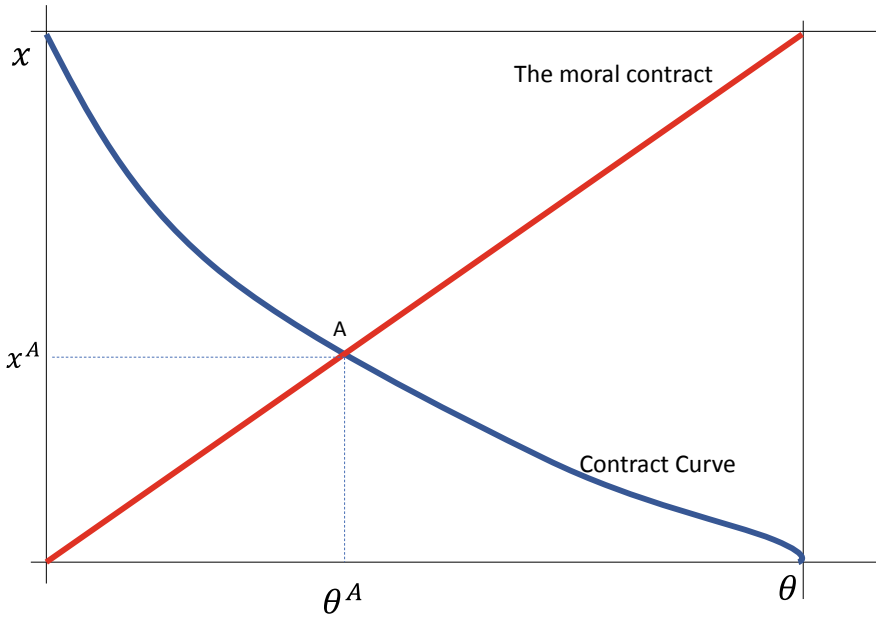

Fig. 7 Ethics and efficiency

presumption here would be that morality and economic considerations are derived from the same principle.

That this is not the case, in general, can be easily demonstrated though this may require a longer discussion which we will not conduct here. For one, economic considerations are consequentialist in nature and moral principles are not only about consequences. For instance, in our above story, the moral principle was about the relationship between effort contribution and distribution of rewards. So it is about causal relationship between two parameters that are determined simultaneously. The two individuals have preferences about the parameters and make their decision about them in terms of what they do to them and not, as morally dictate, in terms of what they mean to the other. The moral contract curve is based on the principle of proportionality irrespective of preferences because it guaranteed that everyone is treated in the same manner. Preferences-based choice of these parameters is only associated with how they affect one's own preferences irrespective of what it means to the other.

Therefore, while someone may be less averse to hard word and less prone to the pleasures of consumption than the other, it may not be morally a good enough reason for him, or her, to work harder and get less of that which they produced. Thus, one may indeed conclude that the association of competitive outcomes with a moral principle like that of proportionality is entirely accidental. Hence, if part of people's expectations from the benefits of the outcome is that it should respect some basic principle of morality, one must recognise through the orthogonality of these allocations that the material benefits of the markets are bought at the cost of morality. It seems to suggest that the problem posed by Mandeville has not really been resolved. Is this something which is sustainable?

2.3 A Note on the Dynamic Aspect of Expectations

So far, we have explored two difficulties with the implicit expectations which lie behind the competitive paradigm. Both are associated with what we called the consequential expectations (or c-expectations). Namely, even if people have a correct expectation about the value of the economic variable (like the price) and even if they have correct expectations about the behaviour of others, the consequences of adhering to such behaviour may not be the one which people anticipate. In the first case, we showed how the expectation of a specific competitive value may not yield the expected material consequences due to problem like missing markets. In the second case, we showed that even if the expected material benefit materialises, the social, or moral, aspect of these benefits may be frustrated. In both cases, one cannot avoid the question: why would people continue to behave in the manner which generates these outcomes and why would they expect others to behave in a similar way?

But one does not have to resort to complex argument about missing markets or the synergy between the formation of moral values and economic expectations in order to question whether the cunning of reason works to support the competitive paradigm. In the Walrasian model, the promise is that if all people behaved rationally

(in the sense of the rational utility maximiser), expected others to behave in the same manner, the equilibrium values they anticipate will materialise and bring with them the benefits they expect encapsulated in the concept of efficiency (both productive and allocative). In general, this raises two immediate questions: firstly, is there a process of price formation which leads to these blessed Walrasian prices the values of which people expect; and secondly, is the promise true?

We already answered the second one in Sect. 2.1 above. But the first one is perhaps even more of a serious problem. In the end, if there is no process which yield the expected prices and the only way for them to materialise is an immediate jump into them (rational expectations), then the whole story changes its meaning. Walras himself was aware of the difficulty associated with how equilibrium prices emerge, and his own depiction of the process (his tatonnement) depended on transactions taking place only when equilibrium prices are reached. As Scarf (1960) demonstrated, there is a problem for competitive prices—in a generalised process, where change in price is a function of excess demand—to reach Walrasian equilibrium and that it is highly likely that prices will oscillate forever without reaching Walrasian prices.[4] To use rational expectation to try to circumvent these difficulties is therefore highly contrived and extremely unlikely if one considers the cognitive abilities and information, required to achieve such a feat. While Muth (1961) may have proposed something which could, in principle, solve the problem of a single market, to employ the mechanism on a large scale of markets appears improbable. Fisher (2011) is quite adamant that we have not resolved the problems of stability in general equilibrium and that such a problem cannot remain unsolved. While it is true that stability and price formations are not the same thing, they do rely on similar processes. It is highly unlikely that in the absence of stability, there may exist a process of price formation that would lead to Walrasian equilibrium prices.

To demonstrate this point, we can look at the simple example which can be found in the case of the cobweb story. Here, we are looking at a single market and ask the question whether there is some reasonable process which would lead us to the competitive equilibrium prices with the benefits which they promise. From the beginning, we accept, in principle, that there exists a price for which quantity demand would be equal to quantity supplied. We also accept that at any price above it there will be excess supply and at any price below it, excess demand. Adding the general widely accepted view that excess demand will lead to an increase in price while excess supply will lead to a fall in it, we seem to be certain that in the end, the price will be pushed towards its equilibrium level.

However, for this to be true, it requires a non-trivial leap of faith. Recall that individuals do not necessarily know all the parameters which determine the position of the demand and supply schedules, and therefore, have no clear idea what the

[4] A good summary of some of these attempts can be found in Manea (2015). One must hasten to say that with some effort, it is always possible to devise some strange mechanism that will achieve the desired result but, like Walras's own auctioneer, the stories remain farfetched and impossible to be universally applicable. Osborne and Rubinstein (1990) capture an attempt at making the process of price formation more realistic through the mechanisms of sequential bargaining but they too, fail to produce the Walrasian prices with their embedded promise of productive and allocative efficiency.

equilibrium price is where they expect their benefits to materialise in the form of a consumer and producer surpluses. They do ***believe*** that if the market is competitive, equilibrium price will have these properties.

What agents observe, as a matter of fact, is the current price in the market. At the end of each trading day, they may believe these prices to be equilibrium prices in the tautological sense according to which the quantity sold always equals the quantity bought. But this does not mean that they are indeed equilibrium prices in the sense of the coincidence of all rational plans.

The cobweb model provides a simple description of the way in which the dynamics behind the static model of demand and supply may be manifested. To give more depth to the story, we distinguish between the way consumer and producers behave. Consumer observes the current price and decide how much to buy but producers, need to prepare for the market in advance so they need to have an expectation about the price in the market. Observing the price at the closing of the previous trading day, they may conclude that this will also be the price tomorrow as they have no information about the general structures of demand and supply. This is, of course, a form of adaptive expectations.

Thus, our model of demand and supply becomes something along the following lines. Suppose that the demand for x is given by:

$$x_t^d = \alpha - \beta \cdot p_t$$

where α and β are parameters, and p_t the current price.

Producers' supply is of the following nature:

$$x_t^S = \gamma + \delta \cdot E_{t-1}(p_t)$$

where γ and δ are the parameters but the price according to which they prepare for the day in the market depends on what they expect the price in the market to be. The simplest and obvious form of such expectation would be the one where they expect the price tomorrow to be the same as the closing price today:

$$E_{t-1}(p_t) = p_{t-1}$$

This means that in equilibrium:

$$x_t^d = \alpha - \beta \cdot p_t = \gamma + \delta \cdot p_{t-1} = x_t^S$$

Which means that we have a dynamic story that explains the evolution of the current price:

$$p_t = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} \cdot p_{t-1}$$

$$p_t = \lambda - \theta \cdot p_{t-1}$$

where $\lambda = \frac{\alpha-\gamma}{\beta}$ and $\theta = \frac{\delta}{\beta}$.

Will this price dynamic lead to a Walrasian equilibrium? Well, the first question is whether or not this dynamic will lead to a convergence of the prices to one price which will become the equilibrium price which is not changing daily. The answer to this depends on the nature of the above difference equation. For prices to converge to an equilibrium price, we would need to examine the condition for a steady state:

$$\bar{p} = \lambda - \theta \cdot \bar{p}$$

$$\bar{p} = \left(\frac{\lambda}{1+\theta}\right)$$

These conditions are given by:

$$p_t = \begin{cases} (-\theta)^t \cdot p_0 + \lambda \cdot \left(\frac{1+\theta^t}{1+\theta}\right) & \text{if } -\theta \neq 1 \\ p_0 + \lambda \cdot t & \text{if } -\theta = 1 \end{cases}$$

Or, in simple terms, there will be a convergence to equilibrium only if: $|-\theta| < 1$, and $\delta < \beta$. This means that only when the slopes of the demand schedule and the supply schedule relate to each other in this manner, will prices converge to an equilibrium price which would indeed be the expected competitive (or Walrasian) price.

It is evidently unlikely that in all markets such a relation will be kept, and therefore Walrasian prices are unlikely to emerge. However, what is most striking about it is the presumption that people will continue to behave in the same manner believing that the benefits of competition are accrued to them. Suppose for a moment that there is no convergence and prices keep oscillating around the equilibrium price never to reach it.

Clearly, the benefits which are accrued to producers and consumers at points *A*, *B*, *C* and *D* in Fig. 8 are very different indeed. It means that at the end of a trading day, someone would find their expectations concerning the available consumer or producer surplus frustrated. Would they then continue to follow the same pattern of rational utility maximisation or would they look for alternative patterns of behaviour?

As may be suggested, this is reminiscent of the prisoner's dilemma problem. There, the rational behaviour of the agents and the expectation that the other agent would behave in a similar rational manner led to an outcome which was not the best for the players. The way economists tried to deal with this was by saying that individuals would not change the principle guiding their behaviour (rational utility maximisation) but would, in fact, change their behaviour (their choice of strategy). In part, this could be done through changes to the structure of the game (infinitely repeated games with certain time preferences and punishment strategies), or through the change in the scope of agents' considerations (either becoming socially minded or learning through evolutionary games that cooperation may be a better strategy). But is

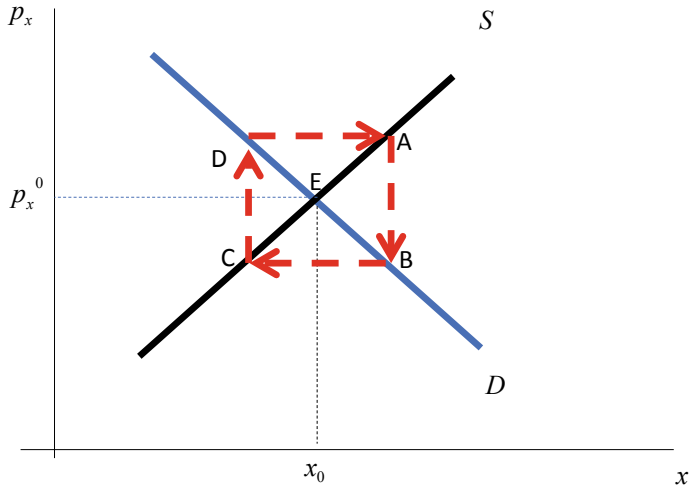

Fig. 8 No convergence

it possible that agents would change the principle which guides their decision-making in the face of such failures? Surely the answer must be in the affirmative.

3 Expectations and Anticipations in Smith's Narrative

Between Mandeville and the modern narrative, which as we suggested, ends up reiterating his dilemma, there is the classical school and, in particular, the work of Adam Smith. In the eyes of some, Adam Smith seemed to have resolved Mandeville's dilemma by positing that it is not the morally repugnant characteristics of human nature that generate material wellbeing but rather a more innocuous one: self-interest, which should not be conflated with selfishness.[5] But even if we accept that the self-interest is not the same as selfishness, the difficulty to which Mandeville alludes, is clearly present in Smith's work and instead of being glossed over, as is the case in modern economics, it is explained and expanded in his work. In other words, not only was Smith aware of the fact that people's expectations with regard to the consequences of their self-interested behaviour would be frustrated but he also offers an explanation as to why in spite of this, self-interest persisted to be a powerful—though failed—pattern of behaviour.

Unlike modern economics, expectations are explicit in every step of the way in his analysis of economics mainly because for Smith, as for most classical economists, economics was part of a social theory with a much broader agenda than the mere creation of plenty. This means that economic motives are never too far from social or

[5]The fallacy behind this reading of Smith is expanded on in Witztum (1998).

moral sensitivities and that the expectations which lie behind the economic project are really social. As a result, the success or failure of the economic system in the eyes of its elements is measured by the way it serves their social objectives. This opens the door to the possibility of successful generation of material wellbeing with a complete failure to deliver on the social purpose of economic interactions. The question that remains is, why would this not lead to a change in the way people behave. The answer, as will be explained below, is that Smith, like J. S. Mill, did not consider the commercial stage of society as the final one. While Smith wrote at the beginning of the era, he was acutely aware of social progress but did not feel the need to elaborate on the next stage of society which would inevitably entail a possible change in the way people behave and interact. To some extent, he laid the foundation for this potential change in human behaviour but did not consider its implications in full. In the case of J. S. Mill, competitive interaction was necessary to rid society of debilitating customs, but as people progress economically, they also develop cognitively and as they do, Mill argued, they will become increasingly more cooperative in their interaction and will also lose interest in the endless pursuit of material accumulation.[6]

The Smithian economics story, very much like the modern one, begins with the observation—as opposed to rational construction—of the phenomena of specialisation and trade. People do not specialise and trade because it is the rational thing to do in order to achieve material wellbeing. Instead, it is a result of natural expression of human sociality: the search for social approbation. "We cannot imagine" writes Smith about the division of labour "[for it] to be an effect of human prudence" (*Lectures on Jurisprudence (LJ)* 492). Nor is it, according to Smith, because of the differences in people's abilities: "This disposition to barter" he claims, "is by no means founded upon different genius and talents" (LJ 492). Instead, "[t]he real foundation of it is that principle to persuade" (ibid). But what does it mean to wish to persuade? According to Smith, persuading the others means to make them think or feel like you; it is a creation of harmony—proximity—between people which is either cognitive or emotional. As harmony of sentiments, in Smith's analysis, is the key for social and moral approbation, the drive behind the division of labour—and the subsequent analysis of its success or failure—is really the search for such approbation.

At an early stage of society, the hunter-gatherer stage, if one finds it easy to forge an arrow from wood, he can easily make plenty of them above that which he needs for his own subsistence and bestow some as presents on another. Giving presents would invoke harmony of sentiments between the giver and receiver and generate social and moral approbation. In return, and for exactly the same reason, he would receive as presents those things, which the other has acquired, above his, or her, need for subsistence. There are two ***unintended*** consequences from this process. Firstly, because people are more likely to have surpluses in those things in which they are good in making, they will realise that through the exchange of surpluses they can acquire more goods than they would have, had they themselves strove to acquire them directly from nature. "By this disposition" writes Smith. "to barter and exchange the

[6]For a more detailed analysis of this, see Witztum (2005b).

surplus of one's labour for that of other people …. he will live better than before and will have no occasion to provide for himself, as the surplus of his own labour does it more effectually" (LJ 493). Secondly, individuals become dependent on increasing number of people to obtain their subsistence. Those with whom one would now need to engage in exchange to acquire all life's necessities may be people who until now have been outside one's immediate social circle. It means that the social distance between them is greater. Nevertheless, as the scope of society increases, the drive for social approbation must now be applied to this larger group.

The early stage of specialisation and trade already contains an element of expectation and a question about whether the institutions facilitating specialisation and competitive trade deliver. So, before we move forward to commercial society and the ***business*** of specialisation and trade where we move into domains much larger than our immediate social circle and where our ability to feel sympathy—and hence, find social approbation—with the others is diminished due to increased social distance, let us examine whether the expectations in this very limited and primal set-up, are fulfilled. To be specific, we ask whether according to Smith, the expectation to find social approbation and sufficient subsistence through exchange of gifts and the subsequent specialisation and trade would be fulfilled. Diagram 1 captures this system.

At first, in Diagram 1, people who are driven by the desire to acquire social approbation, exchange gifts or services to persuade, socialise and establish their position in society. They also realised that by further specialisation they can extend the process of exchange not only to enhance their sense of harmony (through agreements reached in exchange) and hence, sociality, but also better satisfy their material ***needs***.[7]

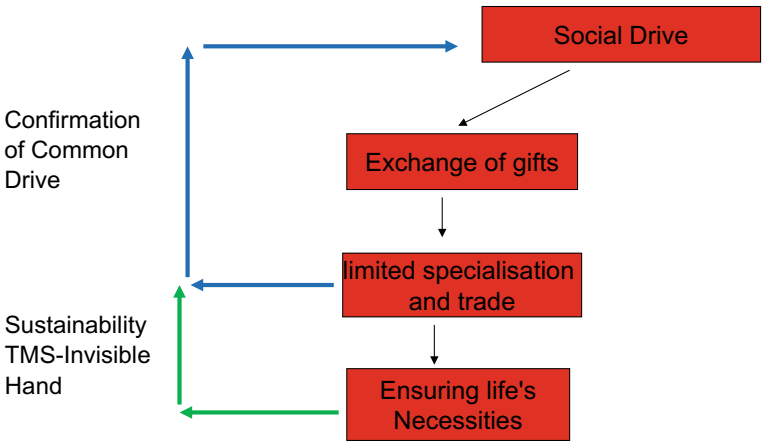

Diagram 1 Fulfilled expectations in the initial cycle of specialisation and trade in Smith

[7]Here, we are referring to what one may need to subsist. Better satisfy their needs does not mean to have as much as possible from everything but have a better way (less exertion—more socialising) to get that which one needs. The notion of needs—or subsistence—should not be taken here literally

Had such exchanges among one's immediate neighbours been sufficient to provide for life's necessities, this is where, in principle, the story would have ended, and we would have probably never entered the domain of an economic system. Namely, through a certain degree of specialisation, both life's necessities and one's social standing would have been satisfied.[8] Notice that the specialisation which takes place in this part of the social scene is not aimed at maximising material wellbeing. It is only aimed at achieving social approbation without harming the ability individuals already have to provide themselves with the basic bundle.

However, Smith recognises that the moment we choose to specialise in order to create a surplus which we could use for the exchange of gifts, we become dependent on an increasing number of people as far as our subsistence is concerned (the second unintended consequence). The question that arises is the extent of this dependency. So far, we mentioned the situation where the immediate social group—the people with whom we find it easier to sympathise in the sense that we would be able to feel as they would have we been in their place—is also the entire group with whom we exchange gifts and which suffices to supply everyone with life's necessities. But as life's necessities develop, the group on which we become reliant to supply our life's necessities (given that we are busy in producing that in which we chose to specialise) becomes much larger and now include people who are socially further from us and with whom, we find it more difficult to feel sympathy as we are less familiar with their particular circumstances.[9] This means that we are now taking a risk by specialising and trading that we may not be able to acquire both the social approbation we seek as well as our life's necessities.

Therefore, for specialisation to deepen, or even take place at all, people must have some confidence that by doing this—in order to achieve their social objectives—they will not become worse off materially than they would have been had they not specialised. They must have confidence in the process of exchange for specialisation to take place:

> And thus the ***certainty*** of being able to exchange all that surplus part of the produce of his own labour, which is over and above his own consumption, for such parts of the produce of other men's labour as he may have occasion for, encourages every man to apply himself to

---

as reflecting physical survival. We are referring here to a bundle which allows people to function socially. What is contained in such a bundle may vary over period and with the wealth of society in general.

[8] The cycle begins with the social drive, goes through exchange of gifts or surpluses to specialisation and trade and feeds back through the blue line which proposes a confirmation of one's social standing through the exchange of gifts or surpluses without harming once ability to satisfy his or her needs. This is the first aspect of the sustainability of the system in the sense that the attempt to acquire social approbation through specialisation has been successful. For Smith, the empiricist, the fact that people did not revert back, in the past, to autarky is the ultimate proof of this.

[9] Indeed, Smith is quite clear that the exchange of gifts—or the specialisation, which is motivated by the desire to persuade—would not be sufficient to supply us with our needs. "In civilised society", he writes, man "stands at all times in need of the co-operation and assistance of great multitudes, while his whole life is scarce sufficient to gain the friendship of a few persons" (*Wealth of Nations (WN)* 26).

> a particular occupation, and to cultivate and bring to perfection whatever talent or genius he may possess for that particular species of business. (*Wealth of Nations (WN)* 28 my italic)

The empiricist nature of Smith's investigation suggests that he thought that at least this first part of specialisation and trade works. Namely, that the confidence which people have that by specialisation they will not undermine their own ability to acquire life's necessities has been quite high. But Smith also provides a more universal reason for this. It is the 'invisible hand' mechanism which is mentioned in the *Theory of Moral Sentiments (TMS)*. It suggests that as the stomach of the rich is never greater than that of the poor, the fact that through specialisation, the same labour time can produce more output means that subsistence would always be distributed in exactly the same manner had all individuals cared for themselves without dividing their labour (see TMS (184) and in the LJ (194-5)). It is, in modern term, what one would call a trickledown theory.

Therefore, it is evident from this analysis that the expectations people have from specialisation and trade to achieve both social approbation—through the exchange of gifts—without harming their ability to supply their life's necessity is fulfilled by the virtue of what he called, the 'invisible hand' as described in the TMS. But the story does not end here. The fact that the social circle had to be extended means that there are people who are sufficiently distant from the person but are now part of his, or her, social/economic circle. Surely the search of social approbation should also be extended to these people too.

To better understand the idea of social approbation and with it, the evolution of the system into the commercial phase, we must say a few words about Smith's ethics. This is not the place to explore the subject in detail, but I would like to draw some general lines.[10] In Smith, we form moral opinion on the basis of a certain harmony of sentiments. This means that we will approve of another person's action or behaviour according to whether we, had we been in their positions, would have felt and acted in a similar manner. This is what Smith means by sympathy. However, in order for such a judgement to be genuine and for our own character not to intrude in the process, Smith resorts to an idea—almost Kantian in nature—of the impartial spectator. Ostensibly, this means that the question we pose when evaluating others is not whether we would have felt or acted the same but whether an impartial spectator would have felt this way. Forming an opinion by means of the idea of an impartial spectator is what people naturally try to do when they come to morally evaluate the other, but this is not an easy task and there are pitfalls on the way. For one, the exercise of considering what would an impartial spectator think is very demanding and requires a considerable amount of interest in the other from those who do the assessment. But as the outcome of the evaluation is a sense of harmony of sentiments (sympathy), some of us tend to conflate the pleasure of harmony in general with a sense of moral judgement. Thus, according to Smith, we may sometime be tempted to believe that because a system is beautiful and well contrived (i.e. invokes the harmony of aesthetics), it must also be morally good.

[10]For a more detailed study of Smith's ethic see, Witztum (1997, 2005a).

The question that arises is whether there are some characters, more than others, who are more prone to conflate the sense of harmony derived from beauty (which Smith calls, utility) with morality. The answer is evidently yes. As we said before, to engage in proper evaluation through the idea of the impartial spectator, one would need to have enough interest in the other to engage in such a complex and demanding process. Those whose interest in the others is limited, are more likely to fall for the harmony from utility as a substitute to proper moral judgement. As I said earlier, Smith did argue that the character behind the commercial system is not necessarily the Mandevillian vile selfish person but rather a person with a prudent self-regard or, self-interest. However, as I demonstrated in Witztum (1998), such a character is also bound to be the one who is least likely to engage in proper assessment of the other and is more likely to use utility, or a sense of harmony due to the beauty of things, to form a moral opinion. This means that such people may judge the system of specialisation and competitive trade as morally good just because of its contrived organisational beauty.

In addition to this, Smith draws our attention to the fact that nature has its own agenda and it is, according to Smith, the multiplication of the species. For this to be fulfilled, nature would like people to create material wealth which will allow the growth of population. To achieve this, nature planted in humans a tendency to find it easier to feel sympathy with the rich (as the sentiments are pleasant) than with the poor. This would mean that in a society which is sufficiently large so that the social distance between individuals is insufficient to generate an informed opinion about the actions of the other (which is what is required for sympathy), accumulation of riches may become a way in which one can gain greater command of other people's admiration than if one were poor. The implications of this for the question of achieving social approbation is that if there is income inequality, only the rich will achieve the approbation they desire while the poor would fail to achieve it. If there is equality, then people would not be able to use their relative wealth to command respect and instead, have to resort to behaviour that would genuinely produce an approval by an impartial spectator.

So, let us see now how the expansion to the commercial phase developed in Smith's analysis and whether the expectations of the agents—which may, in principle, change—have been fulfilled. We now expand Diagram 1 into Diagram 2:

At the end of the red cycle (top left of Diagram 2), individuals have specialised and traded and have achieved social approbation as well as guaranteed supply of life's necessities by virtue of the TMS's invisible hand. The system is, in principle, sustainable. However, as the circle of people on whom one depends for supplying one with life's necessity is greater than the initial social circle for whom the exchange of gifts had been triggered, the question of social approbation within the larger circle remains unresolved. Moreover, material progress and social development further increase the dependency on an increasing number of individuals with whom the ability to find harmony of sentiments diminishes. Evidently, the prospect of finding social approbation seems remote and this could, in principle, trigger a change in expectations and behaviour. But this is not what happens in Smith's analysis. The

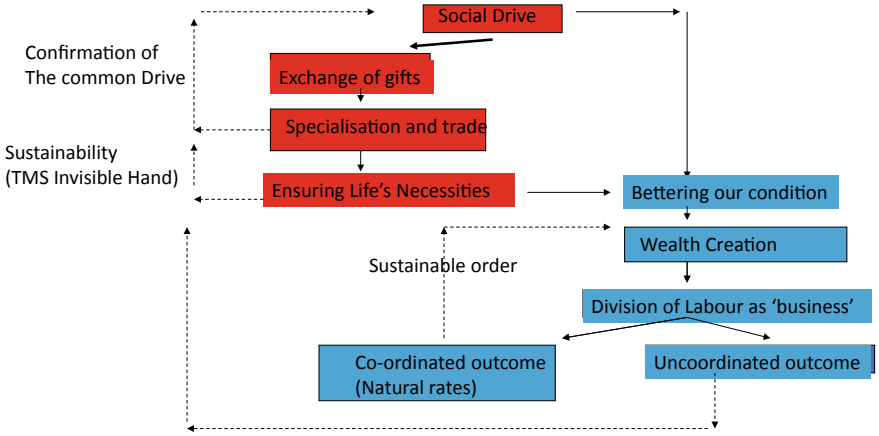

Diagram 2 Commercial stage in Smith

response of people to the expansion of social dependency which characterises commercial societies is to further divide labour (increase specialisation) in the hope of '***bettering their conditions***'. It is where specialisation and trade become, so to speak, a '***business***' (to use Smith's own words).

The motive of 'bettering our condition' which is the motive behind the WN (the second cycle: the right-hand side of Diagram 2) has been interpreted by some as a vindication of the connection between Adam Smith and the modern notion of utility maximisation. However, this cannot be furthest from the truth. In fact, the idea of 'bettering our condition' is not very different from the motive which stood at the heart of the first cycle:

> From whence, then, arises that emulation which runs through all the different ranks of men, and what are the advantages which we propose by the great purpose of human life which we call bettering our conditions? To be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation, are all the advantages which we can propose to derive from it. (TMS 50)

Given the context of Smith's general theory of morals and society this leads him to the following conclusions regarding the way in which such an objective can be achieved:

> We desire both to be respectable and to be respected … To deserve, to acquire, and to enjoy the respect and admiration of mankind, are the great object of ambition and emulation [i.e. bettering our conditions]. Two different roads are presented to us, equally leading to the attainment of this so much desired object; the one, by the study of wisdom and the practice of virtue; the other, by the acquisition of wealth and greatness. Two different characters are presented to our emulation; the one, of proud ambition and ostentatious avidity; the other, of humble modesty and equitable justice. (TMS, p. 62)

Moral and social approbation, in Smith, are the result of the natural tendency which people have to find harmony of sentiments between them and the others (sympathy).

A proper examination of such a process suggests that the road to social approbation should lead to wisdom, virtue, humble modesty and equitable justice. However, as we said before, nature which is concerned with the production of plenty and the multiplication of the species distorts this logic and lures individuals to the second form of behaviour:

> Though it is in order to supply the necessities and conveniences of the body, that the advantage of external fortune [material wealth] are originally recommended to us, yet we cannot live long in the world without perceiving that the respect of our equals, our credit and rank in the society we live in, depend very much upon the degree in which we possess … those advantages. The desire of becoming the proper objects of *this* respect, of deserving and obtaining *this* credit and rank among our equals, is, perhaps, the strongest of all our desires…. (TMS p.213 *my italics*)

So now we have a very different situation to the one we faced in the first cycle. If before we sought approbation while easing the way in which we acquire life's necessities, now we want to acquire more goods in order to receive recognition.[11] The question which then follows is whether or not this pursuit is fruitful. Will people who get into a race of acquiring material wealth succeed in doing so as well as achieve the admiration from the others and the social rank to which they aspire? The focus thus shifts from mere material sustainability to a question of moral acceptability. This is depicted at the bottom right-hand side of Diagram 2.

To see whether the system fulfils the expectations which individuals have with regard to both the acquisition of life' necessities (broadly conceived) and the achievement of social approbation, we must examine two situations which are explicitly analysed by Smith. The first one is a situation where markets reach equilibrium at what Smith calls, natural rates and the other, when prices are just market prices, distinct as they are from their natural rates.

What Smith considers to be the natural rate is not what many modern economists call the long-run prices (which are usually the same as cost of production) but something quite different.[12] It is the rate that will emerge when everyone involved in the production process draws from it only that which is required for social subsistence.[13] There are two implications for the expectations which agents had at the beginning of the process. Firstly, there may not be equality of income but there will be equality in consumption. This, again, means two things: firstly, that each individual would obtain life's necessity almost in the same vein it was guaranteed in the first cycle (top left of Diagram 2); and secondly, that every individual would be equally able to command social approbation through behaviours that would please an impartial spectator. Namely, expectations would be fully fulfilled. The second implication, or perhaps, by-product of this outcome, would be growth maximisation. When all

[11] Note again that while the appearance of wealth accumulation may suggest a resemblance between Smith and modern analysis, the aims of this exercise are very different in Smith.

[12] See a discussion in Witztum (2009).

[13] While Smith himself was concerned with subsistence in the physical sense, one can easily interpret this bundle as that which is socially required to become a member of society (which is somewhat akin to the ideas used by the capabilities approach).

surpluses—above that which is needed for social subsistence—in the system would be used to employ what Smith calls productive labour or, in modern terms, be used for investment, growth will be maximised.[14] Therefore, the system of specialisation and competitive trade may, in principle, fulfil the expectations of all its participants and thus become sustainable in the sense that no one would wish to change their behaviour or bring about a change in the system. This is captured in the left-hand side of the bottom part of Diagram 2.

However, Smith was very sceptical about the emergence of natural rates. At some point, he even comments that market prices may deviate from their natural rates for centuries. Now, what is it that would lead to such a deviation? Well, the answer is embedded in the deception by nature and the corruption of moral sentiments. In Smith's analysis of the move across the various stages of social development, an important element is the emergence of private property. This means that some members of society own property while others do not. Those who live by their wages will get their social subsistence through them but would not be able to influence the use of the surpluses which are normally owned by those who own assets. Consequently, whether the economy reaches the natural rates depends primarily on the behaviour of those who own the surpluses. If they take out of it (by way of consumption) only that which is needed for their social subsistence and saves the rest, prices will be at their natural rates, growth will be maximised, and social approbation would be acquired through good behaviour. If they choose to use parts of the surplus for ostentatious consumption, they will reduce savings and growth, prices will deviate from their natural rates and social approbation will be acquired through luxurious consumption. By implication, those who live of their wages will be relatively poor, admire the rich and fail to acquire social approbation. So, while the market system will, in principle, allow everyone to acquire life's necessities it would, most of the time, fail most people in their attempts to acquire social approbation.

The question that arises is why would such a system be sustained? The Smithian answers lie in the corruption of moral sentiments. The extent of commercial society and the difficulties which many have with properly considering the views of an impartial spectator when social distances are great, allow for nature's deception to work in the sense that people have an incentive to accumulate material wellbeing as a means for commanding social approbation. This means that even people who are poor will still think that working harder towards the acquisition of wealth would gain them the social respect they crave. Modern notions like the American Dream or the celebrity culture we observe are testimonies to the validity of Smith's position. But that which really helps the system to survive is, of course, the corruption of moral sentiment by appreciating the beauty of competitive commercial society where allocations and distributions happen spontaneously without any intervention. The beauty of this machine—manifested in the architectural beauty of financial centres around the world—convinces many of the morality of the system. Consequently, the failure of the poor to command the social respect they crave, in such a beautiful set-up, is perceived to be their own fault rather than a problem with the system.

[14] See Witztum (2009).

In addition to this, the circular argument that promotes self-interested behaviour in a system which (wrongly) claims to allow everyone to achieve that which they want, enhances the social distance between individuals, gives rise to the desire to accumulate material wealth and facilitate the false appreciation of this system; this very argument, provides the system with the sustainability it needs.

4 Conclusions

The main interest in expectations which one finds in modern economics is about the transition process between different points of equilibrium. However, though these expectations are normally confined to the expected value of economic parameters, they hide in them expectations with regard to the behaviour of others (p-expectations) which, in turn, are the results of the expectations which people have about the expected consequences (c-expectations). Economists normally take the latter two for granted; namely, people always behave in a particular frame of rationality and they all expect beneficial outcomes from their participation in the game.

In this paper, we focused our attention on the third aspect of expectation which would normally condition the way we treat the first two. Namely, unless people expect the outcome to be beneficial to them, they would not expect others to behave in that manner that would produce the anticipated value. Therefore, the focus of attention was the promise of beneficial outcomes.

We began by asking a simple question about the benefits which people can expect from a competitive system between the equilibria of which, the discussions about expectations in economics take place. We suggested that there are three reasons to suspect that there is something false in the presumption that there are clear and universal benefits embedded in competitive equilibrium. The first is the endemic problem of missing markets which suggests that we would misinterpret the significance of economic variables with regard to their proposed benefits. The second is the question of the non-economic consequences of economic outcomes. Here, we referred to the question of whether the social and moral view on the activity (i.e. behaviour) that leads to the proposed beneficial outcome, is consistent with prevailing social and moral values. The third is more a methodological problem with regard to the question of whether the values that promise beneficial outcomes are achievable in any form of reasonable dynamics. We showed here that in terms of price formation, it is far from evident that the expected beneficial values will ever materialise or that we may not lose hope of these benefits in the process in which not everyone is benefiting from the system.

We then went back to Adam Smith, as a representative of the classical school to see whether the competitive system fares better in terms of its potential in fulfilling the expectations which agents have. We showed that in the Smithian system, the expectations from the economic system are predominantly social in the sense that people would like to get out of it a living wage (to use a modern term) as well as the ability to acquire social approbation. Through the deception of nature and the

corruption of moral sentiments, we show that only when the system is in equilibrium at natural rate will it fulfil the expectations that bring people to participate in it in the first place. However, these natural values are rare and in all other equilibria, while individuals may still get their living wage (by virtue of the invisible hand from the TMS), most will fail in acquiring the social respect they crave.

All these failures should have led to the conclusion that competitive systems are not sustainable in the long run and that all these disappointed participants would either change their own behaviour or demand a change to the system. Modern economics does not offer an explanation as to why this is the case, but Adam Smith does: that which keeps the system from collapsing is the powerful deception by nature coupled with the self-interested nature of human motivation, which against the power or reason, conflate the beauty of the economy with its social and moral goodness.

References

Crockett, S., Spear, S., & Sunder, S. (2008). Learning competitive equilibrium. *Journal of Mathematical Economics, 44*(7), 651–671.

Fisher, F. M. (2011). The Stability of general equilibrium what do we know and why is it important? In P. Bridel (Ed.), *General Equilibrium Analysis: A century after Walras* (pp. 34–45). London: Routledge.

Goeree, J. K., Lindsay, L. (2012) Stabilizing the economy: Market design and general equilibrium. *Working Paper.*

Hahn, F. (1962). A stable adjustment process for a competitive economy. *Review of Economic Studies, 29,* 62–65.

Hollis, M. (1987). *The cunning of reason*. Cambridge: Cambridge University Press.

Kaldor, N. (1972). The irrelevance of equilibrium economics. *Economic Journal, 82,* 1237–1255.

Lachmann, L. M. (1943). The role of expectations in economics as a social science. *Economica, 10*(37), 12–23.

Mandeville, B. ([1714]1988). *The fable of the bees*. Liberty Press.

Manea, M. (2015). Models of bilateral trade in networks. In Y. Bramoulle, A. Galeotti, & B. Rogers (Eds.), *The Oxford handbook on the economics of networks*. Oxford University Press.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29,* 315–335.

Osborne, M. J., & Rubinstein, A. (1990). *Bargaining and markets* (Chap. 6). Academic Press.

Scarf, H. (1960). Some examples of global instability of competitive equilibrium. *International Economic Review, 1,* 157–172.

Smith, A. (1976a). *The theory of moral sentiments*, (1759). In D.D. Raphael & A.L. Macfie. Oxford: Clarendon Press.

Smith, A. (1976b). An inquiry into the nature and causes of the wealth of nations (1776). In Campbell, Skinner, & Todd (Eds.), *Liberty classics*. Oxford: Oxford University Press.

Smith, A. (1978). Lectures on Jurisprudence. In Meek, Raphael & Stein (Eds.), *Liberty classics*. Oxford: Oxford University Press.

Uzawa, H. (1959). Walras' tatonnement in the Theory of Exchange. *Review of Economic Studies, 27,* 182–194.

Walras, L. ([1874]1954). *Elements of pure economics* (Orion Editions). Lessons: 5, 11–15.

Witztum, A. (1997). Distributional consideration in Smith's concept of economic justice. *Economics and Philosophy, 13,* 242–259.

Witztum, A. (1998). A study into Smith conception of the human character: Das Adam Smith problem revisited. *History of Political Economy, 30*(3), 489–513.
Witztum, A. (2005a). Economic sociology: The recursive economic system of J S Mill. *Journal of the History of Economic Thought, 27*(3), 251–282.
Witztum, A., (2005b) Property rights and the rights of the property less: A note on Adam Smith's Jurisprudence. *Economics and Philosophy*, *21*(2), 279–289.
Witztum, A., (2005c) Social circumstances and rationality: Some lessons from Adam Smith. *American Journal of Economics and Sociology, 64*(4), 1025–1048.
Witztum, A. (2009). A Smithian notion of general equilibrium. In J. Young (Ed.) *Elgar's companion to Adam Smith*.

Expectations, Conjectures and Beliefs. The Legacy of Marshall, Kahn and Keynes

Maria Cristina Marcuzzo

1 Introduction

How people form expectations and make conjectures are issues which have been discussed in the literature along two main lines of investigation: (a) the way to formalize expectations in micro/macro models, when it is assumed they have an important role to play; (b) the evidence from surveys and/or experiments.

For instance, in consumption theory, the life cycle and permanent income approaches bring particular focus to bear on the role of expected future incomes. Investment decisions are assumed to be taken on the basis of expected future prices and sales. Equity prices, interest rates and exchange rates also depend on their expected future values.

These expectations are modelled on the basis of the way people are believed to form their views about the future; generally speaking, there are two alternative approaches: the adaptive hypothesis (of which there are several versions, as we shall see) assumes that expectations are based on what has happened in the past. Rational expectations are a hypothesis formulated in response to the empirical inadequacy and theoretical limitations of adaptive expectations, especially in the context of the Phillips curve (the well-known Lucas critique). Rational expectations state that

> agents' predictions of the future value of economically relevant variables are not systematically wrong in the sense that all errors are random. Equivalently, this is to say that agents' expectations equal true statistical expected values. An alternative formulation is that rational expectations are model-consistent expectations, so the agents inside the model assume the model's predictions are valid (Janžek and Ziherl 2013, p. 172).

A third approach—sometimes presented as a combination of the other two—is the adaptive learning hypothesis, whereby agents adjust and revise their expectations as new evidence turns up:

M. C. Marcuzzo (✉)
Sapienza, Università di Roma, Roma, Italy
e-mail: cristina.marcuzzo@uniroma1.it

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_3

> Adaptive learning models attempt to describe the behaviour of agents faced with repeated decision problems by assuming they use simple learning rules. These models are used in a number of apparently disparate environments. Economic theorists have analysed them in abstract settings. They have been fitted to actual choice data both in economic experiments and the quite different context of the empirical analysis of consumer behaviour (Hopkins 2007, p. 348).

A less theory-laden approach is to be found in survey methodologies—which developed in the late 1930s—where respondents were asked to give qualitative responses on their expectations regarding output, prices and income. Since the early 1990s, moreover, respondents have been asked to report probabilistic expectations of significant events over a wider spectrum of variables, including things like job security, returns to schooling and so forth.

Finally, in more recent times, expectations and conjectures have taken a central role in economics as well as in political science through laboratory experiments and an extensive literature has grown addressing the various statistical and econometric problems connected with it. This literature claims to be able to overcome many difficulties involved in the other approaches (see below).

There are, however, certain other aspects, which may still require attention, namely the specific role of expectations, together with the related notions of conjectures and beliefs pertaining to the degree of confidence attached to those expectations, in any given theory, and how they are conceptualized in alternative theoretical contexts. Indeed, conjectures and beliefs are complementary to the concept of expectations, indicating their hypothetical nature, involving varying degrees of confidence and subjective evaluations of the evidence.

While frequency probability distribution of future events is the framework for analysis of expectations most employed in modern standard economics, there are alternative representations which do not rely on this assumption, either rejecting that particular view of probability or working with less stringent assumptions about the drivers of economic behaviour.

In this paper I will review the ways three economists—Marshall, Kahn and Keynes—belonging to the Cambridge tradition—dealt with these issues, bearing in mind that since Keynes was a pupil of Marshall and Kahn a pupil of Keynes, the family resemblance is bound to be very close. The order of presentation of their views here, however, reflects a less straightforward lineage, which is why I choose to discuss Kahn before Keynes and briefly return to Marshall in discussing Keynes.

Kahn enlarged the scope of Marshall's analysis of the short period in his Dissertation for the Fellowship at King's (where he was elected in March 1929, see Kahn 1989), providing further conceptual elements to the framework, which Keynes would use in the *General Theory* (1936). This work connects—although the literature is divided on how strong this connection is[1]—with Keynes's approach to probability developed in his 1921 book. Kahn's approach to liquidity preference draws on the *General Theory*, but takes the analysis of decision-making behaviour in the speculative demand for money further. While, of course, we owe to Keynes much of

[1]For two somewhat opposite views on the connection, see Carabelli (1988) and Dardi (1991).

our understanding of how expectations impact on speculation and in general on the working of financial markets, Marshall had also come up with significant insights into these matters much earlier on.

The purpose of this paper is to portray a mode of inquiry into expectations by three Cambridge authors in which the expectations are not conceptualized or modelled on the basis of a probability distribution. As to whether this is due to a clearly stated opposition (as in the case of Keynes) or want of the appropriate technique, or indeed a different research approach environment, there may be more than one answer. Within its limited and non-exhaustive scope, this paper offers an interpretation based on the idea that these economists shared a view of the method appropriate to economic theorizing. In Sects. 2, 3 and 4, I present a summary of the main points made by Marshall, Kahn and Keynes on the role of expectations; Sects. 5 and 6 address two issues relevant in contemporary discussion, i.e. the role of expectations in generating market instability and the advantages of taking future markets and experiments as evidence of observable expectations. This latter point leads to a brief discussion on the dividing line between two currents of thought in the Cambridge tradition, namely subjective vs. observable quantities, associated with the followers of the view of the matter taken by Keynes on the one hand and by Sraffa on the other. Some concluding remarks are presented in Sect. 8.

2 Marshall

I would like to start from two points, which have been raised, in the most innovative literature on Marshall:

> Period analysis is machinery that facilitates the analysis of, as well as the formation of, expectations. By means of this machinery we take a market here and now and picture to ourselves the consequences of the different tendencies that we assume market agents *to imagine* to be at work. These tendencies we classify according to whether they are more or less immediate in their effects (Cook 2008, p. 8).[2]

> The impact of a shock, the resulting extent of disequilibrium, and the path to recovery all depend on the agents' ability to interpret current events and *to forecast their consequences*… the key ideas are represented by the two connected concepts of 'normal' states and 'representative' individuals… [both concepts] serve to characterize equilibria as situations in which not only is supply equal to demand, but all divergences between what individuals see and *what they expect to see*, or between what they are and *what they expect to become*, are completely explained by circumstances that do not appear to be persistent enough to make a change in plans worthwhile (Dardi 2006, pp. 221–222, my italics).

What this literature teaches us is that the supply and demand apparatus should not be seen just as a mechanical tool designed to determine equilibrium price and quantity in each market, resulting from maximizing rules followed by consumers and producers, regardless of their expectations, the focus being only on preferences and incentives within their constraints. Rather, the supply and demand apparatus should

[2]The same point is made by Loasby (2002).

also be seen as a means to interpret situations in which expectations are fulfilled. Given their knowledge of the environment, and the routines the various economic agents follow on the basis of that knowledge, if they see no reason to expect a change they behave in such a way that their expectations are confirmed. It is this kind of conformative behaviour that explains equilibrium, not that agents behave as predicted by the model because they know the structure of the model of the economy.

Expectation formation derives from individuals' assessments of the circumstances in a given environment which is characterized by customs and practices related to the specific trade or business. According to Marshall, markets in general function successfully only because the participants observe social norms in trading. In fact, in his view, individuals are not abstract agents, disembodied from their social or economic *milieu*, but representative of a class of people with a well-established identity. As Loasby (2002, p. 5) clarifies:

> The standard practice of deriving equilibria directly from the basic data of the model - goods, preferences, and production sets - requires a comprehensive data set; knowledge must be relevantly complete. In Marshall's system knowledge is never complete, for that would exclude the possibility of generating new knowledge; on the other hand every agent has a history, which has left that agent with a cluster of productive and decision-making skills, including a set of expectations which provide a baseline for conscious thought, and a cluster of connections to other agents and to institutions which may be expected to guide behaviour within groups.

In Marshall's view, expectations mould behaviour, but while they certainly have a subjective element, which explains why they may differ from one individual to another, in an equilibrium situation they are shared by the class of participants in that market, signalling not necessarily optimality, but conformity to their anticipations and forecasts.

For instance, entrepreneurs introduce innovations on the basis of their own subjective evaluation of the future profitability of their investment, the pattern of demand and so forth, so that of the many only those who are able to foresee the actual outcome will be successful. However, this is not so much a matter of individual perfect foresight as ability to adjust—through trial and error—to market twists and turns, assuming that individuals have varying decision-making skills. In Marshall, the role of expectations is to produce market equilibria, conceived as the stopping rules of choices by non-identical agents. This applies equally to the short period, when there is inducement to vary only the level of utilization of capacity, and to the long period, where expectations are such as to induce entrepreneurs to vary capacity.

3 Kahn

Kahn pursued Marshall's distinction between the short period and the long period in his Fellowship dissertation, and in the unfinished book bearing the same name, adding to it some further considerations (see Marcuzzo 2011). He did so, first, by pointing out the difference between the life of fixed capital and the period of production,

the former being considerably longer than the latter, as a *factual* dichotomy; and, second, arguing that the *actual* length of the short period cannot be defined on the basis of technological considerations alone. These elements determine only the upper and lower boundaries of the interval in which the short period varies, because fixed capital cannot remain constant for a period longer than its physical life or be substituted before the current period of production comes to an end. Within this interval, the length of the short period is determined by the beliefs of those who make decisions about capacity, namely the entrepreneurs. Accordingly, depending upon whether the entrepreneurs believe the changes in demand to be transitory or permanent as compared with the level considered "normal", the decision whether to modify plant or organization—typically long-period decisions—will or will not be taken. If the majority of entrepreneurs do not expect that a given increase in demand is likely to last, no investment in capacity will occur in the economy and a general depression will ensue.

The two aspects—the nature of the production process, which is characterized by changes that occur rapidly (in output and employment, for example) and others that occur only slowly (such as alterations in fixed plant), and expectations of changes in demand relative to the level perceived as normal—can explain why capacity is not altered in the short period. This is so because a change in the conditions of demand is not expected to be permanent; in fact, the "ideal" short period is defined as a situation where any change that occurs is not believed to be permanent.

The conditions prevailing in the economy—depression or boom—mould expectations of a return to normal conditions of demand and introduce an asymmetry in the length of the short period. In a boom, short-period equilibrium implies that expectations are such that increasing production, at higher costs, is preferred to building up capacity until the increase in demand is confidently perceived as permanent; the short period, however, can be very short, and innovations are introduced rapidly. In a depression, short-period equilibrium implies expectations that demand will return to its normal level; in this case, the short period can last decades if the firm survives (at a loss) while its physical capital is decaying.

Since what matters are expectations regarding the normal value of the level of demand, it follows that the short period need not be a "short" time interval, nor is it a transitory state before the long-period forces work out their effects. It is, rather, a position that is maintained as long as the set of decisions depending upon the expected level of demand do not change. This attributes to short-period equilibrium a meaning showing a fairly evident analogy with the unemployment equilibrium of Keynes's *General Theory*. The other analogy is with the benchmark for expectations: as in Keynes the "normal" level of rate of interest—in the mind of each the speculator—is the parameter for deciding whether the market rate of interest will rise or fall, so in Kahn the "normal" level of demand is compared by each entrepreneur with the actual level to guess whether or not it will change in the future.

Moreover, this approach to expectations is characterized by the expected duration of an effect and the degree of confidence to which the expectations are held. This latter aspect was particularly stressed by Keynes, but was taken up by Kahn in his own version of monetary theory, in relation to the analysis of the long-term rate

of interest as depending on the expectations for future rates held with uncertainty (Fantacci–Marcuzzo–Sanfilippo 2014).

According to Kahn, in the financial markets there are two kinds of investors: those who hold definite expectations for the future of interest rates and those "who do not have a clue" whether the rate of interest is going to increase or decrease. The same division exists in the mind of the investor himself, who can reveal contradictory preferences and decide to hold money and securities at the same time, with no definite choice between the two. On the basis of this observed behaviour, Kahn argues that a finite elasticity of the demand for money relative to the rate of interest is associated not only with the heterogeneity of expectations held by the public divided between bulls and bears, but also with the lack of conviction individuals show in their own conjectures. It is as if bullish and bearish sentiments "operated inside each person's mind, one being responsible for his holding securities and one for his holding money" (Kahn 1954, p. 247).

Here expectations are interpreted as conjectures described by different degree of beliefs entertained by each individual, who will make his/her choices on the basis of the strength of his/her beliefs and will change them as the strength of those beliefs weakens or grows. There is an obvious similarity with Keynes's "weight of the argument" as presented in his probability theory and applied to his economic analysis of decision-making under uncertainty (on which see below).

Finally, it must be borne in mind that the degree of confidence in expectations should not be interpreted as the variance in the individual distribution of probability, as in Tobin's portfolio model, since the very nature of expectations lies in their being held with notions of uncertainty and confidence that do not lend themselves to being represented by a frequency distribution.[3] Why this is so becomes clearer when we turn to Keynes.

[3]"Instead of assuming that each investor held with certainty some expectation of the future interest rate, Tobin (1958, 1965) [attributed] to each investor a subjective probability distribution over the return on each asset, and preferences over the mean and variance of returns on their portfolios. Given the same information, all investors would hold the same probability distribution over expected returns, but each investor would have his or her preferences, represented by a map of indifference curves over risk and expected return […] Since expectations are represented by a subjective probability distribution, the model deals with risk (as the title of Tobin 1958 states) rather than with fundamental uncertainty in the sense of Knight (1921) or Keynes (1936, Chap. 12), under which no probability distribution could be used because there is not even a complete list of possible outcomes—yet that is still an advance over assuming that each investor holds with certainty a different expectation of the future interest rate" (Dimand 2014, pp. 67–68).

4 Keynes

In turning to Keynes, I would like to start from an oft-quoted passage in the *General Theory:*[4]

> It would be foolish, in forming our expectations, to attach great weight to matters which are very uncertain […]. It is reasonable, therefore, to be guided to a considerable degree by the facts about which we feel somewhat confident, even though they may be less decisively relevant to the issue than other facts about which our knowledge is vague and scanty. For this reason the facts of the existing situation enter, in a sense disproportionately, into the formation of our long-term expectations; our usual practice being to take the existing situation and to project it into the future, modified only to the extent that we have more or less definite reasons for expecting a change (CWK VII, p. 148).

This, of course, is Keynes's argument for distinguishing between long-term and short-term expectations (CWK VII, pp. 46–47): the former are relevant in determining investment in fixed capital, the latter for the level of utilization; the point made is that, unlike the case of short-term expectations, there is no rational basis for forming long-term expectations. In order to understand what is the rational basis for forming short-term expectations, we need to bear in mind what Keynes meant by "rational".

First, we need to recall how this notion is connected with conjectures. De Carvalho (2015, p. 47) summarizes it thus:

> probability has to do with knowledge obtained by argument, which Keynes meant as being obtained as implication of the premises taken as direct knowledge [which] even if true, is seldom complete. Observable premises have to be completed with *conjectures* to enable the decision-maker to derive their logical implications. In other words, one does not start from true knowledge alone but has to complement it with conjectural premises. If more than one set of conjectures is possible, implied results are conditioned by the set of conjectures selected by the decision maker.

The point stressed here is that since conjectures are subjective, they are related to personal beliefs and individual varying degrees of confidence in those beliefs, which in turn reflect how individuals experience reality. Also Ellsberg (1961, p. 657) "remarked that the nature of the individual's information concerning the likelihood of events is a relevant dimension of the decision problem, and proposed to call it the ambiguity of information, 'a quality depending on the amount, type, reliability and 'unanimity' of information' expressing the individual's "degree of confidence in an estimate of relative likelihoods" (Zappia 2016, p. 850). Studies of ambiguity maintain that beliefs have some but not all the structure of a probability distribution, and this is why decisions under *ambiguity* are considered as alternatives to the hypothesis of probabilistic expectations, which have been met with criticism.

Keynes was not alone in believing in fundamental uncertainty, which does not lend itself to being represented by probability distribution, since also Knight held that for some events no probability function can be constructed that can account for

[4]There is a vast literature on Keynes' approach to uncertainty and probability as the ground in which his notions of expectations is rooted (see Roncaglia 2009). Here I wish to make just a few points on the relation between expectations, conjectures and beliefs.

the phenomena. This was not seen as owing to technical constraints, but because the real world is devoid of such facts as would allow for any such construction, so that "the conception of an objectively measurable probability or chance is simply inapplicable" (Knight 1921, p. 231).[5]

Moreover, according to Keynes whenever new information signals that things have developed in a different way than was expected, the need will arise for continuous adaptation to change. As Roncaglia (2012, pp. 448–449) effectively puts it:

> In Keynes's theory of probability, there is no objective rule to establish how the empirical evidence should affect the probability statement, or as to how additional evidence should change it. Thus, no bi-univocal correspondence can be established between evidence and a 'rational' probability statement. However, in Keynes's mind there clearly is the idea that the subject must somehow take the available evidence into account. In fact, together with internal consistency (no contradictions) in the system of beliefs, this is what distinguishes rational from irrational behaviour.

How individuals form their expectations is kept distinct from how confident they are in their expectations, and this makes the analysis less amenable to deterministic results, unlike the case where the confidence in a given expectation is captured by the variance of a subjective probability distribution of the outcomes.[6] The variability in the confidence of expectations, both among individuals and within each individual, has an impact on how they are transmitted to markets, which in their turn may convey signals to individuals to revise these expectations, producing instability.

5 The Role of Expectations in Generating Market Instability

The importance of uncertainty and the role of expectations in generating instability, especially in the financial markets, have taken centre stage in the aftermath of the crisis. Criticism has been levelled against contemporary macroeconomics for eschewing the question of how heterogeneous individual expectations translate into market expectations or, as often referred to, "market sentiment", through the modelling of the representative agent, whose expectations are by definition market expectations, leading to serious misrepresentations of the working of the economy (see Stiglitz 2017).

---

[5] However, the literature diverges on the interpretation of Knight's subjective probabilities (see for instance Lawson 1988; LeRoy and Singell 1987).

[6] "But the distinction between a probability assessment and the degree of confidence in it has no place in a standard probability approach. Even in the Bayesian approach axiomatised by Savage (1954), decision-makers choose as if they were guided by precise numerical probabilities of the consequences of their action, regardless of confidence. Mainstream decision theory considers confidence at most a second-order probability distribution, reduced to a conventional probability measure through the compounding of probabilities. As a result Keynes's distinction has been long ignored" (Zappia 2016, p. 850).

On the other hand, the hypothesis of heterogeneous expectations is a distinguishing feature of heterogeneous agent-based models. It consists in the adoption of a weighted average of different hypotheses of expectation formation, such as chartists' and fundamentalists' expectations,[7] extrapolative[8] and rational expectations, or extrapolative and regressive[9] expectations.

Moreover, experimental economics has gathered a very large number of observations based on laboratory experiments showing that economic agents generally base their decisions resorting to a combination of extrapolative and regressive expectations, rather than rational expectations as in the DSGE models. These results are confirmed by experimental evidence in asset pricing (Sordi and Vercelli 2012).

In the financial markets, two further elements have been seen at work, acting as higher-degree expectations, so to speak: guessing games, such as the beauty contest (Keynes) and reflexivity (Soros).

The following passage from Marshall's unpublished manuscript reveals that he had already anticipated the "beauty contest" analogy made famous later with Keynes' guessing game of "anticipating what average opinion expects the average opinion to be" (Keynes (1936/1971, p. 156). Marshall wrote:

> [speculators] govern their action not by their own forecasts of the distant future, but by their forecasts of the forecasts that will be made by less competent people…by far the larger part of the attention even of the leading operators is given not to the distant future but to the immediate future…Sometimes he [the ordinary professional speculator] does not care so much to anticipate coming events, as coming popular opinion as to coming events (Dardi and Gallegati 1992: 589).

Self-fulfilling effects of expectations have been explored by Soros with his notion of reflexivity, which explains:

> a two-way feedback mechanism between the participants' expectations and the actual course of events. The feedback may be positive or negative. Negative feedback serves to correct the participants' misjudgements and misconceptions and brings their views closer to the actual state of affairs until, in an extreme case, they actually correspond to each other. In a positive feedback loop a distortion in the participants' view causes mispricing in financial markets, which in turn affects the so-called fundamentals in a self-reinforcing fashion, driving the participants' views and the actual state of affairs ever further apart. What renders the outcome uncertain is that a positive feedback cannot go on forever, yet the exact point at which it turns negative is inherently unpredictable. Such initially self-reinforcing, but eventually self-defeating, boom-bust processes are just as characteristic of financial markets as the tendency towards equilibrium (Soros 2010, p. 4).

The notion of reflexivity was also employed by Soros to challenge the idea that markets can overcome financial fragility by incorporating the risk factor in the pricing mechanism. This idea has been discredited since the financial crisis that began with

[7]Fundamentalists devise their strategies from inference of the asset fundamentals, while chartists establish their strategies on observation of past price movements.

[8]Extrapolative expectations are defined as expectations consistent with the observed trend of the variable.

[9]Regressive expectations are defined as expectations consistent with some long-run average value of the variable.

the Lehmann Brothers bankruptcy, while consensus has grown for the view that faith in stochastic models and in the efficient market hypothesis helped to fuel the crisis, which can more aptly be described as a failure of ideas than a market failure.

How much of this failure can be imputed to the modelling of expectations or rather to the structure of the model (incorporating them), which found popularity in the wave of anti-Keynesianism of the 1990s is a much debated question. For instance, Hoover states:

> Rational expectations was such a striking feature of the new classical revolt against 'Keynesian' economics that for a long time economists such as Lucas, Sargent, Wallace, Barro, Kydland and Prescott were commonly referred to as 'rational expectationists'. The misleading belief that their main results derive principally from the rational expectations hypothesis persists in some quarters even today. This belief is misleading because the most characteristic feature of the new classical school is its adherence to the assumption of continually clearing, perfectly competitive markets as a basis for macroeconomics (Hoover 1997, p. 228).

However, the problem of how expectations are formed is still conflated with the chosen theoretical framework. It seems, in fact, that in most of the standard literature the representation of expectations is dictated by the model structure and modelling strategy, rather than the result of observed behaviour.

6 Observation-Based Expectations

Several economic variables in standard economics are derived from individual optimizing choices, in which expectations are also included, but we observe neither; those choices and expectations are conceptual devices with which economists capture intentional, more often than not, maximizing behaviour that is assumed to lie behind prices and the working of any market. Criticism of common assumptions in standard economics runs as follows:

> Economists commonly assume that persons form probabilistic expectations for unknown quantities and maximize expected utility. Hence, the research problem is to infer the subjective probability distributions that express expectations and the utility functions that embody preferences. The difficulty is that observed choice behaviour may be consistent with many alternative specifications of preferences and expectations. Hence, identification of decision processes from choice data must rest on strong maintained assumptions. The prevailing practice has been to assume that decision makers have specific expectations, which are objectively correct (i.e., *rational*). This practice reduces the task of empirical inference to revelation of preferences alone, but has contributed to a crisis of credibility. Researchers performing econometric analysis of choice data often have enormous difficulty defending the expectations assumptions they maintain and, as a consequence, have similar difficulty justifying the findings they report (Manski 2004, p. 1330).

It is claimed that this difficulty can be overcome by performing experiments given the nature of the exercise, i.e. working with observable variables:

> While appropriate modelling of expectation formation on the part of traders is crucial to understanding the behaviour of asset markets, individuals' beliefs about future prices are

> typically unobservable to researchers. However, modern methodological techniques in experimental finance and economics allow researchers to overcome this unobservability, and do permit direct measurement of expectations, for some classes of markets. The procedure for doing so is to elicit predictions of future prices from participants or observers of experimental markets, and to provide monetary incentives for accurate forecasts (Haruvy et al. 2007: 1901).

According to another strand of literature, which possibly goes back to Working (1949), we already have—without performing any experiment—observable variables that allow us to understand expectations. In futures markets, prices are the observable variables (futures prices) that reflect market expectations. Future prices incorporate the risk premium, which is the return an investor receives in exchange for long-term financial exposure to a particular asset-class thereby assuming the risks associated with rises and falls in the market value of this asset-class. The risk premium is paid as a reward and incentive for taking risk above the risk-free rate, not a reward for predicting market movements. It a is a conjecture incorporated in the expected spot price, which calls forth a reward, making the future price—which is observable—diverge from the expected spot price—which is unobservable.

What we observe in any given moment is not the price expected to prevail at a future date, but the price which is sufficient to induce speculators to undertake the risk, given their expectations of what the spot price will be on maturity of the contract. The problem with this approach, however, is that extracting this information is difficult in practice. Nevertheless, according to Keynes, the risk premium is "the best estimate we can make of probabilities" (Keynes 1936/1971, p. 240). The concept is iterated, even more explicitly, in a letter to Hugh Townshend, dated 7 December 1938:

> I am rather inclined to associate risk premium with probability strictly speaking, and liquidity premium with what in my Treatise on Probability I called 'weight'. An essential distinction is that a risk premium is expected to be rewarded on the average by an increased return at the end of the period. A liquidity premium, on the other hand, is not even expected to be so rewarded. It is a payment, not for the expectation of increased tangible income at the end of the period, but for an increased sense of comfort and confidence during the period (Keynes 1979, pp. 293–294).

7 Expectations as Dividing Issue

The role expectations should have in the theory is in fact at the root of the disagreement between Post-Keynesians and Neo-Ricardians, following Sraffa's work (1960), both claiming to be an alternative to standard economics. As Keynes has it decision-making under uncertainty, opinion formation and subjective evaluation of future events are said to be essential elements; in the case on the contrary, of Sraffa, placing the emphasis on the "objective basis" of market price and distribution determination, refuses to appeal to non-observable entities, such as expected utility, marginal productivities and the like. With particular reference to wages and profit,

this approach invites, rather, to look to custom and social norms, and particularly those that are reflected in the money rate of interest, as Sraffa acknowledges (Sraffa 1960).

Moreover, as far as expectations are concerned, rather than deriving them from a subjective distribution of probability, in a Sraffian approach, they are seen as working in patterns, because imitation and mutuality play an important role. As K. Bharadwaj noted: "We need to discover [the] objective basis of different states of expectations and we need to know what systematic objective outcomes arise from these different states of expectations" (quoted in Marcuzzo 2014, p. 59). It follows that broadly uniform behaviour can be made part of the analysis of those aspects of the economic system where customs and socially determined variables play a role. While the theoretical frameworks of Marshall and Sraffa are obviously alternative in the determination of prices and distribution, they seem far less distant from one another when it comes to the role attributed to routines, customs and socially given constraints.

Marshall, Kahn and Keynes placed great stress on expectations in driving behaviour, but they did not endorse the idea that they could be formalized in a model of the economy for the purpose of empirical testing. This was not so much a matter of inadequate mathematical training (as we know, all three were trained mathematicians) as mistrust of statistical inference,[10] as unfit to be applied to economics. Keynes made the most substantial contribution to rejection of the idea that frequency distributions could be applied to understand economic phenomena, providing a theoretical framework that was original and revolutionary. Biddle (2016) shows that under Keynes's influence, prior to WWII several leading economists rejected probability theory as a source of measures and procedures to be used in statistical inference, because the data available to economists did not satisfy the assumptions required for such an approach.

On a different track, Sraffa—in an even more radical vein—rejected the idea that unobservable entities like utility, expectations and beliefs could be modelled *as if* they were observable quantities. He went so far as to expunge them from the analysis, while Marshall, Keynes and Kahn thought they should be part of it. This reflects the different sources their theories were drawn from which explains why they lead to irreconcilable views on how to account for prices and distribution. Marshall, Kahn and Keynes thought subjective elements—identified with unbounded maximizing rationality—need to be part of the *explanans* of the decision-making mechanism that lies behind market outcomes. Sraffa did not deem it necessary as far the price mechanism was concerned, but left open the possibility that norms and customs, embodying non-observable entities, exercise their influence in other parts of the analysis.

[10]For a review of Keynes's objections to statistical inference applied to economic analysis and econometrics, see Garrone and Marchionatti (2009).

8 Some Concluding Remarks

There seem to be at least three different ways to represent expectations in economics: (a) expectations of future spot prices as revealed in futures markets; (b) expectations formalized in probabilistic and deterministic models (adaptive and rational); (c) expectations in surveys and experiments (both qualitative and probabilistic). The most empirically confirmed type appears to be adaptive expectations, with variations: those based not only on the past level of an economic variable, but also on its direction of change, including reversion towards a long-run normal level (extrapolative and regressive, also in combination). However, in micro-founded optimizing models the preference is for rational expectations, i.e. those formed on the basis of the structure of the relevant system describing the economy.

Clearly, the structure of the model and the choice of the relevant variables play a very important part in the choice of the expectations formation mechanism. Thus, the preference given a particular expectation formation hypothesis is often the result of the choice of the model in which to incorporate it; when the expectations formation hypothesis is only conceptualized, but not formalized, the elements deemed to be important can be detected in the structure of the theory, which acts as less binding constraints on the choice of the assumptions.

In this paper, I have sought to present evidence that Marshall, Kahn and Keynes shared a common stark opposition to probabilistic modelling of expectations and attributed a key role to customs and conventions in explaining them, without denying the role of beliefs and expectations in explaining behaviour and market outcomes. If we include Sraffa in the Cambridge tradition, taken as alternative thinking to mainstream economics, then we can add another argument to the reasons for opposing the modelling of expectations, especially in their probabilistic form, i.e. a radical doubt about the possibility of making non-observable quantities part of any economic theory.

While several of the points made are not unique to these authors and the tradition they represent, since there are other examples of no-probability distribution approaches to expectations, their arguments—as I have tried to show—have aspects of particular interest that justify revisiting them.

Acknowledgements I wish to thank Carlo Cristiano, Marco Dardi, Carlo D'Ippoliti, Luca Fantacci, Daniele Girardi, Paolo Paesani, Alessandro Roncaglia, Eleonora Sanfilippo, Alessandro Vercelli and Carlo Zappia for their help in revising an earlier version. I am also grateful to Robert W. Dimand, my discussant at the Guggenheim conference, for his comments. I have not always followed all suggestions I received so the responsibility for any errors and omissions is mine alone.

References

Biddle, J. E. (2016). Statistical inference in economics, 1920—1965: changes in meaning and practice. *Journal of the History of Economic Thought, 39,* 149–173.

Carabelli, A. (1988). *On Keynes's method*. London: Macmillan.
Cook, S. (2008). Faith in the future: the Cambridge school's expectations, Mimeo.
Dardi, M. (1991). Interpretazioni di Keynes: logica del probabile, struttura dell'incerto. *Moneta e Credito,* 44, 59–88.
Dardi, M. (2006). Partial equilibrium and period analysis. In G. Becattini, M. Dardi, & T. Raffaelli (Eds.), *The Elgar companion to Alfred Marshall* (pp. 215–225). Cheltenham: Elgar.
Dardi, M., & Gallegati, M. (1992). Alfred Marshall on speculation. *History of Political Economy, 24,* 571–594.
De Carvalho, F. C. (2015). Keynes on expectations, uncertainty and defensive behavior. *Brazilian Keynesian Review, 1,* 44–54.
Dimand, R. W. (2014). *James Tobin*. Basingstoke and New York: Palgrave Macmillan.
Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics, 75,* 643–669.
Fantacci, L., Marcuzzo, M. C., & Sanfilippo, E. (2014). A note on the notions of risk premium and liquidity premium in Hicks's and Keynes's analyses of the term structure of interest rates. *European Journal for the History of Economic Thought, 6,* 1002–1008.
Garrone, G., & Marchionatti, R. (2009). The role of econometric method in economic analysis. A reassessment of the Keynes-Tinbergen debate 1938–1943. In R. Arena., S. Dow & M. Klaes (Eds.), *Open economics. Economics in relation to other disciplines*. Routledge, London, pp. 148–172.
Haruvy, H., Lahav, Y., & Noussair, C. N. (2007). Traders' expectations in asset markets: experimental evidence. *American Economic Review, 97,* 1901–1920.
Hoover, K. D. (1997). Is there a place for rational expectations in Keynes's general theory? In G. C. Harcourt & P. A. Riach (Eds.), *A 'second edition' of the General Theory* (pp. 219–237). London: Routledge.
Hopkins, E. (2007). Adaptive learning models of consumer behavior. *Journal of Economic Behaviour and Organization, 64,* 348–368.
Janžek, T., & Ziherl, P. (2013). Overview of models and methods for measuring economic agent's expectations. *IFC Bulletin, 36,* 172–179.
Kahn, R. F. (1954). Some notes on liquidity preference. *Manchester School of Economic and Social Studies, 22,* 229–257.
Kahn, R. F. (1989). *The economics of the short period*. London: Macmillan.
Keynes, J. M. (1936/1971). The General Theory of Employment, Interest and Money. In E. Johnson & D. E. Moggridge (Eds.), *The Collected Writings of John Maynard Keynes* (CWK), vol. VII. Macmillan, London.
Keynes, J. M. (1979). The General Theory and after. A supplement. In E. Johnson & D. E. Moggridge (Eds.), *The Collected Writings of John Maynard Keynes* (CWK), vol. XXIX. Macmillan, London.
Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston: Houghton Mifflin Co.
Lawson, T. (1988). Probability and uncertainty in economic analysis. *Journal of Post Keynesian Economics, 11,* 38–65.
LeRoy, S., & Singell, L. (1987). Knight on risk and uncertainty. *Journal of Political Economy, 5,* 394–406.
Loasby, B. (2002). Expectations, uncertainty and beliefs in the Marshallian analysis of markets. *Revue d'économie Politique, 112,* 659–670.
Manski, C. F. (2004). Measuring expectations. *Econometrica, 72,* 1329–1376.
Marcuzzo, M. C. (2011). The 'elusive figure who hides in the preface of Cambridge books'. An appraisal of Richard Kahn's contributions. In C. Gehrke, N. Salvadori, I. Steedman & R. Sturn (Eds.), *Classical economics versus modern theories. Essays in honour of Heinz D. Kurz*, vol. 2, pp 185–206. Routledge, Abingdon.
Marcuzzo, M. C. (2014). On alternative notions of change and choice: Krishna Bharadwaj's legacy. *Cambridge Journal of Economics, 38,* 49–62.
Roncaglia, A. (2009). Keynes and probability: An assessment. *The European Journal for the History of Economic Thought, 16,* 489–510.

Roncaglia, A. (2012). Keynesian uncertainty and the shaky foundations of statistical risk assessment models. *PSL Quarterly Review, 65,* 437–454.
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley and Sons.
Sordi, S., & Vercelli, A. (2012). Heterogeneous expectations and strong uncertainty in a Minskyian model of financial fluctuations. *Journal of Economic Behavior and Organization, 83,* 544–557.
Soros, G. (2010). *Presentation. Anatomy of crisis—the living history of the last 30 years: Economic theory, politics and policy*. INET Conference King's College, Cambridge April, Mimeo.
Sraffa, P. (1960). *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.
Stiglitz, J. E. (2017). *Where modern macroeconomics went wrong*. NBER Working Paper No. 23795.
Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies, 25,* 65–86.
Tobin, J. (1965). Money and economic growth. *Econometrica, 33,* 671–684.
Working, H. (1949). The investigation of economic expectations. *The American Economic Review, 39,* 150–166.
Zappia, C. (2016). Whither Keynesian probability? Impolite techniques for decision-making. *The European Journal of the History of Economic Thought, 23,* 835–862.

Between Pigou and Keynes: Champernowne on Employment and Expectations

Mauro Boianovsky

But employés [sic] cannot as a rule foresee; and they have less power of acting on their knowledge. The consequence is that a rise in wages is seldom or never as fast as that of prices when the cause of the rise is an increase of the currency, that is not accompanied by an increased command over nature (Marshall 1926; minutes of evidence to the 1899 Committee on Indian Currency).

1 Sorting Out Cambridge Expectations

Alfred Marshall's 1899 brief remark about workers' lack of foresight, and its implications for the determination of money-wages, illustrated an opinion shared, but never articulated, by many economists during the Marshallian era. Indeed, workers' expectations are conspicuous by their absence in the then prevailing Cambridge approach to business cycles—advanced by the Marshalls (1879, pp. 152–155, with references to Lord Overstone and J. S. Mill) and fully elaborated by Lavington (1922) and Pigou (1927)—with its focus on waves of optimism and pessimism in assets markets. It was not just that workers occasionally held mistaken expectations, but that, unlike businessmen and dealers in the financial and investment markets, they could not foresee at all. Jevons's (1871) opinion about the supposed inability of the working class to make inter-temporal choices was representative of economists' widespread exclusion of "The Other" from their economic principles (see Dimand 2005).

Lavington (1922) argued that isolated "impulses" are "propagated" to the rest of the economy through cumulative "contagion of confidence". The errors of optimism affect directly and indirectly (through their influence on credit and therefore on

M. Boianovsky (✉)
Universidade de Brasilia, Brasilia, Brazil
e-mail: mboianovsky@gmail.com

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_4

prices) the estimates of future demand, especially in the investment sector. The prosperity phase ends when the gestation period of the new capital goods is concluded and businessmen realize that their actual yield is lower than anticipated, which is followed by errors of pessimism in the downswing. Pigou (1927), building on Lavington, assumed that short-period shifts in the (discounted) demand for labour are caused mainly through changes in expectations of return. Variations in profit expectations are set off by impulses that may be of "real", "psychological" or "monetary" kinds, which lead to the "mutual generation of errors of optimism and pessimism". He rejected the argument that generalized errors of forecast are impossible in the sense that widespread wrong expectations about the movement of a variable are necessarily fulfilled through interaction between agents (see Kregel 1977; Collard 1996; Boianovsky 2005a).

Like other Cambridge economists before him, Keynes (1936) put expectations at the centre of his macroeconomic framework. However, he departed from the view that unemployment was a short-run disequilibrium phenomenon associated with economic fluctuations brought about by incorrect entrepreneurial anticipations. Keynes preferred model, for demonstrating the role of effective demand in the determination of unemployment in equilibrium, assumed away disappointments and shifts in expectations. In that formulation, the (general) state of long-term expectations, which decides investment demand, is given and irresponsive to (individual) short-term expectations, which are always realized. That also applied to income expectations in Keynes's formulation of the consumption function. Regardless of how agents react to disappointed expectations, the economy moves immediately to the point of effective demand, which may happen at less than full employment for a given "state of the news" (Kregel 1976; see also Bateman 1996, Chapters 4 and 5).[1] Just like Lavington and Pigou, Keynes focused on expectations by businessmen, without dealing in any detail with workers' price expectations and their effects on labour-supply decisions and money-wage dynamics.

Pigou and Keynes were both members of King's College. One of the bright economic students at King's in the early 1930s was David Gawen Champernowne (b. 1912; d. 2000), who put forward in 1936 a path-breaking attempt to sort out the unemployment controversy between Keynes (1936) and Pigou (1933), just four months after the publication of the *General Theory*.[2] Champernowne's mathematical gifts granted him in 1931 a mathematical scholarship to King's, where he was supervised together with Alan Turing, the pioneer in modern computing. His incipient interest in economics, based on his reading of Marshall's *Principles*, was confirmed by an encounter with D. H. Robertson at Cambridge, followed by J. M. Keynes's advice to abandon plans of becoming an actuary and turn to economics instead. Champernowne switched to the Economic Tripos by taking the Maths Tripos in double quick time,

[1] Disappointed expectations are, however, relevant for Keynes's (1936, Ch. 22) discussion of the business cycle (instead of equilibrium positions), which is reminiscent of the Lavington–Pigou tradition (see Haberler 1937, Chapter 6 on "psychological theories", and Boianovsky 2005a).

[2] The following two paragraphs are based on Harcourt (2001) and Boianovsky (2017), and references there cited.

and by October 1933 started (together with B. Reddaway) studying under Keynes's supervision and attending his lectures about the forthcoming *General Theory*. His notes of Keynes's lectures would be used as one of the sources of Rymes's (1987, 1989) well-known compilation.

College supervisions formed the core of Cambridge's didactic system, which paid careful attention to the selection, education and relations with students (see Marcuzzo et al. 2008, with mention of several Cambridge economists but not Champernowne). The young Champernowne experienced Cambridge life intensively. He took part in the selective Cambridge Political Economy Club run by Keynes and became, probably under Keynes's guidance, a member and secretary of the Cambridge Conversazione Society, better known as the "Cambridge Apostles". The Apostles was essentially a conversational society, which required of its members that they be ready to question any established views. The 23-year-old Champernowne followed that strategy closely in his piece about "classical" and "Keynesian" approaches to unemployment, where he dared to challenge both Pigou (1933) and Keynes (1936), the only references mentioned in the article.

Although a student of Keynes, as a member of King's College Champernowne was also in close contact with Pigou, as illustrated by his role—encouraged by Richard Kahn—in assisting Pigou (1938, p. 134) grasping the notion that the rate of interest is the mechanism through which changes in money-wages affect employment. Indeed, Champernowne (1936, p. 202) acknowledged suggestions from Pigou and Robertson on earlier drafts of his essay (Aslanbeigui and Oakes 2007). Throughout his long career, Champernowne enjoyed behind-the-scene refinement of ideas and enlightened commentaries on other economists' works (including famously his 1945 introduction to von Neumann), from which Pigou, Keynes, Robertson, Joan Robinson, Nicholas Kaldor, Piero Sraffa and other Cambridge economists benefited from the 1930s to the 1960s. Apart from that, Champernowne's main contributions were in the field of economic statistics (measurement of income distribution and inequality; and probability, decision making and estimation methods in economics). Champernowne (or "Champ", as he was known among his friends) held chairs at both Oxford and Cambridge universities, but even during his Oxford period (1945–1959) he kept close ties with Cambridge economics.

As discussed below, expectations' formation by both businessmen *and* workers is the key to Champernowne's (1936) effort to sort out the differences between Keynesian and Pigouvian analyses of unemployment. Workers' bargains for money-wages in labour contracts reflect their concern with anticipated real wages, with price expectations determined by prices of the previous period—that is, what we now call adaptive expectations. Unexpected changes of the cost of living bring about shifts of the "real supply curve for labour", accompanied by employment levels above (if prices are rising) or below (falling prices) equilibrium. As workers become aware of price-level movements, the rate of inflation or deflation accelerates, which leads to changes of the rate of interest by the central bank in attempt to stabilize the economy. Workers' demand for a certain real wage is then made effective and unemployment

converges to its equilibrium ("basic") level, unless the influence of exogenous long-term expectations (the "state of the news") on investment and money demand prevents the operation of the interest-rate mechanism.

Champernowne described that last scenario as distinctively "Keynesian". From his perspective, "classical" (Pigouvian) macroeconomics applied to the study of the "trend" values of real wages, output and other real variables in long-period full-employment equilibrium, with no relevant role for Lavington–Pigou short-run waves of pessimism and optimism. He would come back to that topic only once, almost 30 years later, on the occasion of the reprint of his 1936 article in the well-known Lekachman (1964) collection. Together with other contributors who had also provided assessments of the *General Theory* in the 1930s and 1940s (J. Viner, G. Haberler, P. A. Samuelson, B. Reddaway, R. F. Harrod, among others), he was asked to "place in print [his] present evaluation of the Keynesian revolution", as stated in Lekachman's preface. Champernowne (1964) was the only contributor who focused on the theme of expectations as the core of the Keynesian revolution, a fitting follow-up to his 1936 essay. But this time Irving Fisher replaced Pigou as the economist he picked to compare with Keynes, as Champernowne moved away from the labour market and focused instead on assets markets and inter-temporal decisions.

2 Labour Supply and Money-Wage Dynamics

The "fundamental difference" between Keynes's *General Theory* and Pigou's *Theory of Unemployment*, pointed out Champernowne (1936, p. 201) in his opening paragraph, is that, while the former argues that the wage bargains decide the money-wage only, the latter maintains that these bargains succeed in determining the real wage rate. This follows from Keynes's (1936, pp. 10–13) "fundamental objection" to the classical postulate that the real wage corresponds to the marginal disutility of labour (workers are on their labour-supply curve). As put by Keynes (p. 13), "there may exist no expedient by which labour as a whole can reduce its *real* wage to a given figure by making revised *money* bargains with the entrepreneurs. This will be our contention". In order to assess Keynes's rejection of classical macroeconomics, Champernowne put forward a general-equilibrium aggregative model, which, unlike Hicks (1937) better-known IS-LM formulation, highlighted the role of the labour market and pointed to the distinct causality structures and expectations mechanisms of "classical" and "Keynesian" frameworks.

Keynes's (1936, pp. 8–10, 12–13) first objection to the classical analysis of labour supply was not "theoretically fundamental" or logical, but factual, related to the "actual behaviour of labour". Based on his observation of collective bargaining by trade unions, Keynes criticized the notion that the influences of wages and prices on the decision to supply labour are symmetrical. A reduction of money-wages would lead to a withdrawal of labour, but a rise in the cost of living would not have the same effect, so that "within a certain range, the demand of labour is for a minimum money-wage and not for minimum real wage". Trade unions are essentially concerned with

relative wages instead of the *general* level of real wages, he claimed, especially if price movements are "small".

Keynes (p. 275) was aware of Pigou's remarks that within some limits workers actually bargain for a given money-wage instead of a given real wage.[3] However, he charged his Cambridge colleague for assuming that this would not entail any significant change for (classical) employment theory, and for sustaining that (non-frictional) unemployment is caused by money-wage rigidity when labour demand fluctuates over the business cycle. Classical economists did not seem to realize, claimed Keynes, that, if labour supply is a function of real *and* money-wages (as admitted by Pigou), classical employment theory is indeterminate. For, "unless the supply of labour is a function of real wage alone, [the classical] supply curve for labour will shift bodily with every movement of prices" (Keynes, pp. 8–9).

Champernowne (1936, p. 202) read Keynes as stating that workers are "always" more conscious of changes in money-wages than in prices, perceived as a "generalization" of Pigou's point that "sometimes" workers are concerned not only with real wages, but money-wages as well. He called it Keynes's "first wave of attack" on classical analysis. However, Champernowne criticized both Pigou's and Keynes's assumptions about money-wage determination. Just like Keynes, Champernowne referred to his observations of workers' actual behaviour, but argued that their concern with money-wages rather than real wages is only temporary, as "pointed out to me by Professor Pigou and Mr. D. H. Robertson" (p. 202, n. 1). "Conversation with a representative wage-earner" had convinced him that it would be "ridiculous" to assume that workers are more interested in their money-wage than in their real wage. The observed lagged reaction to changes in the cost of living was explained by the existence of wage contracts "based on the expectation of a stationary cost of living", transaction costs involved in contract changes, limited information about price changes, and "the habit of thinking in terms of the price level of some earlier date" (ibid.).

Hence, the money-wage rate demanded by workers "today" is the rate that would give them a certain purchasing power "at prices ruling at some date in the past" (p. 203). Prices expected by workers today are those of an earlier contractual period, as expressed in the formula $P_t^e = P_{t-1}$, implicit in Champernowne.[4] This "Keynesian" labour-supply function, as he called it, was described by $N_s(Rw)$, whereas the labour demand function was written as $N_d(R)$, where R and w are (actual) real and money-wages respectively. There is no money illusion in the labour demand function, as producers are (implicitly) assumed to form correct price expectations: $P_t^e = P_t$.

[3] "To a great extent people – employers and employed people alike – think in money. Our income is our money income, and it requires an effort to realize that, provided the price of the things we buy with money has halved, we are really no worse off with a money income that is also halved … Thus, except in periods with violent price oscillations, employers in general fight strongly against upward movements in money rates and workpeople against downward movements. Money wage-rates show themselves in practice highly resistant to change" (Pigou 1933, pp. 294–295).

[4] Champernowne probably had in mind Robertson's (1933) contractual set-up, with the division of periods in "days". See Boianovsky and Presley (2009).

This corresponds closely to Keynes's (1936, pp. 50–51) assumption that firms' short-term expectations (of demand and prices) are correct, to the extent that "expected and realized results run into and overlap one another in their influence". Such assumption of asymmetrical price expectations—which in part goes back to Marshall—plays an important role in Champernowne's employment model.

Champernowne took note of Keynes's criticism that Pigou had overlooked the shift of the labour-supply curve when prices change, but reinterpreted it in terms of *expected* prices. As Keynes (1936, p. 276) pointed out, Pigou's admission that workers demand a certain money-wage instead of a real wage means that the assumption that more labour is not available except at a higher wage, "which is fundamental to most of [Pigou's] argument, breaks down". From Champernowne's viewpoint, that was relevant because a reduction in actual real wages, caused by prices going up faster than money-wages, was associated with an *increase* in labour supply, as workers' *anticipated* real wages increased in the process. That was behind his concept of "monetary employment", defined as the excess of employment beyond its equilibrium level. Such equilibrium—dubbed "basic unemployment"—could only take place if prices had been stationary and workers accordingly demanded the "basic real wage". Symmetrically, "monetary unemployment" resulted from shifts of labour supply when the cost of living had been falling.

Champernowne's unemployment typology cannot be found in Pigou or Keynes.[5] His notion of "monetary unemployment" was "copied from Keynes's 'involuntary unemployment', but differed from that concept" (Champernowne 1936, p. 204, n. 1). While Keynes (1936, p. 15) described movements off the labour-supply curve and situations of excess supply in the labour market, Champernowne depicted points of transitory labour market equilibrium corresponding to different positions of the short-run labour-supply curve. Other young Keynesian economists, such as R. Kahn and Joan Robinson, were also critical of Keynes's treatment of full employment and involuntary unemployment at the time, although they went different ways from Champernowne (see Boianovsky 2005b).

Referring to the *General Theory* as the "Bible", Champernowne wrote to Robinson in 17 April 1936—two months before his article came out—that "as for full employment the Bible says that there is full employment if there is not involuntary unemployment, which means that you can't raise the cost of living and fool the workers into accepting a lower real wage, without causing so much strife that you end up with less men employed than before. I will not commit myself further than this in interpreting the phrase". He announced that "what I want to discuss is how long you can expect the worker to overlook a rise in the cost of living or a fall in it; in order to do this I abandon involuntary unemployment and talk about monetary unemployment, which means unemployment due to the fact that workers behave as though the cost of living were higher than it is" (Joan Robinson Papers, King's College, Cambridge; quoted from Boianovsky 2005b, p. 77).

[5] Workers' price expectations are not integrated into Keynes's or Pigou's labour-supply functions (see Young et al. 2004, p. 11, for another view of Pigou).

Keynes's notion of "full employment", as a fixed upper limit described by the absence of "involuntary unemployment", is distinct from Champernowne's idea of "basic unemployment" as a long-run equilibrium position at which workers' price expectations are confirmed and money-wages do not tend to move. His "basic unemployment" is not an upper limit, since the economy may be *above* that level if "monetary employment" prevails. As much as Keynes, Pigou (1933, 1941, 1943) too regarded full employment as a fixed limit. While arguing for the existence of a long-run tendency to full employment over the business cycle, he remarked that

> This does not, of course, imply that *on the average* full employment … exists. Since we know that employment is sometimes less than full, while it can obviously never be more than full, that would be nonsense. It means that … employment on the average falls short of full employment by a certain quantity attributable to disturbances. (Pigou 1941, p. 79)

Again differently from Pigou and Keynes, Champernowne (p. 204) sustained that periods of monetary employment (unemployment) will be accompanied by rising (falling) money-wages, as workers realize that prices are changing and repair their "oversights" accordingly. Such movements of money-wages will eventually bring the rate of unemployment to its equilibrium ("basic") value provided *real* wages move in the same direction, which brings us to the core of Champernowne's modelling of the Keynes–Pigou dispute.

3 Monetary Policy and the Trend of Real Wages

Keynes's "second wave of attack" on classical Pigouvian analysis was more persuasive than the first one, asserted Champernowne (p. 204). It consisted of the "convincing demonstration" that the effect of a change in money-wages on real wages is indeterminate, unless its (indirect) impact on aggregate demand, employment and the marginal product of labour, through ensuing movements of the rate of interest, is ascertained. Whereas the "first wave" of attack came out in Chapter 2 of the *General Theory*, the "second wave" was the theme of Chapter 19 and its appendix on Pigou's *Theory of Unemployment*. Keynes discussed how falling money-wages (and prices) might increase the real supply of money and, by that, diminish the rate of interest and encourage investment for given long-term expectations. Pigou (1933, Chapter 10), on the other hand, denied that a cut in money-wages might bring about a corresponding fall of the cost of living, leaving unemployment and real wages unchanged. Instead, he claimed that a reduction of money-wages would reduce real wages as well and bring about a movement upward along the labour demand function (see also Cottrell 1994, p. 693).[6]

[6] As mentioned above, it was only later that Pigou (1938), with some help from Champernowne, grasped the interest-rate effect of changes in money-wages. In 1943, he would introduce the famous "Pigou effect"—so named by Patinkin (1965), who preferred the more comprehensive concept of "real balance effect"—as a reaction to Alvin Hansen's "secular stagnation" hypothesis, when the interest-rate mechanism is not operative.

Instead of Keynes's argument about money excess supply—exogenously determined in the *General Theory*—Champernowne discussed changes in the interest rate by the central bank as a reaction to alterations in money-wages and prices. From Keynes's second wave of attack, "it follows that the demand of labour for a certain real wage can only make itself effective in so far as it influences the attitude of the monetary authority and its manipulation of the rate of interest" (Champernowne, p. 204). The central bank's reaction is prompted by the *acceleration* of inflation (deflation) in periods of "monetary employment" (unemployment). When the rate of unemployment is below its "basic" equilibrium level, money-wages increase and, unless real wages increase as well, prices will rise in the same proportion, with another round of rise of money-wages and so forth. The workers' "bargaining power" becomes greater and the pace of revision of money-wage demands will become faster as they get "more accustomed" (p. 205) to the effects on real wages of the rise in prices—that is, to the extent that they revise their price-level expectations in adaptive fashion.

Accelerating inflation puts pressure on monetary authorities to increase the bank rate of interest in attempt to stabilize the economy and bring the rate of unemployment to its "basic" value, as people become concerned that "there should be an inflation 'like in Germany'" (ibid.). There is some, not perfect, symmetry in the deflationary period of "monetary unemployment". The fall in money-wages and prices becomes "accelerated", but the pressure on the central bank to reduce interest rates and stabilize prices and output is not as strong, since agents' "influential opinion" is supposed to be more concerned with the danger of a hyperinflation than with the "prospect of a slump 'like they had in America'" (p. 206). This will not prevent the working of stabilization forces in the downswing, but will turn those periods longer than inflationary ones. Eventually, the economy converges to its "basic unemployment" rate—when the supply price of labour is the same as the demand price—which may be interpreted as the "trend" or long-run value of unemployment. Hence, "provided that the monetary authority does not allow labour to be misled by too long periods of rising or falling cost of living, the 'real supply curve of labour' may be a useful concept for estimating the *trend* of unemployment, real wages, rate of interest and saving" (p. 216). Champernowne apparently had in mind a Wicksellian price stabilization rule, with the bank interest rate converging to Wicksell's natural rate of interest.[7]

Assuming the working of the convergence process to "basic unemployment", classical Pigouvian analysis—of determination of *real* wages by supply and demand for labour—applied. The classical framework is relevant for the investigation of the trend value of unemployment under the assumption that on average monetary employment and unemployment even out in the long-term, but it "breaks down" when actual unemployment is considered. Trend analysis was reminiscent of the study of the classical stationary state, but with net investment going on. Champernowne (1936, p. 207) called it "dynamic equilibrium", characterized by the confirmation of workers' (and

[7]Since employment was also a matter of concern (particularly in the downturn), Champernowne's description of monetary policy is perhaps closer to the later Taylor Rule.

firms') price-level expectations.[8] Both Champernowne and Robinson (1937) were concerned with how to extend the *General Theory* to long-term equilibrium. "I think everybody is a bit puzzled about how to use Keynes's book in studying a long run", he wrote to Robinson in 2 April 1936. Champernowne argued for "a more sophisticated definition of equilibrium" as "a state of affairs in which some particular tendency has worked itself out completely". A tendency relevant for his purposes was "the tendency for there to be an expansion of credit when there is a lot of unemployment, and a contraction of it when there is a boom; then there will be equilibrium when there is just a little unemployment. This is what I think the stationary state's equilibrium ought to mean … This would of course be a moving (dynamic) equilibrium … [so that] one is likely to wobble on either side of it" (quoted from Boianovsky 2005b, p. 84).

Champernowne's interpretation that Pigou's employment theory featured a tendency to full-employment long-run equilibrium is consistent with other accounts (see, e.g., Aslanbeigui 1992, p. 431), even if the equilibration mechanism through interest-rate changes cannot be found in Pigou (1933). Champernowne (1936, p. 211) represented Pigouvian macroeconomics by a system of causal equations, starting with labour market equilibrium: $N = N_d(R) = N_s(R)$. Real wages, employment and income are determined in the labour market. Next, the allocation of output between consumption and investment is decided by the equation of supply and demand of saving: $S = S_d(r) = S_s(r)$, where r is the rate of interest. Finally, nominal variables are determined by a version of the quantity theory of money: $M = wH$, where M and H stand for money supply and real money demand (in wage units), respectively. Champernowne (p. 208) observed that money supply is exogenous only within

[8] Which could only happen if prices were stable in equilibrium, as Champernowne did not entertain the notion of perfectly anticipated inflation (or deflation) rate.

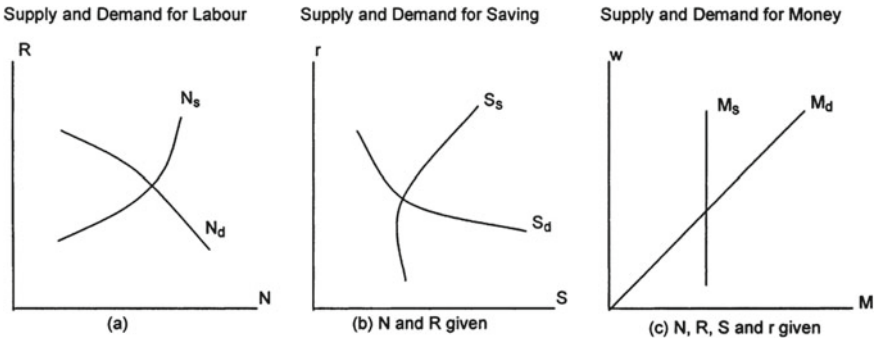

Fig. 1 Classical scheme. *Source* Champernowne (1936, p. 212)

limits, since classical trends assume that the central bank follows a (Wicksellian) stabilization policy. Hence, "in the broader sense, the rigidity of the money-wage-rate determines the price-level and the demand money determines the supply" (ibid.). Figure 1 provides a diagrammatic illustration of the classical system of equations and its sequential construction.

Champernowne's "Keynesian system", as suggested by his interpretation of the *General Theory*, tackles the determination of the same variables as the classical one (employment, real wages, real savings, interest rate, money-wages and quantity of money), but their "logical structures" (p. 209) are distinct. The starting point of Keynesian employment theory is the determination of the rate of interest in the monetary sector: $M = M_s(r) = M_d(rQ')$. The next step is to find the employment level, as decided by the supply and demand for saving (i.e., the multiplier mechanism, which is only implicit): $S = S_s(Nr) = S_d(NrQ)$. The parameters Q and Q' capture the influence on money demand and investment of "general nervousness, the state of the news and effects due to the expectation of changes in the price level, etc." (p. 211). Finally, the equations for the labour market determine the real wage and money-wage rates, as well the price level: $N = N_s(Rw) = N_d(R)$.

The sequential solution of the "Keynesian system" is, therefore, the opposite of the classical Pigouvian one, as depicted in Figs. 1 and 2, especially if "indirect effects", such as the influence of income or employment on the Keynesian money demand function, are excluded. These and other "indirect effects" should be considered in a comprehensive general-equilibrium representation of both systems, but that would blur the distinction between Pigouvian and Keynesian macroeconomics, as Champernowne (p. 211) pointed out. Unlike classical economics, the "Keynesian system" is able to account for "monetary employment" and "monetary unemployment", even if those concepts—and the notion of workers' adaptive expectations on which they are based—are, strictly speaking, alien to the framework of the *General Theory*. Moreover, the key issue, from Champernowne's standpoint, was the logical causal structure of Keynes's system.

Whereas Pigou took the determination of real wages in the labour market as the starting point to find the (trend) employment and output levels, the *General Theory*

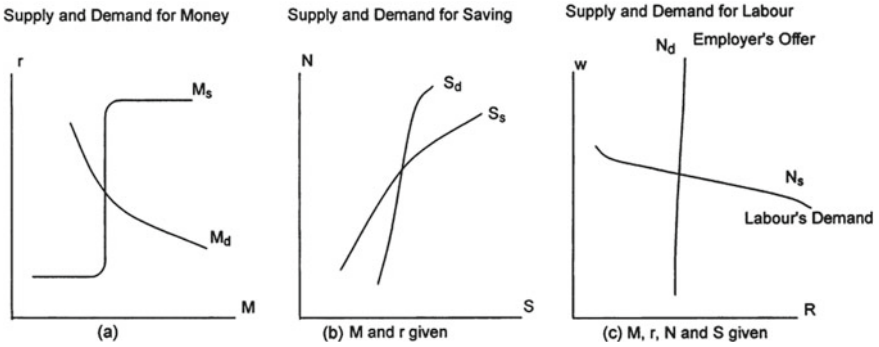

Fig. 2 Keynesian scheme. *Source* Champernowne (1936, p. 213)

started from effective demand in order to determine employment, output and real wages. In both systems, real wages are equal to the marginal product of labour (the "first postulate" of classical economics, which Keynes accepted), but the logic is essentially distinct, as stressed by Pasinetti (1974), Ambrosi (2013) and others after Champernowne. In fact, as pointed out by Skidelski (1992, pp. 575, 603–604), by working out the consequences of Keynes's acceptance of the marginal productivity theory of wages, Champernowne became the "first 'Keynesian' to emphasize that the possibility of increasing employment by demand expansion depended heavily on workers not asking for higher wages as prices rose". But Champernowne's "Keynesian system" was not just about the labour market, as already suggested above and further elaborated in the next section.

4 Keynes Versus Pigou on the Role of Expectations

Pigou's classical "trend" approach is valid provided monetary phenomena such as unemployment or extra employment due to changes in prices are temporary, so that workers' demand for a certain standard of living asserts itself. The Keynesian system would then apply strictly to the short-run only. However,

> This [classical] method will be of no avail if outlets for investment are so scarce or if the employers are so nervous of any increase in the supply of money that they hoard, and it is impossible to lower the rate of interest sufficiently to cause sufficient investment to keep prices and money-wages from falling. (Champernowne 1936, p. 216)

Persistent "monetary unemployment" was, therefore, associated with the variables Q and Q' in the functions of liquidity preference and investment demand reproduced above. This reflected Champernowne's attendance of Keynes's lectures in the 1933 Michaelmas term, when the latter deployed a formalization of the general theory of employment by a system of simultaneous equations featuring the "state of the news" (W) in the money demand and aggregate demand functions (Rymes 1989,

pp. 122–128; Dimand 2007, pp. 85–88). Keynes neither reproduced those equations nor used the term "state of the news" in the *General Theory*, although he did refer to the effects of "a change in the news" on liquidity preference and investment demand functions (Keynes 1936, pp. 155, 198). Uncertainty (or "nervousness", as written by Champernowne) may be interpreted as the inverse of the "weight of argument", a concept originated from Keynes's 1921 *Treatise on Probability* (see also Brady 2017a).

That was a key aspect of Champernowne's 1936 modelling of the Pigou–Keynes debate. As he would recall in correspondence of November 1985 with Warren Young, while comparing his model to Hicks's 1937 IS-LM, "I was aiming to elucidate the relation of Keynes's new model with the Marshall-Pigou-Robertson type of model … My emphasis was on the factors Q and Q' whose changes would shift the curves" (Young 1987, p. 85). Indeed, Champernowne (1936) was the only review article of the *General Theory* at the time that stressed expectations as one of its major features and integrated them into the equations (Boianovsky 2005b; Brady 2017b). Patinkin (1990, pp. 212–213) inaccurately listed Champernowne with other reviewers (Hicks, Lange, Lerner, Hansen, Harrod, Reddaway, Meade) who excluded uncertainty from the main components of Keynes's 1936 book. In any event, as observed by Patinkin (pp. 217–219), it was only after George Shackle's articles and books in the 1960s (e.g. Shackle 1967) that the interpretation of the "central message" of the *General Theory* in terms of uncertainty and expectations started to gain some assent. That was also the time when Champernowne (1964) produced his second (sometimes critical) reading of Keynes, which elaborated on the variables Q and Q', as well as on other "links between the economic future and the present", such as "marginal user cost".

Champernowne (1969, vol. 3, p. 80) was aware that Pigou and other Cambridge economists had ascribed business cycles to waves of optimism and pessimism. He contended, however, that it was not until the *General Theory* "that a *clear* account was given" of the effect of expectations on expenditure decisions and of how "a minor change in the 'state of the news' can play havoc" with the stock exchange and capital investment. Keynes (1936, p. 278) acknowledged that Pigou "speaks, it is true, of fluctuations in the state of demand, much as I do". However, when he came to formalize that notion, Pigou (1933) expressed it in terms of his "real demand function for labour", which was quite far from Keynes's notion of fluctuations in aggregate demand (ibid.). Pigou's complex real demand for labour function assumed a two-sector economy with a wage-goods and a non-wage-goods sector. Its key feature is that it is employment in the wage-goods sector, determined in reference to a given real wage in terms of consumption goods, that is decisive, with the investment-goods sector adjusted to absorb the rest of labour supply. This is the opposite of the priority of investment demand in Keynes's framework. Hence, Pigou's labour demand function is essentially stable and unable to account for employment fluctuations over the business cycle (see Cottrell 1994; Keynes 1936, pp. 278–279). From that perspective, Champernowne's (1936) interpretation—that (long-term) expectations play no active role in the classical analysis of the employment trend—seems warranted:

> The classical analysis can only take account of the forces Q and Q' considered in the Keynesian scheme by superimposing their effects on an equilibrium already found … It is only in such a situation, if such can exist, where only basic unemployment matters, and where uncertainty and nervousness are not very important, that the classical analysis has the advantage over the Keynesian. (Champernowne 1936, pp. 212, 216)

As mentioned above, Champernowne (1936, p. 211) included "expectations of changes in the price level" as one of the influences captured by the variables Q and Q'. Clearly, this refers primarily to the effect of businessmen's price-level expectations on their investment decisions. The expectation of higher prices raises the schedule of the marginal efficiency of capital. "This is the factor through which the expectation of changes in the value of money influences the volume of current output", claimed Keynes (1936, p. 141). Workers' price expectations are not mentioned in the *General Theory*, let alone their influence on output. Likewise, Pigou's (1933, pp. 241–243) chapter on "Reactions via expectations of price movements", based on Fisher's (1896) notion of asymmetrical expectations between lenders and borrowers in the credit market, did not refer to workers' price anticipations. However, as it is well known, Keynes (pp. 142–143) rejected on confused grounds Fisher's hypothesis about the effects of price-level anticipations on the nominal and real rates of interest. Champernowne (1964, pp. 199–200) would criticize Keynes for failing to "appreciate the truth of what Professor Fisher had in fact said" and for suggesting "a 're-writing' of Fisher's theory which is quite nonsensical".

The Fisherian distinction between nominal and real interest rates, however, was not part of Champernowne's 1936 model, possibly because of its dismissal by Keynes.[9] The Wicksellian concept of market and natural interest rates, on the other hand, is implicit in Champernowne's argument, as discussed above. Price expectations, of course, played a key role in Wicksell's ([1898] 1936) cumulative process of price change, especially (but not only) by businessmen (see Boianovsky 1998). Wicksell made clear that if economic agents start expecting changes in the price level, this will have to be taken into account by central banks in their bank rate policy. In particular, if the expected rate of deflation surpasses in absolute value the height of the natural rate (calculated in commodities), monetary stabilization policy may face a zero lower bound problem (see Boianovsky 2013). Such aspects are absent from Champernowne's (1936) discussion, though.

[9] In his later discussion of Chapter 19 of the *General Theory*, Champernowne (1964, p. 190) took note of Keynes's (1936, p. 265) remark about the perverse effects of falling money-wage rates if *expectations* of further falls in money-wages and prices (say, by 2%) arise, which Keynes compared to the effects of an increase of 2% in the interest rate in the same period. Of course, this is close to Fisher's distinction between nominal and real interest rates, which Keynes dismisses elsewhere in the book. Champernowne's (1936, p. 211) reference to the influence of price-level expectations on Q and Q' should be seen in that context.

5 The Champernowne Puzzle

Champernowne (1959, pp. 263–264) put Pigou together with Keynes and Marshall as the main names in the history of Cambridge economists. Pigou was Marshall's pupil and Keynes's colleague and co-protagonist in a "one sided controversy". However, he found Pigou's training in mathematics inferior to that of Keynes and Marshall. Pigou "lacked Marshall's passionate concern with practical and human problems" and Keynes's "brilliance and intuitive sense for pick out the key relations in an economic situation". The strength of Pigou, according to Champernowne (p. 264), consisted of his "sure grasp of logical relations and fanatical intellectual honesty". Hence, had Pigou authored the *Principles of Economics* or the *General Theory*, "they might have been less attractive works, but there would have been far less ambiguity left for lesser economists to resolve".

Champernowne apparently regarded himself as one of this "lesser economists". His 1936 attempted solution of the Pigou–Keynes controversy remained relatively non-influential if compared to Hicks (1937), which dominated macroeconomic textbooks, especially after its restatement by Alvin Hansen in the 1950s. In correspondence, Keynes reacted positively to Hicks (1937) and other formulations of the *General Theory* as a general-equilibrium system of equations (Patinkin 1990, p. 213). There are no records, however, of Keynes's reaction to Champernowne's (1936) employment model. In April 1936, shortly after the publication of Keynes (1936), they corresponded about the argument of Chapter 17 of that book, which Champernowne criticized (see Keynes 1973b, pp. 59–66).[10] However, Champernowne's forthcoming review article, discussed in correspondence with Joan Robinson in that same month, is not mentioned. Pesaran (2004, p. 211) has suggested that Champernowne's 1936 labour-supply function was a "courageous intellectual act that struck at the heart of Keynes's argument" and "could not have helped Champernowne's academic position in Cambridge economics". True enough, Champernowne left Cambridge for the London School of Economics in 1936–38, where he worked with W. H. Beveridge, but was appointed (with both Keynes's and Pigou's support; see Bridel and Ingrao 2005) Lecturer in Statistics in 1938, a position he held until 1940, when his academic career was interrupted by the War. After the War, he became a Fellow of Nuffield College and director of the Oxford Institute of Statistics. He went back to Cambridge in 1959 as Trinity Fellow, until retirement in 1978.

Champernowne's 1936 model failed to impress classical and Keynesian economists alike. Except for his "conversion" to the interest-rate mechanism in 1938, Pigou did not show any signs of accepting Champernowne's concept of "basic" unemployment. That differed from Pigou's (1941) notion of full employment as an upper limit, deployed as well by Patinkin (1965) and others. This started to change with the entrance of the Natural Rate of Unemployment Hypothesis around 1968. Again, the striking similarities between Champernowne's basic unemployment and Friedman's

[10] He would repeat and enlarge those criticisms in his 1964 essay, pp. 194–199, especially in connection with Keynes's contention that wages are necessarily more rigid in terms of money than in terms of other assets.

(1968) natural unemployment rates, with their emphases on workers' price expectations, went unnoticed until historians of thought (Darity and Goldsmith 1995; Darity and Young 1995; Boianovsky 2005b, 2018) discussed it.

Champernowne did not react to Friedman's natural rate concept. He was busy at the time completing his 1969 trilogy on uncertainty and estimations, which tackled fundamental issues in Cambridge probability theory since Ramsey and Keynes. In fact, one cannot help wonder whether Champernowne's apparent lack of interest for Friedman's Presidential Address reflected the fact that his theoretical references came usually from Cambridge (or Oxford occasionally) academics, which he seemed to regard as his audience as well. Champernowne (1969) followed a predominantly Bayesian approach to probability and decision making under uncertainty, which did not square so well with Keynes's vies on probability. Indeed, Champernowne (1964, pp. 192–193) observed critically that the links between the present and the future in the *General Theory* are in one direction only: "Although Keynes has so much to say about the effects of expectations about the future on present economic behaviour, he seems to be not nearly so informative about the causation of these expectations". Champernowne did not feel comfortable with Keynes's treatment of expectations as exogenously given by "psychology and convention", despite their role in his 1936 assessment. Interestingly enough, that brought him closer to Pigou (1950), who acknowledged the importance of expectations in Keynes (1936), but criticized the apparent lack of explanation of how they are formed.

Labour market dynamics was apparently secondary to what Champernowne (1964, pp. 201–202) perceived as the central point of Keynesian macroeconomics, despite lack of its "clear demonstration" by Keynes: the deficiency of inter-temporal coordination between employers' plans regarding productive capacity and future spending decisions of individuals in economies without widespread futures-markets. He contrasted that with the view, adopted by Irving Fisher and others (presumably including Pigou), that each individual's decision to save now carried with it the decision when to spend on consumption goods in future dates.

Acknowledgements I would like to thank Warren Young, Rogério Arthmar, Alex Cukierman, G. C. Harcourt and (other) participants at the 4th Thomas Guggenheim Conference in the History of Economics ("Expectations: theory and applications in historical perspectives", Ben Gurion University of the Negev, 18–19 December 2017) for their helpful comments on earlier drafts. A research grant from CNPq is gratefully acknowledged.

References

Ambrosi, G. M. (2013). *Pigou and Keynes as custodians of the Cambridge School of Economics*. Unpublished paper presented at the "Pigou Mini-Conference", Robinson College, Cambridge, November 29, 2013.

Aslanbeigui, N. (1992). Pigou's inconsistencies or Keynes's misconceptions? *History of Political Economy, 24,* 413–433.

Aslanbeigui, N., & Oakes, G. (2007). The editor as a scientific revolutionary: Keynes, *The Economic Journal*, and the Pigou affair, 1936–1938. *Journal of the History of Economic Thought, 29,* 15–48.
Bateman, B. (1996). *Keynes's uncertain revolution*. Ann Arbor: University of Michigan Press.
Boianovsky, M. (1998). Wicksell on deflation in the early 1920s. *History of Political Economy, 30,* 219–275.
Boianovsky, M. (2005a). Introduction. In M. Boianovsky (Ed.), *Business cycle theory, selected texts 1860–1939, vol. VII—Saving, investment and expectations*. London: Pickering & Chatto.
Boianovsky, M. (2005b). Some Cambridge reactions to *the general theory*: David Champernowne and Joan Robinson on full employment. *Cambridge Journal of Economics, 29,* 73–98.
Boianovsky, M. (2013). Fisher and Wicksell on money: A reconstructed conversation. *European Journal of the History of Economic Thought, 20,* 206–237.
Boianovsky, M. (2017). David Gawen Champernowne (1912–2000). In R. Cord (Ed.), *The Palgrave companion to Cambridge economics* (pp. 767–793). London: Palgrave Macmillan.
Boianovsky, M. (2018). Cambridge anticipations of the natural rate hypothesis? Robertson and Champernowne revisited. *History of Economic Ideas, 26* 15–50.
Boianovsky, M., & Presley, J. (2009). The Robertson connection between the natural rates of interest and unemployment. *Structural Change and Economic Dynamics, 20,* 136–150.
Brady, M. E. (2017a). *J.M. Keynes, the "state of the news (the change in the weight of evidence)", and the weight of the evidence: From Keynes's December 1933 student lectures to chapter 12 of the general theory in February 1936*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3031314.
Brady, M. E. (2017b). *How J.M. Keynes presented the technical analysis of his IS and LM curves in the general theory in 1936: Why D.G. Champernowne got it right in 1936 and Hicks (and Harrod, Meade, Lange) got a big part wrong in 1937 and 1938*. Available at https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2956744.
Bridel, P., & Ingrao, B. (2005). Managing Cambridge economics: The correspondence between Keynes and Pigou (Chap. 5). In M. C. Marcuzzo & A. Rosselli (Eds.), *Economists in Cambridge: A study through their correspondence, 1907–1946*. London: Routledge.
Champernowne, D. G. (1936). Unemployment, basic and monetary: The classical analysis and the Keynesian. *Review of Economic Studies, 3,* 201–216.
Champernowne, D. G. (1959). Arthur Cecil Pigou (1877–1959). *Journal of the Royal Statistical Society (Series A, General), 122,* 263–265.
Champernowne, D. G. (1964). Expectations and the links between the economic future and the present. In K. Lekachman (Ed.), *Keynes's general theory: Report of three decades* (pp. 174–202).
Champernowne, D. G. (1969). *Uncertainty and estimation in economics* (3 Vols.). Edinburgh: Oliver & Boyd. San Francisco: Holden Day.
Collard, D. (1996). Pigou and modern business cycle theory. *Economic Journal, 106,* 912–924.
Cottrell, A. (1994). Keynes's appendix to chapter 19: A readers' guide. *History of Political Economy, 26,* 681–695.
Darity, W., & Goldsmith, A. (1995). Mr. Keynes, the new Keynesians, and the concept of full employment. In P. Wells (Ed.), *Post-Keynesian economic theory* (Chap. 5). Boston/Dordrecht/London: Kluwer.
Darity, W., & Young, W. (1995). IS-LM: An inquest. *History of Political Economy, 27*(1), 1–41.
Dimand, R. W. (2005). Economists and the shadow of the 'other' before 1914. *American Journal of Economics and Sociology, 64,* 827–850.
Dimand, R. W. (2007). Keynes, IS-LM, and the Marshallian tradition. *History of Political Economy, 39,* 81–95.
Fisher, I. (1896). *Appreciation and interest*. New York: Macmillan.
Friedman, M. (1968). The role of monetary policy. *American Economic Review, 58,* 1–19.
Haberler, G. (1937). *Prosperity and depression*. Geneva: League of Nations.
Harcourt, G. C. (2001). David Gawen Champernowne, 1912–2000: In appreciation. *Cambridge Journal of Economics, 25*(3), 439–442.

Hicks, J. R. (1937). Mr Keynes and the 'classics': A suggested interpretation. *Econometrica, 5,* 147–159.
Jevons, W. S. ([1871] 1965). *The theory of political economy*. New York: Kelley.
Keynes, J. M. (1936). *The general theory of employment, interest and money*. London: Macmillan. As reprinted in Keynes's *Collected Writings*, Vol. VII.
Keynes, J. M. (1971–1988). *Collected writings* (30 Vols.). London: Macmillan for the Royal Economic Society.
Keynes, J. M. (1973a). *The general theory and after, part I—Preparation* (D. E. Moggridge Ed.). Vol. XIII of Keynes's *Collected Writings*.
Keynes, J. M. (1973b). *The general theory and after, part II—Defence and development* (D. E. Moggridge Ed.). Vol. XIV of Keynes's *Collected Writings*.
Kregel, J. A. (1976). Economic methodology in the face of uncertainty: The modelling methods of Keynes and the post-Keynesians. *Economic Journal, 86,* 209–225.
Kregel, J. A. (1977). On the existence of expectations in English neoclassical economics. *Journal of Economic Literature, 15,* 495–500.
Lavington, F. (1922). *The trade cycle*. London: P.S. King & Son.
Lekachman, K. (Ed.). (1964). *Keynes's general theory: Report of three decades*. New York: St. Martin.
Marcuzzo, M. C., Naldi, N., Rosselli, A., & Sanfilippo, E. (2008). Cambridge as a *place* in economics. *History of Political Economy, 40,* 569–593.
Marshall, A. (1926). *Official papers* (J. M. Keynes Ed.). London: Macmillan.
Marshall, A., & Marshall, M. P. (1879). *The economics of industry*. London: Macmillan.
Pasinetti, L. L. (1974). *Growth and income distribution*. Cambridge: Cambridge University Press.
Patinkin, D. (1965). *Money, interest and prices* (2nd ed.). New York: Harper & Row.
Patinkin, D. (1990). On different interpretations of the *general theory*. *Journal of Monetary Economics, 26,* 205–243.
Pesaran, H. (2004). Champernowne, David Gawen (1912–2000). In D. Rutherford (Ed.), *The biographical dictionary of British economists* (Vol. I, pp. 210–214). Bristol: Thoemmes Continuum.
Pigou, A. C. (1927). *Industrial fluctuations*. London: Macmillan.
Pigou, A. C. (1933). *The theory of unemployment*. London: Macmillan.
Pigou, A. C. (1938). Money wages in relation to unemployment. *Economic Journal, 48,* 134–138.
Pigou, A. C. (1941). *Employment and equilibrium*. London: Macmillan.
Pigou, A. C. (1943). The classical stationary state. *Economic Journal, 53,* 343–351.
Pigou, A. C. (1950). *Keynes's "general theory"; A retrospective view*. London: Macmillan.
Robertson, D. H. (1933). Saving and hoarding. *Economic Journal, 43,* 399–413.
Robinson, J. (1937). *Essays in the theory of employment*. London: Macmillan.
Rymes, T. K. (1987). *Keynes's lectures 1932–35: Notes of students*. Ottawa: Department of Economics, Carleton University.
Rymes, T. K. (1989). *Keynes's lectures, 1932–35: Notes of a representative student*. Ann Arbor: University of Michigan Press.
Shackle, G. L. S. (1967). *The years of high theory: Invention and tradition in economic thought 1926–1939*. Cambridge: Cambridge University Press.
Skidelski, R. (1992). *John Maynard Keynes—A biography* (Vol. 2). London: Macmillan.
Wicksell, K. ([1898] 1936). *Interest and prices* (R. F. Kahn Trans.). London: Macmillan.
Young, W. (1987). *Interpreting Mr Keynes: The IS-LM enigma*. Cambridge: Polity Press.
Young, W., Leeson, R., & Darity, W., Jr. (2004). *Economics, economists, and expectations*. London: Routledge.

Keynes and Friedman on Expectations Mismatches During the Great Depression

Sylvie Rivot

1 Introduction

At both the analytical level and at the policy level, interest for the Great Depression has never really ceased. After World War Two, it was an explicit aim of Keynesian counter-cyclical policies to ensure that this dramatic episode would never happen again: stabilising aggregate demand through public spending measures was viewed by the Keynesians as a true necessity. In contrast, Friedman built much of his Monetarism on the basis of his understanding of the Great Depression, and especially the way the 1929 financial crisis turned into a recession extremely severe. Even New Classical economists quickly developed an interest for the Great Depression. Despite the huge figures in unemployment rates as well as dramatic falls in GDP, wages and prices, understanding this historical episode as an equilibrium phenomenon was a challenging purpose for Lucas and the other New Classical economists (Lucas and Rapping 1969, 1972; Kydland and Prescott 1982 to take but two critical examples), those who aimed to deny the existence of disequilibria at the micro- as well as the macrolevel. It goes without saying that the recent "Great Recession" (2007–2012) has been the occasion of a renewed interest for the Great Depression.

One can easily mention many theoretical frameworks aiming to explain the Great Depression while ignoring the issue of expectations. First, there is the young Hayek and Robbins, together with the "liquidationists", according to whom the Great Depression corresponds to a clearance crisis (Klausinger 1995, 2003). Along this line of explanation, there might have been disequilibria, but this excess supply was

S. Rivot (✉)
BETA, 61 avenue de la Forêt Noire CNRS, UMR 7522, 67000 Strasbourg, France
e-mail: rivot@unistra.fr

Université de Haute-Alsace, 68100 Mulhouse, France

Université de Strasbourg, Strasbourg, France

Université de Lorraine, Nancy, France

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_5

just compensating excess of opposite sign during the 20s. Another kind of explanation that does not put expectations at the core of the issue is the Fisherian theoretical framework for "debt deflation". As is well known, in Fisher (1933) the stress is led on cumulative effects of bankruptcies and deleverage effects (Pavanelli 2003). Last, one can also mention the role played by innovations: for Schumpeter (1939), the Great Depression was the conjunction of a classic depression of the Juglar type with the reversal of a Kondratiev trend (Dockès, Dal-Pont and Hageman).

The purpose of this paper is to shift the attention towards explanations for the Great Depression that have addressed the issue of expectations, namely Keynes and the late Friedman. A point that should be emphasised from the outset is that we won't focus so much on the 1929 crisis as such. What interests us most is how this crisis turned into the most severe recession ever known at that time. In what follows, we analyse first Friedman's explanation of the Great Depression as phenomenon driven by monetary mismatches. We show that the late Friedman eventually put at the core of his explanations for macroeconomic disequilibria the issue of mismatches in short-term nominal expectations. Next, we turn to Keynes. We show how in the very early 30s Keynes shifted his concern from short-term disturbances to dysfunctionings in long-run expectation. Keynes' appraisal of the Great Depression in terms of coordination failures regarding the intertemporal plans is an alternative to Friedman, an alternative that puts the stress on the indeterminacy of the structure of the economy in the remote future.

2 Friedman and the Case for Purely Monetary Disorders

In our review of Friedman's understanding of the Great Depression, we proceed in two steps. First, we first provide an outline of Friedman's account of the Great Depression, as this *narrative* appears in Chap. 7 of the *Monetary History of the United-States 1867–1960* (1963) co-authored with Schwartz. We then make use of Friedman's theoretical pieces of work to suggest an *analytical* explanation of these real disequilibria (i.e. regarding output and employment).[1]

2.1 Friedman's Narrative of the Great Depression

From the outset, let us precise that, as the title of their book indicates, Friedman and Schwartz focused their attention to the US case. They situated the 1929 episode in a long-run perspective that portrayed the late 20s as a very special period, "a period of high prosperity and stable economic growth" (Friedman and Schwartz 1963, p. 296). In their eyes, the monetary policy carried out during that period

[1]For an extensive inquiry of Friedman's theoretical framework, the reader can refer to the two-volumes book recently published by Edward Nelson, and especially Chaps. 6, 7 and 8.

was highly efficient: "the close synchronism produced much confidence within and without the System that the new monetary machinery offered a delicate yet effective means of smoothing economic fluctuations" (Friedman and Schwartz 1963, p. 296). Strikingly, the diagnosis offered was simply the opposite of the "relative inflation" considered by the Austrians:

> The economic collapse from 1929 to 1933 has produced much misunderstanding of the twenties. […] Far from being an inflationary decade, the twenties were the reverse. And the Reserve System, far from being an engine of inflation, very likely kept the money stock from rising as much as it would have if gold movements had been allowed to exert their full influence.
>
> (Friedman and Schwartz 1963, p. 298)

In the eyes of Friedman (and at the opposite to the view held by Hayek (1929) for example) the 20s appear as a "Golden Age" for the conduct of monetary policy, despite the lack of clear theoretical foundations. From 1922 to 1928, the Reserve System was greatly influenced by the banker Strong (governor of the Federal Reserve Bank of New York). At the very end of the 20s, the implicit reliance (or worse the explicit reliance in the case of the influential Miller, a senior member of the Board of Governors) upon the real bills doctrine proved tragic.

What about the 1929 financial collapse? The tightening of monetary policy that was set up for deterring the use of credit for speculation in Wall Street did also deterred the use of credit for "productive" reasons. Friedman and Schwartz insisted that the first noticeable decline in the total stock of money and in wholesale prices appeared *before* the 1929 financial collapse. No doubt that a restrictive monetary policy precipitated the collapse: the Fed "followed a policy which was too easy to break the speculative boom, yet too tight to promote healthy economic growth" (Friedman and Schwartz 1963, p. 298).

The Great Depression is for Friedman and Schwartz a direct outcome of the—unfortunate—tightening in monetary policy conducted by the Fed that turned an ordinary recession into a dramatic contraction. The money stock (i.e. currency held by the public, demand and time deposits) fell by one-third between 1929 and 1933—a fact unique in the monetary history of the USA. Even if money stock, money income and prices level roughly fell together, Friedman and Schwartz's basic claim is that causality goes from money supply to real income and prices: "the contraction is in fact a tragic testimonial to the importance of monetary forces" (Friedman and Schwartz 1963, p. 300).

To conclude this narrative, how then to explain the revival of 1933? For Friedman interest of policy-makers shifted to the view that "money does not matter", an argument they claimed to have arisen under Keynes' influence. From 1933 onwards, public authorities shifted their matter of concern from monetary policy to fiscal policy. Yet, nothing positive came from fiscal stimulus: in Friedman's eyes fiscal policy is neutral except for nominal interest rates. Paradoxically, in basically ignoring the potency of monetary policy the Fed at least stopped from causing monetary disorders: the Fed did not attempt to alter the high-powered money, either by open-market operations or by rediscounting. As a matter of principle, it cannot be excluded that

the rise in money stock was a loan-demand-induced phenomenon. That is the reason why, again, Friedman and Schwartz took pain to show that recovery came *after* the money stock started to rise, and not the other way round as a (endogenous) result of the recovery. In their analysis, the stress is led on the new monetary arrangements and in particular the large institutional reforms that occurred, although monetary policy (i.e. monetary loosening) was considered as rather impotent by policy-makers.

The *Monetary History* applies a methodology that will be later named the "narrative approach", a methodology that many considered as "the most enduring legacy" (Bordo 2016, p. 147) of that book.[2] Yet, the explanation provided for the Great Depression was criticised for its lack of theoretical background. In that book, there was no *explicit* analytical background on which to rely for getting a proper understanding of the economic forces at work in the processes studied. It was precisely this kind of critiques that gave Friedman strong incentives to sharpen his understanding of the functioning of a monetary economy and to provide theoretical pieces of work. Yet, as acknowledged by Friedman himself, there was an *implicit* theoretical framework behind the *Monetary History of the US*, namely the modern quantity theory money as restated by Friedman (1956), which corresponds to a theory of a demand for money stable in the long run. In that version of the Quantity Theory of Money:

> In the short run, changes in the money stock would produce changes in real output; in the long run, changes in the money stock would be fully reflected in changes in the price level. In modern terms, monetary changes temporarily impact real output, *reflecting nominal rigidities*, but, ultimately, the growth of real output is independent of monetary forces and monetary neutrality prevails.
>
> (Bordo 2016, p. 144; emphasis added)

This theoretical framework that sustains implicitly the transitory non-neutrality of money actually relies on nominal rigidities, lags as well as static expectations. Friedman and Schwartz's case rests on several strong assumptions, all of which having been severely disputed and criticised (Hammond 1996; Rivot 2013a). First, this interpretation supposes that the Fed *did* have a good grasp both of the severe risks of recession and of monetary channels of transmission at work in the economy. Second, the monetary multiplier is supposed to be kept constant for monetary policy to rightly operate. The overall issue regarding Friedman's interpretation of the Great Depression, and regarding business fluctuations in general, refers to what has been called the "identification problem". That is, what remains of Friedman and Schwartz's claim if the creation of money appears to be driven by demand for loans? As shown

[2]Friedman outlined this methodology in "The Methodology of Positive Economics" (1953). Identification of unique historical and institutional circumstances, he argued, provides the closest thing to a controlled experiment in which the direction of influence from money to income can be isolated. Friedman explains this methodology in terms of understanding the causes of inflation. […] Thus Friedman (1953) anticipated the main mission of [*A Monetary History of the US*]: To show that in a long series of examples, stretching across very monetary regimes, the relationships between money, prices and output, suggested by the quantity theory of money, continue to hold (Bordo 2016, p. 147).

by Hammond (1996), causality is the basic issue underpinning the whole debate,[3] a point easily acknowledged by Friedman.

Regarding monetary shocks, this explains well why Friedman later developed an argument to explain cyclical patterns in both money supply and output in his "Theoretical Framework for Monetary Analysis" (1970) when he came to accept that theoretical underpinnings were needed to address the "measurement without theory" objection held by his opponents. Indeed, even if one considers that the causality goes in the right sense (i.e. from the exogenous money supply to aggregate nominal income), there is a critical gap to be filled here, namely the issue of the "transmission mechanism". That is, what are the mechanisms at work to transmit a variation in the money supply to the aggregate nominal income? How to disentangle the effects on prices from the effects on output? And how to explain the huge rates of unemployment during the Great Depression? Regarding this critical issue, what is striking is that this relationship between employment and wages is *not* addressed in the *Monetary History of the United States* (1963). So one needs to go beyond a mere narrative of how the Great Depression developed and to shift the attention to the functioning of a monetary economy. As we will see below, when he eventually concedes the need to provide an explicit theoretical framework to his account for monetary disturbances, Friedman's attention shifted dramatically to the issue of uncertainty and the formation of private expectations.

2.2 *Expectations and Friedman's Rationale for the Great Depression*

Regarding the "transmission mechanism" of changes in money supply, one has to combine two types of reasoning in Friedman's writings to get a complete picture: first the effect of a change in money supply on nominal aggregate income, output level and price level; second, the effect of these changes on the level of employment. The analytical explanation of unemployment during the Great Depression can then be shaped in a symmetrical manner to the analysis of a monetary expansion: the starting point of the causal chain is a monetary contraction as the one undertaken by the Fed in the early 1929. As we will see below, when he addressed explicitly the issue of the "transmission mechanism", the idea of nominal rigidities is replaced by the idea of inelastic expectations: one kind of stickiness is substituted for the other.

Let us start with the effect of a rise in money supply on the aggregate income, an analysis that can be found in "The Optimum Quantity of Money" (Friedman 1969), "A Theoretical Framework for Monetary Analysis" (Friedman 1970) and the *Monetary Trends* (Friedman and Schwartz 1982). When the rate of growth of money increases, the rate of inflation will increase at the same pace in the long run: the Quantity

[3]Friedman seems to have become aware of the possible futility of giving empirical evidence a primary role in demonstrating causation to his fellow economists soon after he and Schwartz began presenting their results (Hammond 1996, p. 210).

Theory of Money will do well. But in the short run, things are more complicated. What happens if people take time to understand what is happening and do not revise their inflationary expectations immediately? In the case of adaptive expectations, individuals are not necessarily able to disentangle at the very beginning of the process a transitory rise in the money growth rate from a permanent one. The real rate of interest being constant, the perceived cost of holding money is not fully adjusted and the current balances exceed their long-term desired level. That is, the nominal interest rate does not rise as much as in the case where inflationary expectations are revised without delay. Once the rise in the money supply is perceived as permanent, expectations are revised, which leads to a higher nominal interest rate. The desired cash balances fall because of the increased demand for goods. There is no reason for the process to be smooth, so that cyclical patterns are possible.

The first key point to be emphasised here is that the decision-making process applies to the holding of money. That is, people buy goods because they hold the cash balances they wish to (i.e. their cash constraint is saturated). Second, the whole analysis relies on the hypothesis of adaptive expectations. Considering rational expectations would imply an instantaneous adjustment to the new rate of growth in the money supply.

Next, what is the effect of this rise in aggregate demand on employment and output? From his 1967 AEA lecture "Role of Monetary Policy" (published in 1968 in the AER) to the *Monetary Trends* (1982), Friedman offered several explanations of this second part of the argument regarding the "transmission mechanism" of a monetary expansion. Yet, some contributions have to be disregarded at the benefit of others if one aims to get a proper idea of Friedman's "definitive" views on this issue.

Regarding first Friedman's "Role of Monetary Policy" (1968), the theoretical argument developed in that paper has been proved rather inconsistent and misleading (De Vroey 2001; Laidler 2012; Forder 2018; Rivot 2018). Consequently, in what follows we will ignore that paper. Another contribution that we will ignore below is Friedman's essay dated 1975 entitled *Unemployment versus Inflation*. Indeed, it is in that essay that can be found Friedman's most extreme views regarding the expectations issue. Adaptive expectations are presented there as meaning that "the anticipated rate of inflation is an exponentially weighted average of past rates of inflation, the weights declining as one goes back in time" (Friedman 1975, p. 25). And Friedman's sympathetic assessment of rational expectations in the 1975 paper is rare enough to be worthy of note. In offering his critique of the adaptive expectations hypothesis, Friedman pointed out that economic agents are fundamentally forward-looking and that they do not build their anticipations "on the basis of a weighted average of past experience with fixed weights" (Friedman 1975, p. 26).[4] The rational expectations hypothesis is then acknowledged as follows:

[4]But people who are forming anticipations are not fools – or at least some of them are not. They are not going to base their anticipations solely on the past history of prices. Is there anybody in this room whose anticipation of inflation next year will be independent of the result of the coming British election? (Friedman 1975, p. 27).

> Therefore, said Muth, we should assume that people form their anticipations on the basis of a correct economic theory: not that they are right in each individual case but that over any long period they will on the average be right. Sometimes this will lead to the formation of anticipations on the basis of adaptive expectations, but by no means always. (Friedman 1975, pp. 27–8)

Yet, sticking to this very sympathetic assessment of the rational expectations approach to macroeconomic dynamics would entail a severe misperception of Friedman's overall understanding of the functioning of a monetary economy as well as his overall policy advices. That is, in this New Classical version of the Phillips curve with rational expectations and imperfect information (such as in Lucas 1972), the forward-looking Phillips curve is non-vertical in the short run only because of unanticipated changes in money supply (i.e. "surprises"), and not because of delays in adaptation to a new monetary regime, even in the case the change in money supply is announced (Rivot 2017). As is well known, with rational expectations monetary expansion is not only neutral but also super-neutral. Noticeably, one cannot find after 1975 an assessment of the rational approach as sympathetic as in the essay dated 1975, quite the contrary. In a section of the *Monetary Trends* (1982) precisely dedicated to rational expectations, one can find the following statement:

> The formalization in the theory of rational expectations of the ancient idea that economic actors use available information intelligently in judging future possibilities is an important and valuable development. But it is not the open sesame to unravelling the riddle of dynamic change that some of its more enthusiastic proponents make it out to be. (Friedman and Schwartz 1982, p. 630)

If one searches for a representative review of Friedman's appraisal of the transmission mechanism from changes in aggregate demand to changes in output and employment, and if one seeks to apply this appraisal to Friedman's assessment of the Great Depression, one has to focus on Friedman's writings after 1975. That is the reason why we concentrate below on Friedman's 1976 "Nobel" lecture (published in 1977) and on the *Monetary Trends* (1982). The quotes from Friedman's "Inflation and Unemployment" (1977) that is crucial for our assessment run as follows:

> … we developed an alternative hypothesis that distinguished between the short-run and the long-run effects of *unanticipated* changes in aggregate nominal demand. Start from some initial stable position and let there be, for example, an *unanticipated* acceleration of aggregate nominal demand. This will come to each producer as an *unexpectedly* favorable demand for his product. In an environment in which changes are always occurring in the relative demand for different goods, he will *not know* whether this change is special to him or pervasive. It will be *rational* for him to interpret it as at least partly special and to react to it, by seeking to produce more to sell at what he now *perceives* to be a higher than expected market price for future output.
>
> … Both workers and their employers are likely to adjust more slowly their perception of prices in general – because *it is more costly to acquire information* about that – than their perception of the price of the particular good they produce. (Friedman [1977] (1987), pp. 352–53; emphasis added)

What the quotations above show is that individuals take time to understand events, to move from an uncertain context to a risky context, i.e. to gain critical knowledge

about the probability distribution they face. They really need time to adapt to a new environment, for example to disentangle a change in relative prices from a scalar effect.

Regarding monetary expansion, the argument runs as follows. Monetary expansion leads to a rise in prices that employers wrongly consider as applying to their own activity (i.e. they confuse a rise in the general price level with a modification in relative prices), so that they want to produce more and hire more labour accordingly. Firms do increase their supply price because of an increased marginal cost (including higher money-wages necessary to cope with the hiring of additional labour). As for them, workers wrongly perceive the rise in money-wages as a rise in real-wages (which means that they perceive the increase of the general price level with delay), so that they offer more labour. There is a crucial hypothesis made here that a perceived increased real-wage is necessary for the labour force to work more. The—necessarily temporary—macroeconomic disequilibrium is rendered possible only by mischiefs in price perceptions. As soon as new prices (of goods, labour and so on) are incorporated in individual expectations, we are back to the initial position, i.e. to the natural rate of unemployment—the general price level excepted. Here, macroeconomic disequilibrium is consistent with microeconomic equilibrium: people are on their curves while misperceiving current prices and wages.

To conclude our review of Friedman's appraisal of the Great Depression, let us apply his explanation of the dynamics around the natural rate of unemployment (i.e. the argument of his "Nobel" lecture) to get a rational of "abnormal" unemployment during the Great Depression. Monetary tightening in 1929 lowered the nominal aggregate demand and thus the general price level as well as the output level. In Friedman's monetary framework, the aggregate supply curve is not vertical in the short run because of price misperceptions that allow for variations in the output level. This means that private actors reacted to the contraction in nominal aggregate demand in a twofold way. Employers wrongly perceived the fall in price as a relative decline in the price of the good they produced and decided accordingly to produce less. They dismissed workforce because of the increase in the perceived real-wage they respectively faced (i.e. the real-wage paid in terms of the product price they sold). A symmetrical argument applies to the labour market: the fall in the general price level led to a fall in money-wages, but workers wrongly perceived the money-wage cut as a real-wage cut (which means that they perceive the fall of the general price level with delay), so that they offered less labour. Were employers and workers in capacity to correct their misperceptions and to adapt accordingly their inflationary expectations to the new level of money supply, the macroeconomic disequilibrium would have disappeared. In Modigliani's (1977) words: "output falls, not because of the decline in demand, but because of the entirely voluntary reduction in the supply of labour, in response of erroneous perceptions" (Modigliani 1977, p. 4).

How can we then explain the huge rates of unemployment attained during the Great Depression? In Friedman's own words in his Chap. 12 of *Price Theory* ([1976] 2008) dedicated to the labour market functioning, "the answer is not entirely clear" (Friedman [1976] 2008, p. 236). Friedman acknowledged that "there was a series of sharp, unanticipated declines in aggregate demand, so that the recurrent and even bigger

readjustments in anticipations were required" (Friedman ([1976] 2008, p. 236). The rationale behind Friedman's implicit analysis of the labour market is the following: since it pursued its monetary tightening from 1929 to 1933, the Fed did not give private actors a chance to correct their wage and price misperceptions. From 1933 onwards, the attention of policy-makers shifted towards fiscal stimulus. Paradoxically, the fact that the monetary policy potency was roughly ignored from that time on allowed monetary disorders to disappear. Price expectations adapted to the new level of money supply growth, and the economy went at last back to the natural rate of unemployment as defined by the long-run vertical Phillips curve.

To sum up our argument regarding Friedman, we have shown that the *Monetary History of the US* implicitly relied on Friedman's (1956) restatement of the Quantity Theory of Money based on nominal rigidities. When Friedman finally surrendered to the criticisms of his opponents who compelled him to provide an explicit theoretical background, and certainly under the influence of the rational expectations approach to macroeconomics that was launched precisely at that time, Friedman progressively shifted his attention towards the issue of short-term mismatches in the forecasting process. In the late Friedman, adaptive expectations are at the core of the explanation for macroeconomic disequilibria.

3 Keynes and the Case for Self-feeding Real Disorders

Keynes lived the Great Depression as a contemporary so one can follow virtually day after day how he grasped the developing of the 1930 recession and how he situated himself by comparison with his contemporaries. As it was the case with Friedman, we focus on the expectations issue. Precisely, we try to show that in the developing of the Great Depression, the conversion of a standard (although rather severe) recession into the worst economic depression led Keynes to shift his matters of concern and to build his case for an economics of depression. In the meantime, Keynes shifted his matter of concern from deleverage effects and mismatches in short-term expectations to an interest for long-term expectations and less-than-full-employment equilibrium.

3.1 1929–30 and Keynes' Initial Diagnosis for an "Abnormal" Slump

Let us start with Keynes' appraisal of the economic climate that prevailed before the slump. As Friedman, Keynes did not consider the 20s in the USA as an inflationary period. Regarding the causes of the crash on the New York stock exchange, Keynes held an interpretation very close to Friedman's. For Keynes too, the Fed unfortunately deterred credit for "productive" investment when the dear money policy was implemented with the purpose of curbing speculation on financial markets.

At the very beginning of the slump, Keynes appeared rather optimistic. For example, he wrote to the Philips Electronics firm: "there are a fair number of indications that we may be somewhere in the neighbourhood of the bottom point" (letter to the Philips company, April 1930, quoted in Dimand 2016). In particular, "the continuance of cheap money, and even more the *expectation* of such continuance, is bound to be effective in the situation in the course of a few months" (id.; emphasis added). Here is to be found a dominating aspect of Keynes' overall understanding of the efficiency of economic policy: an economic policy, and especially monetary policy, is all the more efficient when it modifies the expectations of private agents regarding the economic climate in the periods ahead, including of course expectations about the economic policy carried out by public authorities occurring in future (Rivot 2013b). In the early 1930, Keynes considered that monetary loosening would be all the more efficient if private actors were convinced that this policy would be pursued for as long as needed.

Now, Keynes quickly changed his mind about the depth of the recession that followed the financial crash.[5] At the Economic Advisory Council in July 1930, he argued that the recession would be probably much more severe than expected at that time. What appeared in Keynes's reasoning were the mechanisms of secondary deflation because of deleverage effects that worsen the initial recession. This kind of argument is in line with the reasoning developed by Keynes in the *Treatise on Money* stated, in which he argued:

> Thus I attribute the slump of 1930 primarily to the deterrent effects on investment on the long period of dear money which preceded the stock-market collapse, and only secondarily to the collapse itself. But the collapse having occurred, it greatly aggravated matters, especially in the United States, by causing a disinvestment in working capital. Moreover, it also promoted the development of a profit deflation in two other ways – both by discouraging investment and by encouraging saving. The *pessimism* and the *atmosphere of disappointment* which the stock-market collapse engendered reduced enterprise and lowered the natural rate of interest; whilst the '*psychological*' *poverty* which the collapse of paper values brought with it probably increased saving.
>
> (Keynes 1930, CW 6, p. 176; emphasis added)

At the global level, Keynes acknowledged the existence of an "abnormal" depression. From that moment on, he proved virtually systematically more pessimistic than contemporary commentators (Dimand 2016).[6] Keynes started also very soon to disentangle the case of the USA from the case of Great Britain. The USA had

[5] A the present time, as compared with, say, eight months ago, I should put in the forefront of the causes of our industrial position the worldwide international slump in prices, profits, volume of trade, employment and the output of enterprise. The trade slump which has developed since the Wall Street crash last autumn is amongst the most important which have ever occurred in modern economic history. Both the business world and the outside public are, I think, very much underestimating the magnitude of this factor (Keynes 1930, CW 20, p. 370).

[6] It is now fully clear the world is in the middle of an international cyclical depression of unusual severity… a depression and a crisis of major dimensions. I believe that the prevailing opinion in the United States is still not pessimistic enough and is relying too much on a recovery in the early autumn, an event which is, in my opinion most improbable. Nothing is more difficult than to predict the date of recovery (letter to the Philips company, July 1930, quoted in Dimand 2016).

experienced a long period of stable growth during the 20s, whereas Great Britain already experienced stagnation during that period. This led Keynes to argue in one of his statements for the Economic Advisory Council:

> Even if we had had no troubles whatever of our own, we should at this moment be suffering severely. Coming on top of our previous troubles, this has given people the *impression* of its being a further stage in a definite downward progression, instead of its being the downward movement of a swing – which I am quite sure it really is – to be followed in due course by an upward reaction. It might be possible, and would be serviceable, to *try to make this vivid to the public*.
>
> (Keynes July 1930, CW 20, p. 370; emphasis added)

The idea stated above is stated in terms of the economic climate prevailing at the time the economy is hit by a deflationary shock. Yet, the argument foreshadows that would later be made about the long-term expectations regarding the economic prospects in the remote future. The point is that expectations were probably not so pessimistic in the USA in comparison to Great Britain when the crisis appeared, so that deflationary expectations were probably not so anchored despite the violence of the shock.

In December 1930, Keynes published in two papers his essay entitled *The Great Slump of 1930* (Keynes 1930, CW 9, pp. 126–34), in which he acknowledged "the extreme violence of the slump" (Keynes 1930, CW 9, p. 127). On can observe in that essay the practical implementation of the theoretical insights established in the *Treatise on Money* to provide an explanation for the slump. First, there is unemployment because "industrialists do not expect to be able to sell without loss" (Keynes 1930, CW 9, p. 129). Second, this disincentive to hire workers and to produce appeared because prices had fallen more than costs. The third argument is to deny the relevance of Says' law, even if the wording is not used.[7] Still in line with the two fundamental equations model of the *Treatise on Money*, the fourth argument is that the discrepancy between sales and costs forecastings is eventually explained by the discrepancy between savings and investment. This corresponds to the existence of negative "windfall profits" (or losses) due to mismatches in sale proceeds. Since short-term forecasting errors in sales proceeds imply a discrepancy between saving and investment, the fifth argument foreshadows what will be later on the core argument in Keynes' explanation for the Great Depression, namely the non-profitability of the capital goods market. On the capital market profits depend on "whether the public prefer to keep their savings liquid in the shape of money or its equivalent or to use them to buy capital goods or the equivalent" (Keynes 1930, CW 9, p. 130), what will become the inadequacy of liquidity preference with the marginal efficiency of capital.

> By this time a vicious will be set up and, as a result of a series of actions and reactions, matters will get worse and worse until something happens to turn the tide.

[7] It is not true that what the business men pay out as costs of production necessarily comes back to them as the sale proceeds of what they produce. It is the characteristic of a boom that their sale proceeds exceed their costs; and it is the characteristic of a slump that their costs exceed their sale proceeds (Keynes 1930, CW 9, p. 130).

> […] If, then I am right, the fundamental cause of the trouble is the lack of new enterprise due to an unsatisfactory market for capital investment.
>
> (Keynes 1930, CW 9, p. 131)

So at that time, Keynes was mainly interested in deleverage effects due to bankruptcies and mismatches in short-term expectations (i.e. expectations regarding sales proceeds). In case of "the obstinate persistence of a slump" (Keynes 1930, CW, p. 332), the Keynes of the *Treatise* would advocate "the purchase of securities by the central bank until the long-term market rate of interest has been brought down to the limiting point" (Keynes 1930, CW 6, p. 332), which corresponds to open-market operations "*à outrance*" (Keynes 1930, CW 6, p. 331). The Friedmanian tone of this statement is noticeable. But because of the special case experienced by Great Britain, Keynes progressively came to shift his mind about monetary loosening and started to put forward again his case for public works, a case that he already made in the pamphlet *Can Llyod George Do It?* (1929) co-authored with Henderson (Keynes 1929, CW 9, pp. 86–125). That is, if in normal circumstances "the stimulating effect of cheap money and abundant credit on the new issue market is exceedingly rapid" (Keynes May 1930, CW 20, p. 346), such a policy would be inefficient in special circumstances: "when enterprise and confidence have collapsed to the extent that they have today, the response to what would have been in other circumstances a strong stimulus may be very reluctant" (Keynes May 1930, CW 20, p. 346).

3.2 1931–32 and the Developing of the Slump

In June 1931, Keynes moved to the USA to give lectures at the Harris Foundation of Chicago. These lectures were published as *Unemployment as a World Problem* (Keynes, CW 13, pp. 343–67). Keynes started with a dramatic tone, the first sentence being: "we are today in the middle of the greatest economic catastrophe—the greatest catastrophe due almost entirely to economic causes—of the modern world" (Keynes, CW 13, p. 343).

Keynes insisted on the severe decline in investment as the fundamental explanation of what was happening. At the occasion of a round-table organised by the Harris Foundation, Keynes took the opportunity to precise his mind:

> There is a certain point where almost everybody in charge of funds reaches the stage of what I call '*abnormal psychology'*. In an ordinary way, any kind of financial institution has a certain cushion of some kind, reserves and margins, and is prepared to run *reasonable risks*, prepared to be *sensible on the evidence*, but when those margins run down to a certain point they get into a state of mind where they are not prepared to run even an actuarial risk or better than an actuarial risk. *They just won't run any risk at all*, because they have got to the end of their margins.
>
> […] That *morbid psychology*, though quite intelligible and natural, is a tremendous obstacle to a right development of affairs when it exists. […] Willingness to run sound risks is the (only) possible basis for all progress, and in order to restore a normal state of mind, to get rid of the abnormal psychology, it may be quite right to use methods of comforting which

> would be unsound in any ordinary conditions. (Keynes 1931, CW 20, pp. 536–37; emphasis added)

Accordingly, the big issue was to restore the profitability of business.[8] The theoretical underpinnings of this appraisal are still the ones of the *Treatise on Money*, namely the discrepancy between savings and investment: "[…] it is out of the disequilibriums between savings and investment, and out of nothing else, that the fluctuations of profits, of output, and of employment are generated" (Keynes 1931, CW 13, p. 355). And Keynes relied on the existence of equilibrating forces "for expecting the decline to reach a stopping-point" (Keynes 1931, CW 13, p. 356), through the variations in the level of output and changes in the rate of savings.

In the lecture for the Harris Foundation called "The Road to Recovery", Keynes identified two critical means needed for the recovery: first "a return of confidence to the business world so as to incline them to borrow on the basis of normal expectations of the future"; second "a drastic fall in the long-term rate of interest so that full advantage may be taken of any recovery of confidence" (Keynes 1931, CW 13, pp. 358–59), confidence being required for both lenders and borrowers for investment to rise again. A third means, which is a *consequence* of the first two arguments (and not a preliminary), is the "restoration of prices to a higher level" (Keynes 1931, CW 13, p. 359). In these lectures, one can find also Keynes' opposition to wage cuts, with the main argument of the "impossibility" to implement the required wage cuts. Another argument that still belonged to the theoretical framework of the *Treatise on Money* is the "burden of monetary indebtedness" (Keynes 1931, CW 13, p. 355).[9] Noticeably, public works (which Keynes called "new construction programmes under the direct auspices of the government or other public authorities" (Keynes 1931, CW 13, p. 355) are presented as a way to restore confidence, to "break the vicious circle" for "a government programme is calculated to improve the level of business profits and hence to increase the likelihood of private enterprise again lifting up its head" (Keynes 1931, CW 13, p. 364). Noticeably, the argument here is not stated in terms of a direct effect on aggregate demand: there is no reasoning in terms of effective demand.

At the end of his journey, Keynes sent a letter to Henderson, in which he acknowledged that a factor that he had "most underestimated before [he] came was the position of many banks in the country" (Keynes 1931, CW 20, p. 555). And when he went back to Great Britain, he noticed in a memorandum for the Economic Advisory Council the "many banks and of many depositors" as well as an "absolute mania

[8] This is my secret, the clue to the scientific explanation of booms and slumps (and of much else, as I should claim) which I offer you. For you perceive that when the rate of current investment increases (without a corresponding change in the rate of savings) business profits increase. Moreover, the affair is cumulative. For when business profits are high, the financial machine facilitates increased orders for and purchases of capital goods, that is, it stimulates investment still further; which means that business profits are still greater; and so on (Keynes 1931, CW 13, p. 354).

[9] For the burden of monetary indebtedness in the world is already so heavy that any material addition would render it intolerable (Keynes 1931, CW 13, p. 361).

for liquidity" (Keynes 1931, CW 20, p. 568; emphasis added).[10] This worsening of the lending market has catastrophic implications for the new construction industry, which depends crucially on the terms of the loan and which is also considered as a leading sector that triggers many other industries. That is the reason why the recovery of confidence became a matter of huge concern for Keynes. But the recovery of confidence he is concerned with does not only apply to the monetary and financial side of the economy, but also at a more general level to "a return of confidence to the business world" (Keynes 1931, CW 13, p. 358).

The next episode reflected vividly the international interdependencies of the countries affected by the crisis. In August 1931, Great Britain went under the pressure of a speculative attack against the Sterling. Great Britain left the Gold Exchange Standard in September 1931, which significantly aggravated the recession, through outflows of gold and subsequent disorders on currency markets. At that time, a deflationary policy, which implied cuts in prices and wages as well as a deflationary budget (what Keynes called an "economy budget"), was called for allowing Great Britain to go back to the gold standard at the previous parity. As is well known, Keynes strongly opposed this kind of deflationary remedies as involved by an economy budget: "in some directions its immediate effects will be to make matters worse and it may be expected (unless it is accompanied by a tariff) to aggravate unemployment" (Keynes 1931, CW 20, p. 597).

During 1932, the economic crisis continued to develop, stock exchange indexes to decline, costs and prices to contract and unemployment rates to worsen all over the developed countries. In a lecture given in Hamburg in January 1932 and called "The economic Prospects" (Keynes 1932, CW 21, pp. 39–48), Keynes noticed the fall of the money value of almost every kind of assets. For him, there was at that time "a competitive panic to get liquid" (Keynes 1932, CW 21, p. 40), for individuals of course, but also for institutions and governments. At that time, Keynes appeared extremely pessimistic regarding the economic prospects.[11] He also complained about deflationary policies as well as beggar-my-neighbour policies that were implemented at that time:

---

[10]The *anxiety* of many banks and of many depositors throughout [the US] is a dominating factor, the importance of which I had not fully estimated before visiting the United States. It is, I think, one of the biggest obstacles overhanging the situation, in the way of the normal process of recovery.

A considerable number of the member banks and a fairly substantial proportion, measured in assets (perhaps as much as 10 per cent), are probably not solvent today, if their assets were to be valued strictly.

[…] The inevitable result is an *absolute mania for liquidity* wherever liquidity is possible.

[…] This atmosphere affects perfectly good banks as well as the bad ones." (Keynes 1931, CW 20, p. 568; emphasis added).

[11]One can see him, for example, arguing in February 1932: "The immediate problem for which the world needs a solution today is different from the problem of a year ago. Then it was a question of how we could lift ourselves out of the state of acute slump into which we had fallen, and raise the volume of production and employment back towards a normal figure. But today the primary problem is to avoid a far-reaching financial crisis. […] The restoration of industry must come second in order of time." (Keynes 1932, CW 21, pp. 50–1).

> Practically all the remedies popularly advocated today are of this internecine character. Competitive wage reduction, competitive tariffs, competitive liquidation of foreign assets, competitive currency deflations, competitive economy campaigns, competitive contractions of new developments—all are of this beggar-my-neighbour description. The modern capitalist is a fair-weather sailor. As soon as a storm rises he abandons the duties of navigation and even sinks the boats which might carry him to safety by his haste to push his neighbours off and himself in. (Keynes 1932, CW 21, pp. 52–3)

During the second part of 1932, Keynes started to consider other kinds of disruptions that added to the initial monetary disturbances. That is, he started to consider real disturbances of a self-feeding kind, each fall in investment and aggregate demand preparing and encouraging the next one, precisely because of self-realising pessimistic expectations. Keynes also continued to strongly condemn the implementation on both sides of the Atlantic Ocean of deflationary policies, which eventually did nothing but aggravated the economic depression. Accordingly, Keynes shifted his policy advices towards the erection of a National Investment Board, which would work in close cooperation with the Bank of England. But this shift of interest for a direct leverage of aggregate demand undoubtedly called for renewed theoretical underpinnings of the ups-and-downs endured by a monetary economy.

To conclude on that period, what appears in Keynes' initial appraisal during the years 1931–32 regarding the development of the Great Depression is that the theoretical underpinnings of Keynes' diagnosis as well as his policy advocacies still mainly relied on the theoretical arguments provided in the *Treatise on Money*. The very originality of the *Treatise on Money* was the idea that saving and investment usually do not coincide, because of the role played by financial markets as an intermediary between savers and investors. The trade cycle is explained by the existence of windfall profits or losses, namely the short-term forecasting errors regarding sale proceeds and costs. Besides, the *Treatise on Money* accounted for cumulative self-feeding mechanisms due to deleverage effects, i.e. the lack of re-equilibrating competitive forces at some point of the time. But there were two defects that prevented a consistent explanation of the severe recession of the time. First the *Treatise on Money* was mainly concerned with the issue of fluctuations around a long-term equilibrium because of mismatches in short-term expectations. The deflationary shock might have been severe; according to the *Treatise on Money,* the economic system eventually reaches its full employment level. Second, if it well accounted for the role played by destabilising short-term expectations in these self-feeding dis-equilibrating forces, the *Treatise on Money* did not address the issue of the stabilising forces that prevented an economy to completely collapse after a protracted deflationary shock.[12] This required shifting the attention towards the construction of a case for unemployment equilibrium.

The quest for theoretical underpinnings of Keynes' understanding of the development of the Great Depression precisely led to the revision of the *Treatise on Money*,

[12]The banana parable of the *Treatise on Money* showed how an economy that produce and consume only banana would completely collapse in case of rise in the propensity to save banana. The identification of the deficiencies of the reasoning in this parable proved crucial in the way towards the elaboration of the *General Theory*.

a process that took Keynes and the young economists around him about 3 years and resulted in the publication of the *General Theory* (Dimand 1988).

3.3 From 1932 to 1933 Onwards: The Great Depression and Mismatches in Long-Term Expectations

In 1933, at a time the bulk of the theoretical arguments that would be put forward in the *General Theory* were well developed and also articulated the ones with the others, Keynes made significant alterations in his policy advices too. First, from 1933 onwards the argument relied on the lack of re-equilibrating forces in the market for goods and not only on the credit market and on financial markets. First, it was argued that the destabilising forces at work feed themselves and that a laissez-faire economy would inevitably plunge into a deep recession. Second, Keynes put even more the stress on the expectations issue. The expectations involved in the disease to be explained are not merely short-term mismatches in forecasting errors regarding sale proceeds and costs (as in the *Treatise on Money* and as in Friedman). The basic defects in the formation of expectations applied henceforth to long-term expectations regarding the state of effective demand in the periods ahead. Third, this alteration in Keynes' perception of expectations mismatches led him to shift emphasis from monetary loosening to public works. Let us argue all these three points.

Regarding first the lack of competitive re-equilibrating forces on the market for goods, Keynes narrated the consequences of a "state of financial tensions"[13] as follows:

> The reduced demand, which is the same thing as reduced purchasing, causes prices to fall; the fall of prices diminishes profits; and the entrepreneurs of the world, whether they are in difficulties or not, have a diminished incentive to produce output or to make the purchases and create the incomes which would accompanied it. Thus the declines in demand, in prices, in profits, in output and in incomes feed on themselves and one another.
>
> When financial tension leads to a diminution in demand, the decline necessarily feeds on itself, because each step which an individual (or a community) takes to protect himself and to relieve his own tension merely has the effect of transferring the tension to his neighbour and of aggravating his neighbour's distress. The course of exchange, as we all know, moves around a closed circle. When we transmit the tension, which is beyond our own endurance, to our neighbour, it is only a question of a little time before it reaches ourselves again travelling round the circle. (Keynes 1932, CW 21, pp. 212–13)

There is thus the idea is that deflationary mechanisms can easily feed themselves in the real sphere of the economy as well. When the initial shock is severe and also protracted, the competitive forces on which the liquidationists rely to advocate

[13] A state of financial tension means that individuals and communities suddenly find much increased difficulty in putting their hands on money to meet their obligations, with the result that they take various measures to reduce their purchasing. Others, not actually in difficulty, fear that the same thing may overtake them later, and from precaution reduce their purchasing also (Keynes 1932, CW 21).

deflationary policy might simply vanish. Shocks are transmitted from one market to the others, with amplifying effects:

> The reduced demand, which is the same thing as reducing purchasing, causes prices to fall; the fall of prices diminishes profits; and the entrepreneurs of the world, whether they are in difficulties or not, have a diminished incentive to produce output or to make the purchases and create the incomes which would have accompanied it. Thus the decline in demand, in prices, in profits, in output and in incomes feed on themselves and one another.
>
> (Keynes 1932, CW 21, p. 212)

Second, expectations play a critical role in this self dis-equilibrating and self-feeding mechanisms. There is first the issue of the expectations applying to the credit market. For the investors to accept to borrow and to take what they would consider "actuarial" or "normal" risks, and for the borrowers to accept to lend at cheap rates, a significant confidence must prevail regarding the state of the economy in the periods ahead. But there is also the issue of the expectations prevailing on the goods markets, regarding the state of "effective demand" in the periods ahead. Here, the government plays a critical role in the control of these expectations. The point to be emphasised here is the self-fulfilling character of the fiscal policy undertaken by the government. Keynes stated in a paper for *The Times* (which was his answer to the critiques of his Means to Prosperity):

> Unfortunately the more pessimistic the Chancellor's policy, the more likely it is that pessimistic anticipations will be realised and vice versa. Whatever the Chancellor dreams, will come true! We must begin by resuscitating the national income and the national output. (Keynes 1933, CW 21, p. 184)

What appears here is a critical aspect of the argument developed in the *General Theory*. Expectations regarding the state of the economy in the remote future are indeterminate, for they mainly depend on "conventional" views. Besides, expectations regarding the effective demand in the periods ahead are self-realising: if private agents (are crucially entrepreneurs and investors) expect a low level of effective demand in the future, they adapt to these expectations in having today a low demand price for capital goods, with the eventual result of a low level of effective demand today. Accordingly, the government has a critical role to play in the control and in the stabilising of these long-term expectations. A government that appears itself as pessimistic about the state of the economy in the periods ahead (and which develop accordingly pessimistic expectations regarding tax proceeds) would be tempted to reduce its level of public investment. Through their effects on private effects, these pessimistic expectations from the part of public authorities would be simply self-realising.

Third, what about economic policy? By the end of 1933 the emphasis was reversed from monetary loosening to public works. In a paper published in *The New York Times,* 31 December 1933, Keynes argued: "In the field of domestic policy, I put in the forefront, for the reasons given above, a large volume of loan expenditure under government auspices" and about monetary policy: "I put in the second place the

maintenance of cheap and abundant credit, in particular the reduction of the long-term rate of interest" (Keynes 1933, CW 21, p. 307). With no surprise, economic policy aims for Keynes precisely to stabilise expectation.

Through the study of Keynes' political writings after the worst of the Great Depression, when in particular Great Britain slowly started to recover, it is possible to disentangle two issues. There is first the policy schemes called for during the Great Depression, public works and monetary loosening for economic recovery, with a reversal of priorities between the two sides of economic policy that occurred around 1932. In that case, "the object is to start the ball rolling" (Keynes 1934, CW 21, p. 307). Second, there is the issue of the prevention of recession, the policy schemes designed by Keynes after the *General Theory* to ensure that the Great Depression would never happen again: "emphasis should be placed primarily on measures to maintain a steady level of employment and thus to prevent fluctuations" (Keynes 1943, CW 27, p. 323).

What appears very clearly after the *General Theory* is that, first, Keynes was not merely concerned with inflationary expectations (as Friedman), but with long-term expectations regarding the economic prospects in the remote future (Rivot 2017).[14] It can be argued that the market failures trouble that prevails in a monetary economy à la Keynes is a twofold coordination problem connected to a long-term perspective. An economy might be trapped in a less-than-full-employment equilibrium either because private agents (crucially wealth owners and investors) expect a low level of effective demand or because they attribute a very low weight to these expectations. On the one hand, pessimistic expectations lead to a low marginal efficiency of capital and this heading will be roughly addressed by fiscal policy. On the other hand, the interest rate as it is actually determined is not able to ensure the indirect coordination of savers' and investors' plans: low confidence placed in expectations leads to a high liquidity preference, and this heading will be roughly addressed by monetary policy.

Keynes' overall policy advices can be understood as a way to prevent the Great Depression from recurring, which means the stabilisation of the economy around its full employment level. Eventually, Keynes' economic policy is a double-sided weapon aiming at the stabilisation of long-term expectations at their full employment level as well as the establishment of a strong degree of confidence in these expectations (Rivot 2017).

[14] As famously stated in his 1937 paper published as an answer to the critiques addressed against the *General Theory*: "By 'uncertain' knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. […] The sense in which I am using the term [uncertainty] is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth owners in the social system in 1970. […] We simply do not know." (Keynes 1937, CW 14, pp. 113–14).

4 Conclusion

Despite the fact that they are usually considered as poles apart on almost everything, in both Keynes and Friedman expectations are at the core of their understandings of disequilibria. And for both of them, the *monetary* character of the economy implies some fundamental alteration in its functioning. But yet, this is not to deny the significant and encompassing differences between these two economists, regarding the role played by money, the relevant disturbances at work and by the way the potential existence of endogenous correcting forces. For Keynes and as well as for Friedman, the Great Depression is a dramatic episode that helped them to shape their mind about the functioning of a monetary economy and to build their policy advices accordingly. So what Keynes and Friedman can offer us are dissimilar but yet complementary ways of addressing this riddle, while focusing in each case on the expectations issue.

Concerning precisely Friedman, it is misleading to considering him as the "inflation economist", with a research programme anchored in the attempts to cope with the inflationary tensions of the 1970s. The young Friedman had been deeply impressed by the Great Depression (Snowdon and Vane 1999). The provision of a convincing rationale for that dramatic episode (at the theoretical level) and the call for devices that would ensure that the Great Depression would never appear again (at the policy level) are at the core of his matters of concern. Noticeably, the *Monetary History of the US* not provided any analytical arguments for an understanding of the forces at work during the Great Depression. Besides, the issue of unemployment was not directly addressed in that book. What the *Monetary History of the US* provided was a narrative of the monetary disturbances that eventually led to severe decline in aggregate demand. But that book remained silent on the successive transmission mechanisms from shocks in money supply to aggregate demand, and from that to unemployment.[15] The reading of Friedman's theoretical pieces of work leads to focusing the attention on the role played by expectations in his explanation for the Great Depression. Precisely, in Friedman the point is that short-term inflationary expectations take time to adapt to a changing monetary environment: the information set is not complete and the speed of adjustment in price forecasts is not instantaneous. Since it repeated mistakes in monetary policy through its inability to raise money supply at the required level to stabilise nominal aggregate demand, mismatches in private inflationary expectations too were repeated until at least 1932. In our view, here is to be found the rationale behind Friedman's k per cent rule. In the final analysis, thanks to the stabilisation of money supply, price devices are able to play their coordinating role in a decentralised market economy and mismatches in inflationary expectations are eventually removed.

By contrast, Keynes' rationale for the Great Depression is as a general proof of the inefficiency of a decentralised market economy. In Keynes, the trouble at stake regarding expectations is rather the lack of a unique intertemporal relative price

[15] By the same token, the *Monetary Trends* did not consider the Great Depression as a peculiar episode and a Chapter that could have look like the "Great Contraction" was removed in the *Monetary History*.

system. As we have shown, the Great Depression was the occasion for Keynes to reshape his mind about the functioning of a monetary economy. The Keynes of the *Treatise on Money* considered the role played by mismatches in sale forecasts (i.e. mismatches in short-term inflationary expectations as in Friedman). But Keynes added to that the deflationary effects played by bankruptcies and deleverage effects. This allowed him to take into consideration destabilising and self-feeding deflationary effects on credit market and on financial market. In the road towards the *General Theory*, Keynes shifted interest towards long-term expectations regarding the structure of the economy in the remote future as well as towards the degree of confidence of private actors in these expectations. He also extended the role played by destabilising and self-feeding deflationary effects on the goods market. Eventually, Keynes built the case for an economics of depression: a monetary economy might easily be trapped *ad infinitum* in unemployment equilibrium. Keynes' policy advocacies aim accordingly to stabilise a monetary economy at its full employment level. First, regarding fiscal policy there is the need to implement public works in order to escape the slump but also the erection of a capital budget to prevent the occurrence of slumps. This is to cope with the first kind intertemporal coordination failure, namely the lack of forward markets for many goods and assets, which lead to pessimistic long-term expectations. Second, monetary policy consists in monetary loosening and cheap money as a weapon of second importance in order to escape the slump but also to prevent the occurrence of slumps the implementation of a slowly decreasing long-term interest rate, which is the sign that private actors have firmer and firmer confidence in their expectations. Economic policy is thus in Keynes a double-sided weapon that aims to anchor long-term expectations regarding the economic prospects in the remote future.

Acknowledgements I am grateful to the organisers of the 4th Thomas Guggenheim Conference in the History of Economic Thought that took place at the Ben-Gurion University of the Negev, 18 and 19 December 2017, especially Arie Arnon, Karine Van der Beek and Warren Young. I am also indebted to the participants of this conference for fruitful critiques and comments.

References

Bordo, M. (2016). Milton Friedman and US monetary history. In R. A. Cord & J. D. Hammond (Eds.), *Milton Friedman. Contributions to economics and public policy* (pp. 143–155). Oxford: Oxford University Press.

De Vroey, M. (2001). Friedman and Lucas on the Phillips Curve: From a disequilibrium to an equilibrium approach. *Eastern Economic Journal, 27*(2), 127–148.

Dimand, R. (1988). *The origins of the Keynesian revolution, the development of Keynes's theory of employment and output*. Aldershot U.K.: Edward Elgar.

Dimand, R. (2016). John Maynard Keynes narrates the Great depression: his reports to the Philips electronics company. *Paper presented at the 16th Charles Gide Conference*, University of Strasbourg, April 14–16, 2016.

Dockès, P. (2017). Le Capitalisme et ses rythmes: quatre siècles en perspective. Tome I, Sous le regard des géants. Paris: Classiques Garnier.

Fisher, I. (1933). The debt-deflation theory of great depression. *Econometrica, 1*(4), 337–357.
Forder, J. (2018). Two lectures by Friedman: One famous, one good. *History of Economic Ideas* (forthcoming).
Friedman, M. (1956). The quantity theory of money – a restatement. In Friedman (Ed.), *Studies in the quantity theory of money* (pp. 3–21). Chicago: University of Chicago Press.
Friedman, M. (1962) [2007]. *Price theory*. New Brunnswick, New Jersey: Transaction Publishers.
Friedman, M. (1968). The role of monetary policy. *The American Economic Review, 58*(1), 1–17. Reprinted in Friedman (1969), pp. 95–110.
Friedman, M. (1969). *The optimum quantity of money and other essays*. Chicago: Aldine Publishing Company.
Friedman, M. (1970). A theoretical framework for monetary analysis. *Journal of Political Economy, 78*(2), 93–238. Reprinted in R. J. Gordon (Ed.) (1974), pp. 1–62.
Friedman, M. (1975). *Unemployment versus inflation*. Institute for Economic Affairs, London.
Friedman, M. (1976) [2008]. *Price theory*. New Brunnswick, New Jersey: Transaction Publishers.
Friedman, M. [1977] (1987). Nobel lecture: Inflation and unemployment. *Journal of Political Economy, 85*(3), 451–72. Reprinted in K. R. Leube (Ed.) (1987), pp. 347–69.
Friedman, M., & Schwartz, A. J. (1963). *A monetary history of the United-States 1867–1960*. Princeton: Princeton University Press.
Friedman, M., & Schwartz, A. J. (1982). *Monetary trends in the United States and the United Kingdom. Their relation to income, prices, and interest rates, 1867–1975*. Chicago: University of Chicago Press.
Hammond, J. D. (1996). *Theory and measurement, causality issues in Milton Friedman's monetary economics*. Cambridge: Cambridge University Press.
Hayek, F. A. [1929] (1933a). (In German) *Monetary theory and the trade cycle*. London and Toronto: Jonathan Cape.
Keynes, J. M. (1971–89). *The collected writings of John Maynard Keynes* (30 Vols.). London: Macmillan for the Royal Economic Society.
Klausinger, H. (1995). Schumpeter and Hayek: Two views of the great depression re-examined. *History of Economic Ideas, 3,* 93–127.
Klausinger, H. (2003). The Austrians on relative inflation as a cause of crisis. *Journal of the History of Economic Thought, 25*(2), 221–237.
Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica, 50*(6), 1345–1370.
Laidler, D. (2012). *Milton Friedman and the evolution of macro-economics. Introductory essay for Milton Friedman's collected writings, revised January 2012*. Available at: http://economics.uwo.ca/centres/epri/workingpapers.asp#2012.
Legrand, M. Dal-Pont, & Hagemann, H. (2017). Business cycles, growth and economic policy: Schumpeter and the great depression. *Journal of History of Economic Thought, 39*(1), 19–33.
Leube, K. R. (Ed.). (1987). The essence of Friedman. Stanford, California: Hoover Institute Press, Stanford University.
Lucas, R. J. (1972). Expectations and the neutrality of money. *Journal of Economic Theory, 4*(2), 103–124.
Lucas, R. E., & Rapping, L. A. (1969). Real wage, employment, and inflation. *Journal of Political Economy, 77*(5), 721–754.
Lucas, R. E., & Rapping, L. A. (1972). Unemployment in the great depression: Is there a full explanation. *Journal of Political Economy, 80*(1), 186–191.
Modigliani, F. (1977). The monetarist controversy or, should we forsake stabilization policies? *American Economic Review, 67*(2), 1–19.
Nelson, E. (2020a). *Milton Friedman and Economic Debate in the United States, 1932–1972* (Vol. 1). University of Chicago Press.
Nelson, E. (2020b). *Milton Friedman and Economic Debate in the United States, 1932–1972* (Vol. 2). University of Chicago Press.

Pavanelli, G. (2003). The great depression in Irving Fisher's work. *History of Economic Ideas, 11*(1), 151–167.
Rivot, S. (2013a). *Keynes and Friedman on Laissez-Faire and planning: where to draw the line?*. London: Routledge.
Rivot, S. (2013b). Gentlemen prefer liquidity: Evidence from Keynes. *The Journal of History of Economic Thought, 35*(3), 397–422.
Rivot, S. (2017). Economic policy as expectations management: Keynes' and Friedman's complementary approaches. *European Journal for the History of Economic Thought, 24*(5), 1053–1084.
Rivot, S. (2018). Friedman (1968) versus Phelps (1968) about policy-making. *History of Economic Ideas, 3*(suppl.), 93–121.
Schumpeter, J. A. (1939). *Business cycles. A theoretical, historical and statistical analysis of the capitalist process* (Vol. 2). New York: Mc Graw-Hill.
Snowdon, B., & Vane, H. R. (1999). *Conversations with leading economists, interpreting modern macroeconomics*. Cheltenham: Edward Elgar.

Expectations in Tobin's Macroeconomics: The Fisherian and Keynesian Roots of Tobin's q and Corridor of Stability

Robert W. Dimand

1 Introduction

Expectations were central to James Tobin's macroeconomics. His q theory of investment rested on treating equity prices as an observable measure of expectations of profitability (Brainard and Tobin 1968, Tobin and Brainard 1977). His post-1971 analysis of macroeconomic stability or instability, whether the economic system is self-adjusting after large demand shocks, depended crucially on how responsive expectations are to new observations (Tobin 1975, 2003). Along with Hicks (1935), whose influence on Tobin did not primarily involve expectations, two economists were major influences in shaping Tobin's approach to economics: John Maynard Keynes and Irving Fisher. Keynes's *General Theory* (1936) was the first economics book Tobin ever read, assigned to the 18-year-old Harvard sophomore in September 1936 for a weekly tutorial, and Tobin always declared himself an unreconstructed "Old Keynesian"[1] (Tobin 1992, 1993, 1997, Colander 1999, Shiller 1999, Dimand 2014). Tobin shared Keynes's emphasis on animal spirits driving long-period expectations about returns on investment, with Tobin's q providing a channel through which such changes in expectations had real effects. Fisher had been admired by Tobin's teacher Joseph Schumpeter, and after joining Yale in 1950 as an associate professor (Tobin was never an assistant professor), Tobin developed a deep appreciation

[1] Ironically in light of his subsequent interpretation of Keynes, Tobin's first publication, based on his undergraduate thesis (Tobin 1941), had attributed money illusion to Keynes's analysis of the labor market—at a time when Milton Friedman was still writing about fiscal policy to control inflation, rather than monetary policy (Shoup, Friedman and Mack 1943).

Presented at the Thomas Guggenheim Conference in the History of Economic Thought, Ben-Gurion University of the Negev, December 2017. I am grateful to Mauro Boianovsky for helpful comments.

R. W. Dimand (✉)
Department of Economics, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada
e-mail: rdimand@brocku.ca

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_6

of Fisher's economics (e.g., Tobin 1987a) and took the lead in rebuilding Fisher's reputation, battered by Fisher's unforgettably pithy and quotable conviction that in October 1929 stock prices had reached a permanently high plateau. Fisher (1896) had emphasized expectations of inflation as the wedge between real and nominal interest rates, and later used distributed lags of price changes to show how slow adjustment of inflation expectations changed real interest to change, as in Fisher (1926), making unemployment and output fluctuate (Dimand 1999). Fisher's 1906 concept of the net present value of the expected stream of earnings as the market value of assets provided the numerator of Tobin's q. The tension in Fisher's work between Fisher's neoclassical theory of interest and capital, showing the coordinating role of financial markets, and Fisher's debt-deflation theory of great depressions (Fisher 1933, Tobin 1980a), showing how such coordination could break down, was mirrored by a tension in Tobin's work between an appreciation of the technical efficiency of the financial system (e.g., Tobin 1969, 1971) and a concern that large demand shocks could push the economy outside the corridor of stability within which it was self-adjusting.

2 Keynes's Q, Tobin's q, Fisher's Net Present Value

The driving force of the economic system in John Maynard Keynes's *Treatise on Money* (1930) was Q, profits or windfalls, equal to I−S, the *ex-post* difference between investment and saving. The symbol Q for above-normal profits brought Alfred Marshall's quasi-rents to the reader's mind (Keynes's 1933 and 1934 lectures defined Q as expected quasi-rents, Rymes 1987, 1989). Profits in the production of consumption goods were $Q_1 = I' - S$, while profits in the production of investment goods were $Q_2 = I - I'$, where I is the market value of newly produced investment goods and I' the cost of producing those investment goods. Q in that sense was a measure of surprise, causing entrepreneurs to change their expectations of profitability and thus affecting their investment in the next period. In one passage, however, Keynes (1930, Vol. I, p. 159, italics in original) remarked that, "We have spoken so far as if entrepreneurs were influenced in their prospective arrangements entirely by reference to whether they are making a profit or loss on their current output as they market it [but] it is obviously the anticipated profit or loss on new business just concluded, which influences them in deciding the scale on which to produce and the offers which it is worthwhile to make to the factors of production. Strictly speaking, we should say that it is the *anticipated* profit or loss which is the mainspring of change, and that it is by causing anticipations of the appropriate kind that the banking system is able to influence the price-level." Tobin's q (or Tobin and Brainard's q, since it first appeared in Brainard and Tobin 1968, written with his then Ph.D. student and later colleague William Brainard), the ratio of the market value of equity to the replacement cost of the capital underlying the equity, would be greater than, equal to, or less than one as Keynes's Q_2 was greater than, equal to, or less than zero. In the notation of Keynes (1930), and ignoring for the moment that Tobin's q refers to the market value and replacement cost of the stock of capital rather than of

the period's new investment goods,[2] Tobin's q would be I/I′ where Keynes's Q_2 was I−I′. Tobin's q was a measure of expected profit because the market value of equity, the numerator of Tobin's q, is the net present value of the expected stream of income from owning the equity. Net investment would be zero when $q = 1$. If q exceeded one, a company could increase its net worth of its existing equity by additions to its capital stock financed by issuing equity. The rate of investment would depend on how far q was from zero.

Klaus J. W. Schmidt (1995, p. 175) states that honesty requires that Tobin's q should be called Myrdal's Q, since Gunnar Myrdal used a similar concept and even the notation Q in an essay in Swedish in 1931 (expanded in a 1933 German version that was, unknown to Schmidt, translated into English as Myrdal 1939, reissued in Kelley's Reprints of Economic Classics in 1965). Being apparently unaware of Keynes's Q in *A Treatise on Money* (1930), Schmidt attached no significance to Myrdal's multiple citations of Keynes (1930) in 1931, repeated in 1933 and 1939, or to being told by Tobin of the influence on him of Keynes's *Treatise on Money*, the only mention of Keynes (1930) in Schmidt's article. Myrdal's Q was an explicit reference in 1931 to Keynes's Q of the previous year and indirectly to Marshall's quasi-rents (see Dimand 2014, pp. 75–76, 79–81[3]). Tobin with Golub (1998, p. 150 n3) followed Schmidt in stating that Myrdal "long anticipated q, even calling it Q! However, his Q was not a ratio but the absolute difference between market value and replacement cost."[4] However, Myrdal was only following the notation of Keynes (1930) and attempting to clarify Keynes's handling of *ex-ante* and *ex-post* concepts. Wicksell ([1898] 1936) influenced both Keynes (1930) and Myrdal (1939), especially Wicksell's distinction between natural and market rates of interest (see Jüttner 1987 on affinities between Wicksell and Tobin on investment, and note the reference by Tobin with Golub 1998, p. 264, to "the Wicksellian, *q* ratio story"). The lineage of Tobin's q is Keynesian, both to the *Treatise on Money* and *The General Theory*, and indirectly Wicksellian, through Wicksell's influence on Keynes's *Treatise on Money*.

The numerator of Tobin's q, the market value of equity, is the net present value of the income stream that investors expect to obtain by owning that equity, a concept that was the central message of Irving Fisher's *The Nature of Capital and Income* in 1906 (reprinted in Fisher 1997, Volume 2). For Fisher, the time pattern of expected income was fundamental; the stock of capital was simply the discounted value of that stream. Tobin, who was consulting editor for William Barber's 1997 edition of

[2]Tobin (with Golub 1998, p. 153 n5) credited Abba Lerner (1940, p. 334) as "the first to point out that there was a stock-flow confusion in Keynes's (1936, Ch. 11) investment function."

[3]That section of Dimand (2014) is based on joint work with Harald Hagemann.

[4]Tobin (in Tobin with Golub 1998) stated that he had not known Myrdal's Q (and followed Schmidt in thinking that Myrdal's monograph had never been translated into English). He only narrowly missed learning of Myrdal at a formative point in Tobin's career: The Canadian economist and public servant Robert Bryce, one of the translators of Myrdal (1939), was a graduate student at Harvard from 1935 to 1938, while Tobin was an undergraduate there. Bryce, who had attended Keynes's lectures from 1932 to 1934, brought the Keynesian message to Harvard (in the words of Schumpeter, Bryce's supervisor at Harvard, "Keynes is Allah and Bryce is his prophet") at a time when Tobin was discovering Keynes.

The Works of Irving Fisher, wrote the editorial introduction and afterword for only one of the fourteen volumes: *The Nature of Capital and Income*. In Tobin's writings on Fisher, such as his 1987 *New Palgrave* entry on Fisher (Tobin 1987a) or his 1985 article on Fisher and John Bates Clark for the *American Economic Review* special issue on the centenary of the founding of the American Economic Association, *The Nature of Capital and Income* always figured prominently.

Tobin always insisted on the importance of the numerator of q being observable, a summary of the expectations of investors. Fumio Hayashi (1982) offered a neoclassical interpretation of marginal q as the shadow price of installed capital in an optimizing model of investment subject to internal, strictly convex costs of capital stock adjustment (see Buiter 2003, p. F599). If both the production function and adjustment costs were linear homogenous in their arguments, marginal q (the shadow price of installed capital) would equal the average value of installed capital (Tobin's q). Tobin (interviewed by Shiller 1999, p. 887) objected that Hayashi's marginal q "is a shadow price of an optimal program solution … not something you could actually measure as a market variable."[5] Tobin was concerned with the Fisherian discounting of the stream of expected earnings as shown by asset prices.

According to Tobin (with Golub 1998, p. 152), "Empirically, it is quite obvious that stock market *q*s and formal implicit *q*s are not the same animals. Variations in marginal cost of adjusting capital stocks by investment would have to be implausibly large to be consistent with fluctuations in observed market valuations. Like Keynes's view, the position here is that the stock market does not grind out values by mirroring the rational optimization of informed managements but generates values of its own. These nevertheless provide incentives or disincentives for investment. Tobin's *q* is so far from being a thoroughgoing neoclassical theory that it is quite consistent with recognition that corporate managers and other economic agents respond to market noise and are in any case sluggish in responding to the arbitrage opportunities of large deviations of *q* from par."

3 Liquidity Preference as Behavior Toward Risk

"Nearly two decades of drawing downward-sloping liquidity preference curves in textbooks and on classroom blackboards should not blind us to the basic implausibility of the behavior they describe," stated Tobin (1958, p. 65). "Why should anyone hold the non-interest-bearing obligations of the government instead of its interest-bearing obligations?" For both Keynes and Tobin, the answer was that the price of bonds could fluctuate, but they interpreted that characteristic of bonds in contrasting ways.

[5] In addition to emphasizing the market prices of equities as a measure of expectations of returns on investment, Tobin also took an interest in survey data on consumer's intentions and expectations (e.g., Tobin 1959).

The title of Tobin (1958), "Liquidity Preference as Behavior Toward Risk," nodded to Keynes by referring to money demand as liquidity preference, but the title was exact in stating that the article was analyzing demand for money as an asset as behavior toward risk, not fundamental uncertainty. Tobin (1958) represented expectations of returns on risky assets by subjective probability distributions, whereas Keynes (1937) had insisted that expectations about an uncertain future could not be reduced to a probability distribution (but the account of the speculative motive in Keynes 1936 had investors holding a point estimate of the future interest rate with certainty, notwithstanding the emphasis on uncertainty pervading the rest of the book). Nonetheless, Tobin (1958) did build upon Keynes's account in *The General Theory* of the speculative motive for holding money and extended it in a way that brought it into closer agreement with the facts of how people hold wealth. Keynes (1936, pp. 170–74) posited that each wealth-holder had some expectation of what the interest rate would be in the future. Comparing that expectation of the future interest rate with the current interest rate yields a prediction of a capital gain or loss from holding a bond. If the expected capital loss exceeded the interest to be received on the bond, the wealth-holder would be better off owning no bonds and holding only cash. If the expected capital loss was less than the interest, or if the wealth-holder expected a capital gain on the bond, the wealth-holder should hold all of her or his wealth in bonds and none as money. The market price of securities will be the one that balances the sales of the "bears" who expect bond prices to fall (the interest rate to rise) with the purchases of the "bulls" who expect bond prices to rise and the interest rate to fall (Keynes 1936, p. 170). Each person holds an undiversified portfolio of all securities or all money, but expectations of the future interest rate vary across individuals, so liquidity preference is a function of the interest rate (as well as of the level of income), with a small rise in the interest rate switching a few people from being bearish about bond prices to being bullish. Keynes (1936, p. 172) found it "interesting that the stability of the system and its sensitiveness to changes in the quantity of money should be so dependent on the existence of a *variety* of opinion about what is uncertain. Best of all that we should know the future. But if not, then, if we are to control the activity of the economic system by changing the quantity of money, it is important that opinions should differ. Thus, this method of control is more precarious in the USA, where everyone tends to hold the same opinion at the same time, than in England where differences of opinion are more usual" (Keynes's italics). Keynes (1936) was the first to write money demand as a function of interest and income (except for Walras's *encaisse desirée* half a century before), or indeed to explicitly write any asset demand function (as distinct from verbal statements).

Tobin (1958) found it not merely interesting, but also unsatisfactory, that people with the same information should hold different expectations, and that a tiny change in asset prices should make investors switch all their wealth from cash to bonds or vice versa (see Tobin 1983, 1984b). Tobin was an admirer of Harry Markowitz's application of linear programming to optimal portfolio diversification: when the Cowles Commission for Research in Economics left the University of Chicago in 1955 to move to Alfred Cowles's alma mater, Yale University, with Tobin as director of the new Cowles Foundation, Markowitz moved with Cowles, working on the

Cowles Monograph published as *Portfolio Selection* (Markowitz 1959). Markowitz (1952) had already published the central concept of that study, combining risky assets in a diversified portfolio to minimize variance for each given expected rate of return on the portfolio, in his Chicago doctoral dissertation and an article that Tobin later chose to include in *Landmark Papers in Macroeconomics Selected by James Tobin* (Tobin 2002). Tobin derived a money demand function that responded to the interest rate like Keynes's liquidity preference function, but instead of assuming a distribution of people's point estimates about what the interest rate would be, Tobin assumed a probability distribution over what the interest rate would, with people who held the same information sharing the same subjective probability distribution about asset returns: "My theory of liquidity preference as behavior towards risk was built on a rational expectations model long before the terminology," Tobin told Shiller (1999, p. 878). This led Tobin to be, in the words of Buiter (2003, p. F587), "not an unqualified admirer" of the extension of the Tobin–Markowitz mean-variance approach by William Sharpe, John Lintner, and Jan Mossin into the Capital Asset Pricing Model (CAPM): The assumption of homogenous beliefs made CAPM, in the eyes of Tobin (1983), a representative-agent model, the sort of single-agent, no-trade model that Tobin considered unhelpful for understanding the economy (on which see Tobin 1987b, and on two-agent overlapping-generations models, Tobin 1980b). As with Tobin's 1952 statement of what later became known as debt neutrality or Ricardian equivalence (see Buiter 2003, p. F609, Dimand 2014, p. 50), Tobin here at least partially anticipated an approach whose later uses and extensions did not meet with his approval.

The separation theorem of Tobin (1958) showed, taking money as a riskless asset (in nominal terms) with an exogenously fixed return strictly lower than the expected return on risky assets, risk-averse investors would choose the fraction of their portfolio to hold in the riskless asset by trading off risk against expected return, but all would hold the same combination of risky assets, a result summarized by Buiter (2003, p. F587) as "regardless of your degree of risk aversion and caution, you will only need *two* baskets for all your eggs." Fisher (1928) had tried fervently to dissuade people from thinking of money as riskless, given fluctuations in the purchasing power of money; perhaps US Treasury Inflation-Protected Securities (TIPS) or UK or Canadian government real-return bonds would be more appropriate as riskless assets. Comments by Karl Borch and by Martin Feldstein in 1969 objected that Tobin's mean-variance analysis was exact only if asset returns were normally distributed (so that the probability distribution of returns was fully described by its first two moments) or if investors had quadratic utility functions (so that they only cared about the first two moments of the probability distribution of returns). Tobin (1971, p. 269) was unimpressed by the comments of Borch and Feldstein: "I do not believe it is an exaggeration to say that, until relatively recently, the basic model of portfolio choice in economic theory was a one-parameter model. Investors were assumed to rank portfolios by reference to one parameter only—the expected return, possibly corrected by an arbitrary 'risk premium,' constant and unexplained … This extension from one moment to two was never advertised as the complete job or the final word, and I think that its critics in 1969 owe us more than demonstrations that it rests on

restrictive assumptions. They need to show us how a more general and less vulnerable approach will yield the kind of comparative-static results that economists are interested in. This need is satisfied neither by the elegant but nearly empty existence theorems of state preference theory nor by normative prescriptions to the individual that he should consult his utility and his subjective probabilities and then maximize."

4 Expectations and Macroeconomic Instability: An "Old Keynesian" View

Tobin (1980a, 1980b, 1987b) was an outspoken "Old Keynesian" critic of rational expectations and of New Classical claims to have established rigorous choice-theoretic general equilibrium microeconomic foundations for New Classical macroeconomics (see Lucas 1981b for a New Classical riposte). Like his younger colleague Robert Shiller (2000), Tobin (1984) doubted that financial markets were efficient in any macroeconomic sense of establishing asset prices that in reflected underlying fundamental values, or in preserving macroeconomic stability (see Colander 1999 and Shiller 1999 for connections between Shiller and Tobin). Tobin (1975, 1980a, 1992, 1993, 1997) argued that faster adjustment of prices and money wages could well be destabilizing (see Driskill and Sheffrin 1986, De Long and Summers 1986, Chadha 1989). Contrary to the conclusion by Don Patinkin (1965) that the Pigou–Haberler real balance effect ensured that, as a matter of theory rather than practical policy, a sufficiently low price level and money wage would always suffice to restore full employment, Tobin (1975, 1980a, 1992, 1993, 1997) held that the effect of falling prices, and the increased default risk associated with a higher real value of inside debt, could swamp the real balance effect of an increase in the real value of the small amount of outside money. In arguing so, Tobin drew on the analysis of the potentially destabilizing effect of money wage cuts advanced in Keynes (1936, Chap. 19 "Changes in Money Wages"), and, from Tobin (1980a) onwards, also on the debt-deflation theory of depressions of Fisher (1933).[6]

The crucial step in the development of Tobin's thought on macroeconomic stability came with the 1971 reprinting of his 1965 Irving Fisher Lecture to the Econometric Society about "Money and Economic Growth." Following the observation that "An accelerated decline in prices means a more attractive yield on money and encourages a further shift in portfolio demand in the same direction as the original shock," Tobin (1971, p. 145) deleted the original stability analysis and added, "The issue depends on the speed with which actual price movements are translated into expectations. If the process is sluggish—expectations are inelastic—then the stabilizing Pigou effect will win out. But, if current experience has a heavy weight in formation of expectations, the system can be unstable." Tobin (1975) provided a simple model demonstrating the possibility of instability if prices and money wages adjusted rapidly (see Bruno

[6]Minsky (1975) drew on Keynes (1936, Chap. 19), as did Tobin (1975), and on Fisher (1933), which did not appear in Tobin's writings until Tobin (1980a).

and Dimand 2009 for a derivation of the stability condition for Tobin's 1975 model, and Palley 2008). Tobin (1978, p. 524) elaborated on this possibility of instability when presenting his proposal for a tax on international currency transactions: "As a technical matter, we know that a rational expectations equilibrium in markets of this kind is a saddle point. That is, there is only a singular path that leads from disequilibrium to equilibrium. If the markets are not on that path, or if they do not jump to it from wherever they are, they can follow any number of paths that lead away from equilibrium—paths along which, nonetheless, expectations are on average fulfilled. Such deviant paths are innocuous in markets—as for rare coins, precious metals, baseball cards, Swiss francs—which are sideshows to the real economic circus. But, they are far from innocuous in foreign exchange markets whose prices are of major economic consequence." Where Milton Friedman (1968) had invoked adjustment of expectations to argue against the possibility of government intervention to improve macroeconomic outcomes, and Robert Lucas (1981a) invoked expectations that were satisfied on average to argue against any systematic effects of government macroeconomic policy, Tobin pointed to rapid adjustment of expectations as a source of instability and to the possibility of multiple paths for the economy, along each of which expectations would be satisfied on average. As Mishkin (1983) showed, policy ineffectiveness did not necessarily follow even if expectations were right on average.

Together with Hyman Minsky (1975, 1982, 1986), Tobin was responsible for directing attention to those two long-neglected works, Keynes's Chap. 19 and Fisher's debt-deflation theory (although neither Minsky nor Tobin cited Fisher 1932, on which Fisher 1933 was based, apart from Tobin's 1987 *New Palgrave* survey of Fisher's career). Tobin and Minsky both studied with Joseph Schumpeter and Wassily Leontief at Harvard, their studies at Harvard overlapping from 1946 to 1949 (Tobin received his Ph.D. in 1947, the year of Minsky's master's degree, but Tobin remained as a Junior Fellow for two more years). Nonetheless, Minsky (1981, 1986) and such followers as James Crotty (1990) upbraided Tobin for supposedly taking classical rather than Keynesian positions on each of the issues mentioned above (a denial of Tobin's credentials as a Keynesian that paralleled Tobin's denial, when interviewed by Colander 1999, that New Keynesians such as Gregory Mankiw were Keynesians). Minsky (1981) and Tobin (1989), their review articles of each other's books, reveal a failure to reach agreement about whether they agreed about Keynes, expectations and macroeconomic stability (see Dimand 2004).

Tobin (1989, p. 107) vehemently objected that Minsky (1986, pp. 5n, 133–138) "accuses the misguided Keynesians of embracing the Pigou–Patinkin real balance effect as a proof that flexibility of wages and prices ensures full employment so that government macroeconomic interventions are not needed. This is just not true. I, for example, say the opposite in publications that Minsky knows and actually cites," including Tobin (1975, 1980a). Tobin (1989, p. 106) concluded that "this 'post-Keynesian' theory [mark-up pricing] is not convincingly linked to the central

message of the book [Minsky 1986[7]], the financial theory of business cycles. Minsky's excellent account of asset pricing and investment decisions is separable from his theory of prices, wages and profits. It sounds like 'q' theory to me … He is right to stress that 'inside' monetary and financial institutions and markets make a big difference, and to reject 'Modigliani–Miller theorems that assets and debts which wash out in accounting aggregations wash out in economic effects as well. Minsky's classifications of debt finance—'hedge', 'speculative', and 'Ponzi'—are suggestive and helpful."[8] Tobin did not, however, succeed in persuading Minsky and Crotty that Tobin shared Minsky's view of asset pricing, investment, animal spirits, macroeconomic instability, and systemic financial fragility or even that he had claimed to do so.

James Crotty (1990) contrasted the Keynes-based approach of Hyman Minsky (1975, 1986) to systematic financial fragility, fundamental uncertainty, and shifting long-period expectations with the allegedly neoclassical q theory of investment, as presented in Tobin and Brainard (1977). Tobin and Brainard (1990, p. 543) responded indignantly, insisting that they had always shared Keynes's "stress in Chap. 12 of the *General Theory* on the inevitable role of non-rational attitudes—optimism and confidence or their opposites—in forming estimates of the marginal efficiency of capital … Nothing excuses [Crotty's] charge that 'Tobin places Keynes's stamp of approval on rational expectations, efficient-markets general equilibrium models that are the modern extension of the classical theory Keynes so vehemently opposed'" (a charge that would have surprised Minsky's former Berkeley student Robert Lucas 1981b).[9]

Crotty's phrase "Tobin's stable and efficient financial markets" particularly offended Tobin and Brainard (1990, p. 549): "We did not use the word 'stable'. Our word 'efficient' referred only to technical market-clearing efficiency. We did not say or mean that stock markets come up continuously with fundamental valuations. In our 1977 article, which Crotty cites, and in others on 'q', we followed Keynes in believing that speculation makes prices diverge from fundamental valuations. Again putting his own words in Tobin's mouth, Crotty says in his footnote 9 that in his 1984[a] article, 'Tobin appears to recant his belief in the valuation efficiency of financial markets.' The term 'valuation efficiency' does not appear in our 1977 article, and no other writing of ours, individual or joint, asserts such a belief. Tobin has nothing to recant." Tobin (1984a) viewed financial markets as informationally efficient in the weak and semi-strong senses that one cannot systematically profit by using generally available public information, and technically efficient in the sense that

[7]Minsky (1986) was published by the university press of Tobin's own university, Yale, and, as a Twentieth Century Fund study, listed opposite the title page the Twentieth Century Fund trustees, including James Tobin.

[8]Tobin made a similar remark when the present author wrote a term paper on Minsky's theory of asset pricing and investment for Tobin's Money and Finance course in 1979–80, a year after graduating from a Post-Keynesian-learning program at McGill.

[9]However, Colin Rogers (1989, pp. 119–122) tracked down a few instances where Tobin, and Brainard and Tobin (1968, pp. 353, 365), used the term marginal productivity of capital rather than marginal efficiency of capital.

one can trade large quantities of financial assets with low transactions costs, at little or no notice and without significant effect on market prices. He emphatically did not accept that financial asset prices necessarily reflect rational expectations of future payments from owning the assets ("valuation efficiency"), let alone that financial markets necessarily support Pareto-efficient economy-wide outcomes, what Tobin (1984a) termed Arrow-Debreu full insurance efficiency (see Buiter 2003, pp. F589, F604–F605, Shiller 2000).

5 Conclusion

In his q theory of investment, in his analysis of money demand and portfolio choice, and in his investigation of macroeconomic instability, Tobin focused on expectations, expectations of returns on assets and of inflation. His analysis was informed by his close study of Keynes (1930, 1936) and Fisher (1997), two economists for whose work he felt a strong affinity and whose work often served as his starting point even when, as with Tobin (1958) on liquidity preference as behavior toward risk, he was quite prepared to significantly modify their approach.

References

Brainard, W. C., & Tobin, J. (1968). Pitfalls in financial model-building. *American Economic Review: AEA Papers and Proceedings, 58*(2), 99–122.

Bruno, R., & Dimand, R. W. (2009). The corridor of stability in Tobin's Keynesian model of recession and depression. *International Journal of Applied Economics and Econometrics, 17*(1), 17–25.

Buiter, W. (2003). James Tobin: An appreciation of his contributions to economics. *Economic Journal, 113*(491), F585–F631.

Chadha, B. (1989). Is increased price inflexibility stabilizing? *Journal of Money, Credit, and Banking, 21*(4), 481–497.

Colander, D. (1999). MD dialogue: Conversations with James Tobin and Robert Shiller on the 'Yale tradition' in macroeconomics. *Macroeconomic Dynamics, 3*(1), 116–143.

Crotty, J. (1990). Owner-management conflict and financial theories of investment instability: A critical assessment of Keynes, Minsky, and Tobin. *Journal of Post Keynesian Economics, 12*(4), 519–542.

De Long, J. B., & Summers, L. H. (1986). Is increased price flexibility stabilizing? *American Economic Review, 76*(5), 1031–1044.

Dimand, R. W. (1999). Irving Fisher and the Fisher relation: Setting the record straight. *Canadian Journal of Economics, 32*(3), 744–750.

Dimand, R. W. (2004). Minsky and Tobin on the instability of a monetary economy. In M. Lavoie & M. Seccareccia (Eds.), *Central banking in the modern world*. Cheltenham, UK, and Northampton, MA: Edward Elgar Publishing.

Dimand, R. W. (2014). *James Tobin*. London, Basingstoke and New York: Palgrave Macmillan.

Driskill, R. A., & Sheffrin, S. M. (1986). Is price flexibility destabilizing? *American Economic Review, 76*(4), 802–807.

Fisher, I. (1896) *Appreciation and interest* (Vol. 1). New York: Macmillan for the American Economic Association; (Reprinted in Fisher (1997)).

Fisher I (1926) A statistical relation between unemployment and price changes. *International labour review*, *13*(6), 785–792, Reprinted in Fisher (1997) Vol. 8, and as "Lost and found: I discovered the Phillips curve. *Journal of Political Economy*, *81*(1973), 496–502.
Fisher I (1928) *The money illusion*. New York: Adelphi; (Reprinted in Fisher (1997), Vol. 8).
Fisher I (1932) *Booms and depressions*. New York: Adelphi; (Reprinted in Fisher (1997), Vol. 10).
Fisher I (1933) The debt-deflation theory of great depressions. *Econometrica*, *1*(3), 337–57; (Reprinted in Fisher (1997), Vol. 10).
Fisher I (1997) *The Works of Irving Fisher*, 14 volumes, edited by William J. Barber assisted by Robert W. Dimand and Kevin Foster, consulting editor James Tobin, London: Pickering & Chatto.
Friedman M (1968) Presidential address: The role of monetary policy. *American Economic Review*, *58*(1), 1–17; (Reprinted in Tobin, ed. (2002)).
Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica, 50,* 213–224.
Hicks J. R. (1935) A suggestion for simplifying the theory of money. *Economica* new series, *2*(1), 1–19, (Reprinted in Tobin, ed. (2002)).
Jüttner, D. J. (1987). The Wicksell-Tobin investment model. *Journal of Macroeconomics, 9,* 457–462.
Keynes, J. M. (1930). *A treatise on money, 2 volumes*. London: Macmillan.
Keynes J. M. (1936) *The general theory of employment, interest and money*. London: Macmillan.
Keynes J. M. (1937) The general theory of employment. *Quarterly Journal of Economics*, *51*, 209ff. Reprinted in Keynes (1971–89) Vol. XIV, In S. E. Harris (Eds.) (1947), *The new economics* (pp. 181–93). New York: Knopf.
Keynes J. M (1971–89) *The collected writings of John Maynard Keynes*, 30 volumes, In D. E. Moggridge & E. A. G. Robinson, volume editors E. S. Johnson and D. E. Moggridge, London: Macmillan, and New York: Cambridge University Press, for the Royal Economic Society.
Lerner, A. P. (1940). *The economics of control*. New York: Macmillan.
Lucas, R. E., Jr. (1981a). *Studies in business cycle theory*. Cambridge, MA: Oxford, Basil Blackwell, MIT Press.
Lucas, R. E., Jr. (1981b). Tobin on monetarism: A review article. *Journal of Economic Literature, 19*(2), 558–567.
Markowitz H (1952) Portfolio selection. *Journal of finance*, *7*(1), 77–91 (Reprinted in Tobin, ed. (2002)).
Markowitz, H. (1959). *Portfolio selection*. New York: Wiley for the Cowles Foundation for Research. in Economics.
Minsky, H. P. (1975). *John Maynard Keynes*. New York: Columbia University Press.
Minsky, H. P. (1981). James Tobin's *Asset accumulation and economic activity*: A review article. *Eastern Economic Journal, 7*(3), 199–209.
Minsky, H. P. (1982). *Can "IT" happen again? and other essays on instability and finance*. Armonk, NY: M. E. Sharpe.
Minsky, H. P. (1986). *Stabilizing an unstable economy*. New Haven, CT: Yale University Press for the Twentieth Century Fund.
Mishkin, F. S. (1983). *A rational expectations approach to macroeconometrics: Testing policy ineffectiveness and efficient-market models*. Chicago: University of Chicago Press.
Myrdal G (1939) *Monetary equilibrium* (R. B. Bryce & N. Stolper, Trans.) Glasgow: William Hodge; (Reprinted New York: Augustus M. Kelley, Reprints of Economic Classics, 1965).
Patinkin, D. (1965). *Money, interest and prices* (2nd ed.). New York: Harper & Row.
Palley, T. I. (2008). Keynesian models of recession and depression revisited. *Journal of Economic Behavior & Organization, 68*(2), 167–177.
Rogers, C. (1989). *Money, interest and capital*. Cambridge, UK: Cambridge University Press.
Rymes, T. K. (Ed.). (1987). *Keynes's lectures, 1932–35: Notes of students*. Ottawa: Department of Economics, Carleton University.
Rymes T. K. ed. (1989) *Keynes's lectures, 1932–35: Notes of a representative student*, London and Basingstoke: Macmillan, and Ann Arbor: University of Michigan Press.

Schmidt, K. J. W. (1995). "Tobin's q? Myrdal's Q! Ein Fallbeispiel für den Wert von Fremdsprachenkenntnissen" [Tobin's q? Myrdal's Q! An Example of the Value of Knowing Foreign Languages]. *Kredit und Kapital, 28,* 175–200.

Shiller, R. J. (1999). The ET interview: Professor James Tobin. *Econometric Theory, 15,* 867–900.

Shiller R. J. (2000) *Irrational exuberance* (2nd ed.). Princeton, NJ: Princeton University Press (2005).

Shoup, C., Friedman, M., & Mack, R. P. (1943). *Taxing to prevent inflation*. New York: Columbia University Press.

Tobin, J. (1941). A note on the money wage problem. *Quarterly Journal of Economics, 55,* 508–516.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies, 25*(1), 65–86.

Tobin, J. (1959). On the predictive power of consumer intentions and attitudes. *Review of Economics and Statistics, 41*(1), 1–11.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit, and Banking, 1*(1), 15–29.

Tobin, J. (1971). *Essays in economics* (Vol. 1). Amsterdam: North-Holland.

Tobin, J. (1975). A Keynesian model of recession and depression. *American Economic Review: AEA Papers and Proceedings, 65*(2), 195–202.

Tobin, J., & Brainard, W. C. (1977). Asset markets and the cost of capital. In R. Nelson & B. Balassa (Eds.), *Economic progress, private values, and public policy: Essays in honor of William Fellner* (pp. 235–262). Amsterdam: North-Holland.

Tobin J (1978) A proposal for international monetary reform. *Eastern Economic Journal*, *4*(3–4), 153–159; Reprinted in *Eastern Economic Journal*, *29*(4), 519–526.

Tobin J (1980a) *Asset accumulation and economic activity*. Oxford, Chicago: University of Chicago Press, Blackwell.

Tobin, J. (1980b). The overlapping generations model of fiat money: Discussion. In J. H. Kareken & N. Wallace (Eds.), *Models of monetary economies*. minneapolis, MN: Federal Reserve Bank of Minneapolis.

Tobin, J. (1983). Liquidity preference, separation and asset pricing. *Zeitschrift für Betriebswirthschaft, 53*(3), 236–238.

Tobin, J. (1984a). On the efficiency of the financial system. *Lloyds Bank Review, 153,* 1–15.

Tobin, J. (1984b). A mean-variance approach to fundamental valuations. *Journal of Portfolio Management, 20,* 26–32.

Tobin, J. (1987a). Irving Fisher. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new palgrave: A dictionary of economics*. London: Palgrave Macmillan.

Tobin, J. (1987b). 'Microfoundations' and the impasse in macroeconomics. *Mathematical Social Sciences, 14,* 195–198.

Tobin, J. (1989). Review of *Stabilizing an unstable economy*, by Hyman P. Minsky. *Journal of Economic Literature, 27*(1), 105–108.

Tobin, J., & Brainard, W. C. (1990). On Crotty's critique of q theory. *Journal of Post Keynesian Economics, 12*(4), 543–549.

Tobin, J. (1992). An old Keynesian counterattacks. *Eastern Economic Journal, 18*(4), 387–400.

Tobin, J. (1993). Price flexibility and output stability: An old Keynesian view. *Journal of Economic Perspectives, 7*(1), 45–65.

Tobin J. (1997) An overview of the general theory. In G. C. Harcourt & P. A. Riach (Eds.), *A "Second Edition" of the general theory* (Vol. 2, pp. 3–27). London and New York: Routledge.

Tobin, J., & Golub, S. S. (1998). *Money, credit, and capital*. Boston: Irwin McGraw-Hill.

Tobin, J. (Ed.). (2002). *Landmark papers in macroeconomics selected by James Tobin*. MA, Cheltenham, UK, Northampton: Edward Elgar Publishing.

Tobin J. (2003) *World finance and economic stability: Selected essays of James Tobin*, with a foreword by J. Yellen, Cheltenham, UK, and Northampton, MA: Edward Elgar Publishing.

Wicksell K. ([1898] 1936), *Interest and prices* (R. F. Kahn with introduction by B. Ohlin, Trans.). London: Macmillan.

Growth Without Expectations: The Original Sin of Neoclassical Growth Models

Michaël Assous and Muriel Dal Pont Legrand

1 Introduction

The official (hi-)story of the early developments of growth theory is presented basically as a two-step process, identified by two major contributions: Roy Harrod's "Essay in Dynamic Theory" and Robert Solow's 1956 "contribution." Specifically, Harrod is considered as being the first to point to the pervasive instability in macrodynamics, while Solow showed that this instability disappeared with flexible production function coefficients. Thus, Harrod's vision, based on the "principle of instability," was seen as if he would have hence developed a pessimistic vision of economic growth, while Solow would have on the contrary produced an "optimistic" understanding of economic development.[1]

Such an interpretation is problematic for several reasons. First, it leads to the neglect of a literature which emerged between 1948 and 1963, a period which is however identified by Young (1989) as progressive, characterized by an intensive

[1] In fact, many textbooks, although making some efforts to present Harrod's view correctly, adopt "a linear" interpretation of the connection between Harrod and Solow.

A first version of this paper was presented at the IVth CNRS GREDEG, Guggenheim Conference on the History of Economic Thought, Ben-Gurion University of the Negev, December 18–19, 2017. A second version was presented at the HES annual conference, Chicago, 2018. The authors thank Arie Arnon, Mauro Boianovsky, Robert Dimand, Harald Hagemann, Esteban Perez and Anwar Shaikh for useful comments. The usual disclaimers apply.

M. Assous
Université Lyon 2, CNRS TRIANGLE, Lyon, France
e-mail: Michael.Assous@univ-lyon2.fr

M. Dal Pont Legrand (✉)
Université Cote d'Azur, CNRS GREDEG, Nice, France
e-mail: muriel.dal-pont-legrand@univ-cotedazur.fr

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_7

cross-fertilization of ideas.[2] Second, it drastically reduces the scope of growth theory and ignores most of the "solutions" and options proposed by their protagonists which though partial and fragile, might have contributed to re-defining the growth economics frontier.

There are already papers which have contributed to remove a few misunderstandings that may have otherwise subsisted. Besomi (2000) demonstrates that Harrod's initial project was to model business cycles not economic growth,[3] while Bruno and Dal Pont Legrand (2014) show that the possibility of cyclical growth is supported by the arguments developed by Harrod in 1939. Hoover and Halsmayer (2016) examine how a "culture of misunderstanding" guided Solow's reading of Harrod and his (Solow's) own modeling strategy. We want to point out that this "culture" was exacerbated by problems related to the incorporation of expectations into his growth models and the building of robust investment functions which aspects were discussed at some length in the early 1960s by economists such as Sen (1963b), Hahn (1960), Hahn and Matthews (1964), as well as by Solow himself. In addition, we want to stress that the modeling choices implemented subsequently were implicitly based on the disconnection between expectations and growth. For instance, unpublished papers and correspondence between Sen and Samuelson and between Solow and Hahn show that those economists were perfectly aware of the limitations of their work and the fragility of their respective models—elements that seem to have been overlooked by economists who contributed subsequently to the field. How did expectations come to be ignored, and what have been the consequences for analyses of growth stability? What kind of issue(s) could growth models address without the integration of expectations? How can they describe short-run and long-run dynamics?

The paper is organized as follows. The first section discusses the debates triggered by Harrod's 1939 paper and shows how the instability issues were tackled by his contemporaries. It pays special attention to growth analysis[4] and to Solow's growth model which on the assumption that whatever is saved is invested, makes any treatment of expectations redundant. Section 2 examines published and unpublished papers and less formal exchanges among the macroeconomists of that period. A review of the contributions (and sometimes also of the reactions) of various economists from Sen—who questioned the internal consistency of the so-called neoclassical growth model, to Hahn—who explicitly addressed the question of stability in relation to expectations, and even Meade, contributes to our understanding of the process involved in building the so-called neoclassical growth model. The paper sheds

[2] Young (1989) provides an analysis of Harrod's Trade cycle group. He then investigates how Harrod went from *The Trade Cycle* to the '*Essay*.' Investigating this long period, Young identifies (p. 174) three distinct stages in the growth research program: a formative period (1924–1939), a 'take off' or 'progressive' period (1948–1963) and finally, 'degenerative' period from 1964 onwards.

[3] Besomi (2000) explains that influenced by Keynes, Harrod revised his seminal 1939 paper to focus on the stability of his proposed equilibrium and moving away gradually from business cycles analysis.

[4] Recall that Harrod's instability principle was used also to build business cycles models (Hicks 1950). We deliberately restrict the scope of this paper to the analysis of early growth models.

light on the nature of the compromises made which *in fine* led to the disconnection of expectations from growth analysis.

2 Stumbling Toward Instability: Harrod's Attempt to Incorporate Expectations

Harrod's concerns over expectations emerged initially in 1936 when he wrote *The Trade Cycle*. It was then that he explicitly introduced[5] the terms "realization" and "disappointment." Entrepreneurs have expectations about additional effective demand and determine the level of investment; in turn, investment via the multiplier affects the level of effective demand. Then, the revision of expectations following a disappointment based on a multiplier-accelerator effect gives rise to a cumulative mechanism which leads to further deviations. In 1939, cumulative dynamics were still present, and Harrod developed arguments to show that instability played a key role.[6] In his "Essay on dynamic theory," Harrod emphasized several issues. First, he examined the possibility of steady growth in a model with fixed capital-output and fixed savings-output ratios. Second, he introduced a sort of investment function[7] before examining possible sources of the instability of this steady growth path. More precisely, he analyzed under what conditions entrepreneurs' expectations are verified, and what type of dynamics emerges if they are not. In addition, he defined G_w (the warranted growth rate) as a moving equilibrium.

> (…) as actual growth departs upwards or downwards from the warranted level, the warranted rate itself moves, and may chase the actual rate in either direction. (…) Indeed, there is no unique warranted rate; the value of warranted rate depends upon the phase of the trade cycle and the level of activity. (ibid.: 29–30)

G_w as a moving equilibrium implies a dynamic approach to the warranted ICOR. It has been shown (Bruno and Dal Pont Legrand 2014) that this along with Harrod's suggested introduction of corrections to both short-term (depending on the different phases of the cycle) and long-term expectations (comparing G to G_n, respectively, the effective and natural rates of growth), makes it possible to dampen the instability.[8] Harrod coined the term "instability principle" to describe possible adjustments between effective accumulation (i.e., that realized by entrepreneurs) and warranted accumulation. The incorporation of expectations then raises two issues: (i) Under what conditions are the entrepreneur's expectations realized? (ii) What happens if

[5] In Chapter "Expectations and its Variants: The Nuanced Role of Expectations in Classical Economics."

[6] See also on that point Harrod (1948, 1960).

[7] This sort of investment function was interpreted by Sen (1970) as an accelerator type and by Yoshida (1999) as an accelerator-multiplier type.

[8] In fact, in line with Harrod's own comments, the model (Bruno and Dal Pont Legrand 2014) shows that different types of dynamics patterns, including cyclical convergence to the steady state, growth cycle and corridor of stability can emerge.

these expectations are not realized? Although Harrod deals with the first question in detail, his treatment of the second is superficial and opened the way to numerous interpretations. Baumol (2000: 1039), for instance, concludes that "the main achievement of his [Harrod's] model lies in the ideas it inspired in those who did not fully understand it." Among the many different interpretations, most textbooks privilege Solow's view that due to the rigidity of the warranted rate of growth, the economy faced a so-called razor problem which vanishes with the assumption that technology is flexible.[9] Harrod persistently rejected that interpretation[10] and never considered the warranted growth rate G_w, as fixed but instead, times and again, insisted on the fact that G_w was a moving equilibrium over the different phases of the business cycle. Few contemporaries were aware that Harrod[11] was relying heavily on how entrepreneurs form and adapt their expectations (Baumol 1959; Hahn and Matthews 1964) and ultimately on the dynamics of investment. Nevertheless, it was not that road which was privileged. By introducing perfect substitutability between labor and capital in a world where whatever was saved was automatically invested, Solow proposed a model that was interpreted as a solution to Harrod's instability analysis. This functional approach determined the rate of growth of the economy and eliminated the need for an investment function. However, this modeling choice transformed Harrod's initial problem. Indeed, as pointed out by Hahn and Matthews (1964), it was necessary (Hagemann 2009: 85) to "[distinguish]clearly … between two different notions of knife-edge problem, the inequality between the warranted and the natural rates of growth and the instability of the warranted rate itself but also observed that Solow, in focusing exclusively on the first, had missed Harrod's emphasis on the second" (Hagemann 2009: 85).[12]

[9]Note that at that time, many economists interpreted Harrodian instability very 'negatively' and were always keen to dampen it (cf. Bruno and Dal Pont Legrand 2014: 467–468).

[10]He found the razor edge interpretation unpalatable; it was spread not only by neoclassical economists but also by Joan Robinson. Harrod never ceased to contest this interpretation, and in 1973, when he published his *Economic Dynamics*, he referred to the corridor concept proposed that same year by Axel Leijonhufvud in an article on effective demand failures, as being more appropriate. Bruno and Dal Pont Legrand (2014) follow this line of research and show that Harrod's dynamics was indeed compatible with the corridor concept in the context of growth cycles dynamics.

[11]More precisely, the instability was interpreted as an obstacle to growth and as a possible opportunity for business cycles; however, in both cases, it was considered excessive and requiring of dampening. The (only) person who seems properly to have understood Harrod's project was Jacob Marschak who made a clever comment in his 1939 article. For a more detailed analysis, see Sember (2010).

[12]In a private conversation with Muriel Dal Pont Legrand in 2009, Solow said that the real challenge was the investment function, and unfortunately, no one so far had come up with a satisfactory answer to this issue.

3 Getting Rid of Expectations: The Original Sin of Neoclassical Growth Models

The neoclassical approach was understood as a double departure from Harrod's initial contribution. The first departure can be summarized as rejection of Harrod's knife-edge which led macroeconomists to assume permanent equality between full employment saving and full employment investment. To "solve" Harrod's instability problem, Solow assumes full employment.

In addition, because his model ignored the demand side, it was at the origins of the dichotomy which emerged in modern macroeconomics between business cycles and growth theories, i.e., again between short- and long-run dynamics.[13] Solow confessed later that at the time, he had not realized that in building this model he was redirecting the growth research program substantially and had contributed to its definitive disentangling from business cycles analysis:

> it is clear to me that I oversimplified matters in 1956. The model was new, and I didn't understand all its implications. Some of what Harrod called instability is, of course, a matter of the behavior of the effective demand, off equilibrium paths. Harrod never specified clearly what he had in mind, and indeed there is very little literature even now that marries the theory of growth with effective demand. What I was getting at in 1956 was this: the special character of Harrods' model rests in the fact that the natural and warranted rates of growth are independent numbers… That characteristic of the model rests on fixed proportions. (It is immaterial whether Harrod believed that factor proportions are technically fixed or simply never change.) In turn, at least some aspects of "instability" arise because the economy is always being pulled away from the warranted path because it differs from the natural path. (Solow 1966,[14] cited by Boianovsky and Hoover 2009: 6)

Solow was not alone in proceeding this way. In fact, neoclassical growth theory and its so-called Solow–Swan–Meade type model relied on a plausible scenario explaining how the "economy" could be on the long-term equilibrium path. Swan (1963 [1970], 205) decided to assume "either that the authorities have read The General Theory or that they are socialists who did not need to." Meade made a similar assumption in the introduction to his 1961 book where he points out that his analytical framework was "based on the assumption of an ideally successful Keynesian policy which at every point of time manages to keep the value of investment at the desired level" (ix). Meade (1961) is even more explicit in adding:

> We desire to watch this system grow through time (…) But in this book we shall confine ourselves to watching this process of growth on the assumption that the growing system remains in equilibrium. We must explain (…) and this explanation can perhaps best be made in terms of the monetary system which we shall be assuming. (ibid.: 3)

He even provides a more detailed justification: "(…) we shall assume that there is banking system with a central bank and that the rate of interest is thereby always

[13] Cf. Bruno and Dal Pont Legrand (2014).

[14] Boianovsky and Hoover (2009) quote a letter from Solow to Ernst Helmstädter, dated February 11, 1966, from the Duke University Rare Book, Manuscript and Special Collections Library, Durham, N.C.

set at such level as to preserve a constant cost-of-living index… we assume that full employment of labor and land available at any moment of time is achieved by adjustment of the money wage per worker and the money rent per acre of land" (ibid.: 3).

Finally, Meade appears to be expressing some annoyance in confessing that by proceeding this way "we (they) are ignoring all the dynamic problems involved in ensuring that our economy does not leave the path of equilibrium growth" (fn. 1, p. 4). Setting aside short-run rigidities does not mean that those growth models ignored the importance of short-run disequilibria,[15] but it clearly is a non-neutral modeling strategy, i.e., it transformed how Harrod's (1939) definition of the line of steady advance was understood.[16] Clearly, business cycles were eliminated, and the research program focused on analysis of the long-run equilibrium path.

The second departure from Harrod's initial contribution was that the introduction of perfect substitutability between capital and labor resulted in a perfect adjustment between the warranted and natural rates of growth. The direct consequence of this hypothesis was that while the equilibrium growth path was only a possible solution in Harrod's and post-Keynesian models which was stressed also by Joan Robinson's "creeping platinum age,"[17] it became the unique solution and direction in Solow's model, i.e., in the so-called canonical neoclassical model.

However, it became obvious at a second stage that to complete the Solow–Swan approach required proof that the dynamic system was stable. Hahn (1960) analyzed the stability of the full employment growth equilibrium. The issue was revisited again by Hahn and Matthews (1964 [1972], p. 34) who thought that the dynamics needed to be based on the convergence process between G and G_w and needed also to show that the equilibrium path was stable both locally and globally. These proofs of stability were indeed seen as necessary conditions before one could be allowed to derivate implications for real life but also a more difficult task than the analysis of the equilibrium growth path properties (ibid. pp. 41–42). The point is to ensure that whatever the initial condition, there was a mechanism that would guarantee that the system was able to define and then to achieve a full employment growth equilibrium. They came to the conclusion there are situations for which the system is unstable (ibid. p. 42).

---

[15]This view was shared by Solow as evidence in his insistence on the fact that the absence of short-run disequilibria should not be interpreted as his willingness to neglect these issues: "It is not my contention that these problems don't exist, nor that they are of no significance in the long run" (Solow 1956: 91) but rather as an unavoidable distinction to clarify complex (and related) issues.

[16]The line of steady advance (i.e., the warranted rate of growth) appears in Harrod's 1936 book when he defines the line of steady advance as the growth path compatible with producer long-run equilibrium. However, at that time the "Harrodian cycle oscillates around the line of steady advance, it would appear that Harrod uses the steady growth path as a dynamic reference for his analytical framework and not as a medium which might formally link cycles and growth" (Bruno and Dal Pont Legrand 2014: 471). It was not until 1939 that the warranted rate of growth played a more active (dynamic) role.

[17]On this specific point, see Sen (1970: 21).

In various published and unpublished papers (Sen 1963a, b, 1964), Sen also paid attention to the stability issue. Sen (1963b) explored the possibility of price movements playing an equilibrating role in a context of a fixed money rate of interest. The neoclassical model appeared robust in that context, "the process does not make it possible to have warranted growth with full employment, but in the long run, the warranted rate of growth gives a stable proportion of unemployment" (Sen 1963b: 280). The second and more sensitive issue was how the neoclassical model behaves if the assumption of constant equality between the warranted and actual growth rates was relaxed. Sen (1963a) introduced an independent investment function based on an expected rate of growth in a Solow–Swan type model and finally obtained "the old-fashioned Harrodian problem" (ibid.: 278). Like Baumol[18] before him, Sen observed a clearly destabilizing effect emanating from lack of adaptation via entrepreneurs' expectations. Herein, for Sen (and others), lies the fundamental failure of the neoclassical growth theory:

> (…) the absence of an investment function and the consequent failure to assign a major role to entrepreneurial expectations about the future. It may be convenient at this point to recall that Harrod was concerned with the instability problem in equilibrium growth precisely in this context. The balance between 'warranted growth' and 'natural growth' is only one of Harrod's problems, and this is the only one that the Solow-Swan model takes up, leaving out the question of balance between 'warranted growth' and 'actual growth' related to entrepreneurial expectations. Once an independent investment function is introduced, the instability problem of Harrod quickly reappears in the Solow-Swan model, in spite of replacing the assumption of a constant capital-output ratio by a neo-classical production function (see Eisner 1958; Hahn 1960; Sen Reading 10). The assumption of substitutability does not seem to be a key difference between neo-classical and neo-Keynesian studies of growth, though it is sometimes thought to be so, and the main difference seems to lie in the investment function. (Sen 1970, p. 23)

In an unpublished paper, Sen (1964: 279) reiterates that "if growth theory is to have any relevance to policy, it cannot do without an investment function, and once that is given a fair play, it is easy to recognize that anything that reduces the 'knife-edge' balance between G_n and G_w will tend to highlight the 'knife-edge' balance between G and G_w."[19] So, fully aware of the two distinct (Harrodian) instability problems, Sen knew that they were not independent of each other. Sen (1963b: 277) confessed that what kind of investment function should be introduced was not clear to him since "this deals with one of the most untraceable elements in capitalist economy." Finally, he concludes: "The difficulty is usually concealed by doing without an independent investment function in the growth models and thereby by eliminating the influence of expectations. It is a dodge, and like all clever dodges, it has its usefulness, but it is easy to outlive that" (ibid. 280).

Samuelson had intensive exchanges with Sen on this issue. He was in quest of a consistent global view which would reconcile short-run Keynesian interventionism and long-run neoclassical growth analysis. Samuelson (1955) was convinced that the

[18] 1959.

[19] Sen (1963b) had observed that: "(…) anything that makes G_w move towards G_n will itself raise some questions about the equality of the actual rate of growth with G_w" (ibid. 280).

economy could not adjust automatically in a long-run equilibrium growth path, and that such a long-run equilibrium path could only be achieved through appropriate fiscal and monetary policies. He then explicitly refers to Harrod's dynamics issue as a problem which could "easily" be solved "(…) laissez faire harrodian discrepancies can lose much of their terror and relevance in a properly managed mixed economy" (1976: 754). In this way, he circumvented the difficulty related to introducing expectations.[20]

This view (or strategy) was questioned by Arrow (1967)[21] who rightly pointed out that even if such a mechanism worked and allowed the economy to achieve a full employment equilibrium path, there would be no guarantee that an economy which had achieved this path artificially would then behave like an economy which had reached it automatically, i.e., in the absence of economic policy. This was a crucial critique which reduces the scope of growth analysis cf. Assous et al. (2018).

4 Conclusions

This paper shows that in not incorporating expectations, early growth models broke with Harrod's project. Specifically, unpublished papers and correspondence reveal how the difficulties experienced by economists such as Solow, Hahn and Sen in formulating an autonomous investment function based on analysis of expectations led them to examine economic dynamics under the hypothesis of full employment and ultimately to admit that saving–investment coordination problem could be ruled out in the context of the long run.

This watershed moment which occurred in the 1960s reduced the scope of dynamics quite drastically. Following Solow's 1956 contribution, the literature had focused mostly on the behavior of equilibrium paths with business cycles analysis clearly disconnected from growth. Though partly responsible for it, Solow regretted that turn. In his Nobel Prize lecture, he stressed this explicitly:

> it is impossible to believe that the equilibrium growth path itself is unaffected by the short – to medium-run experience. In particular the amount and direction of capital formation is bound to be affected by the business cycle… So a simultaneous analysis of trend and fluctuations really does involve an integration of long run and short run of equilibrium and disequilibrium. (1987, 311–312)

There is no doubt that this separation between cycles and growth dynamics most likely contributed much to the success of Solow's growth model: "(…) theoretical attention shifted from fluctuations with growth to growth without fluctuations" (Punzo 2009, p. 101).

Building a growth model which was distinct from business cycles dynamics and without an independent investment function led Sen to think that the neoclassical

[20] See Assous et al. (2018).

[21] Arrows' comments (1967) were written at the occasion of the publications of Samuelson collected papers edited by Joseph Stiglitz.

growth research program needed to be redefined. He saw it necessary to consider three possible options (Sen 1970: 23–24): (i) the model describes the (real) working of the economy in which thanks to judicious government intervention "ex ante investment and ex ante savings are brought in line with each other" (ibid. 23); (ii) it provides "a description of the consequences over time of maintaining full employment, rather than a causal model which may have explained what would actually happen" (ibid.); and finally, (iii) the model is more useful "to trace a full employment path rather than describe what would in fact happens in a capitalist economy, with or without control" (ibid., 23–24). Sen concludes with the words: "This is less heroic but also less objectionable" (ibid.: 24).

References

Arrow, K. (1967). Samuelson collected. *Journal of Political Economy, 75*(5), 730–737.

Assous, M., Dal Pont Legrand, M., & Mansieri, S. (2018). *Walking a tightrope: The neoclassical synthesis in action* (presented at the STOREP Conference, 28–30 June 2018, GREDEG Working Paper) (Forthcoming).

Baumol, W. (1959). *Economic dynamics* (2nd ed.). New York: Macmillan Company.

Baumol, W. J. (2000). Book review of D. Besomi. *History of Political Economy, 32*(4), 1037–1039.

Besomi, D. (2000). On the spread of an idea: The strange case of Mr. Harrod and the multiplier. *History of Political Economy, 32*(2), 347–379.

Boianovsky, M., & Hoover, K. (2009). The neoclassical growth model and 20th century economics. In Introduction of the special issue *Robert Solow and the development of growth economics*. *History of Political Economy, 41*(Annual Suppl.), 1–23.

Bruno, O., & Dal Pont Legrand, M. (2014). The instability principle revisited: An essay in Harrodian dynamics. *European Journal of the History of Economic Thought, 21*(3), 467–484.

Eisner, R. (1958). On growth models and neo-classical resurgence. *Economic Journal, 68*(272), 707–721.

Hagemann, H. (2009). Solow's 1956 contribution in the context of the Harrod-Domar model. *History of Political Economy, 41*(Annual Suppl.), 67–87.

Hahn, F. (1960). The stability of growth Equilibrium. *The Quarterly Journal of Economics, 74*(2), 206–226.

Hahn, F., & Matthews, B. (1964). The theory of economic growth: A survey. *The Economic Journal, 74*(296).

Harrod, R. (1939). An essay in dynamic theory. *Economic Journal, 49,* 14–33.

Harrod, R. (1948). *Towards a dynamic economics*. London: Macmillan.

Harrod, R. (1960). Second essay in dynamic theory. *The Economic Journal, 70,* 277–293.

Harrod, R. (1973). *Economic dynamics*. London: Macmillan.

Harrod, R. (2003). The collected interwar papers and correspondence of Roy Harrod. Vol. 1: Correspondence 1919–35; Vol. 2: Correspondence 1936–39; Vol. 3: Essays and press items. In D. Besomi (Ed.), *The papers of Roy Harrod*. Cheltenham: Elgar. Available from: https://www.researchgate.net/publication/228646011_The_papers_of_Roy_Harrod. Accessed June 18, 2017.

Hicks, J. (1950). *A contribution to the theory of the trade cycle*. London, Oxford: Clarendon Press.

Hoover, K., & Halsmayer, V. (2016). Solow's Harrod: Transforming cyclical dynamics into a model of long-run growth. *European Journal of the History of Economic Thought, 23*(4), 561–596.

Meade, J. (1961). *A neo-classical theory of economic growth* (pp. ix + 146). New York: Oxford University Press.

Punzo, L. F. (2009). A nonlinear history of growth and cycle theories. *History of Political Economy, 41*(Annual Suppl.), 88–106.

Samuelson, P. (1955). *Economics* (3rd ed.). New York: MacGraw-Hill.
Samuelson, P. (1963a). A brief survey of post-Keynesian developments. In R. Lekachman (Ed.), *Keynes' general theory: Reports of three decades*.
Samuelson, P. (1963b). *Inconsistent and incomplete theories of distribution (mimeographed)*.
Samuelson, P. (1976). *Economics* (10th ed.). New York: MacGraw-Hill.
Samuelson, P. A., & Modigliani, F. (1966). The Pasinetti paradox in neoclassical and more general models. *The Review of Economic Studies, 33*(4), 269–301.
Sember, F. (2010). Closing the model? The Harrod–Marschak correspondence on the draft of the essay in dynamic theory. *Journal of the History of Economic Thought, 32*(4), 583–608.
Sen, A. (1963a). Neo-classical and neo-Keynesian theories of distribution. *Economic Record, 39*(85), 53–64.
Sen, A. (1963b). The money rate of interest in the pure theory of growth. In F. Hahn & F. Brechling (Eds.), *Theories of the rate of interest*. London: Macmillan.
Sen, A. (1964). *Full employment, perfect foresight and the neo-Keynesian distribution theory* (Unpublished).
Sen, A. (1970). Introduction. In A. Sen (Ed.), *Growth economics, selected readings*. Penguin Books.
Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics, 70*(1), 65–94.
Solow, R. M. (1988). Growth theory and after. *American Economic Review, 78*(3), 307–317.
Swan, T. W. (1963). Growth models: Of golden ages and production functions. In K. E. Berrill (Ed.), *Economic development with special reference to East Asia*. London: Macmillan.
Yoshida, H. (1999). Harrod's 'knife-edge' reconsidered: An application of the Hopf-bifurcation theorem and numerical simulations. *Journal of Macroeconomics, 21*(3), 537–562.
Young, W. (1989). *Harrod and his Trade Cycle Group: The origins and development of the growth research program*. London: Macmillan and New York University Press, USA.

Inflation Expectations and the Phillips Curve: Then and Now

Harald Hagemann

1 Inflation Expectations: Recent Importance and Interest

The Monetarist Counter-Revolution in the late 1960s and early 1970s is still remembered as being particularly important with respect to the role of inflation expectations in the Phillips curve discussion. While the originators of the Phillips curve are believed to having neglected inflation expectations, the introduction of adaptive inflation expectations by Phelps and Friedman at the end of the 1960s put them at the center of the discussion. The concept of rational expectations in the early 1970s soon followed and became a cornerstone of modern macroeconomics. Recent approaches to expectation formation, such as rational inattention and sticky information, however, are in some important elements much closer to the debates in the 1950s and 1960s than to the concept of rational expectations. This paper will analyze these links of the modern discussion to the debates over time, highlighting similarities, differences, and ways ahead regarding the formation of inflation expectations.

There can be no doubt that the Phillips curve was in the focus of many key debates in macroeconomics ever since Samuelson and Solow (1960) modified the original curve, in which Phillips (1958) had documented the empirical relationship between unemployment and the rate of change of money wages in the UK from 1861 to 1957, linking the unemployment rate to the inflation rate and bringing the curve to the attention of a wider audience among economists as well as extending to policy debates. Although they did not advocate explicitly a higher rate of inflation as a means of reducing unemployment, the idea of a price inflation-unemployment tradeoff seemed to follow. The modified Phillips curve of Samuelson and Solow illustrated the difficulties to simultaneously achieve the two central macroeconomic goals of high employment and price stability. According to James Forder's *Macroeconomics*

For valuable suggestions and comments on an earlier version I thank Johannes Schwarzer.

H. Hagemann (✉)
Institute of Economics (520), University of Hohenheim, 70593 Stuttgart, Germany
e-mail: hagemann@uni-hohenheim.de

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_8

and the Phillips Curve Myth (2014) and several other articles in recent years, there is not a piece of evidence that macroeconomists in the 1960s recommended to exploit an inflation-unemployment trade-off. He thereby challenges the widespread belief that there existed a permanent and stable negative trade-off between inflation and unemployment, which could be exploited for policy purposes. Helmut Schmidt, the later German Chancellor's statement, in the election campaign to the Bundestag in 1972 "that 5% inflation is better than 5% unemployment" is often interpreted that this "menu of choice" view some years later had reached the political sphere and that a social utility function in inflation and unemployment could be maximized. Forder attributes the genesis of this myth to Milton Friedman who, most clearly in his Nobel Prize lecture of 1977, given in a period of stagflation, when *both* inflation and unemployment increased, wanted to emphasize the originality and importance of his own presidential address on the role of monetary policy (Friedman 1968). After Friedman's presidential address, any postulated trade-off between unemployment and inflation was explicitly short-term, whereas the long-run Phillips curve was taken to be vertical at the natural rate of unemployment which later mutated to the non-accelerating rate of unemployment (NAIRU).

There exist some controversies about the amount Friedman (1968) and Phelps (1967) introduced inflationary expectations into the Phillips curve analysis. Surely the timing of their contributions was almost perfect since they were made at a time when wage and price inflation were increasing at a global level. Although Forder's statement that Phillips' 1958 article had only little contemporary impact is simply not true (as most students of economics in the 1960s could confirm), he is surely right that contemporary economists were not so stupid to confuse the difference between money and real wage increases, nor were trade unionists. The presence or absence of money illusion in the labor market was considered as an empirical question, including the existence of power relations.

Leeson (1997: 166) goes even as far as talking of the "Phillips–Friedman–Phelps critique," since Phillips in an earlier paper had examined the role of expectations about future price changes.[1] Leeson (1997: 171) is right in his concluding statement that Phillips was not an advocate of the proposition that ongoing and nontrivial inflation would purchase sustainable reductions in unemployment." However, in his famous 1958 article Phillips did not follow up the role of inflationary expectations, whereas Samuelson and Solow in their subsequent article reflect on the possibility that "a period of high demand and rising prices molds attitudes, expectations, even institutions in such a way as to bias the future in favor of further inflation" (1960: 193).

Although short-run and long-run expectations and their changes played already a central role in Keynes' *General Theory*, and expectations were considered as a key determinant for inflation, output, and employment ever since the hyperinflationary processes in many countries after World War I (which had a stronger impact on Austrian theories of the business cycle such as those of Mises and Hayek) or the deflationary period of the Great Depression in the early 1930s, in the Phillips curve

[1] See Phillips (1954: 311).

debate for a longer period there was no agreement how to model the evolution of expectations, due to practical difficulties, until a radical change took place with the so-called rational expectations revolution. Whether this was good for the economics discipline or economic policy is still discussed controversially.

Shortly before this conference, in September 2017 the outgoing chairwoman of the Federal Reserve Janet Yellen pointed out: "Key among current uncertainties are the forces driving inflation, which has remained low in recent years despite substantial improvement in labor market conditions."

In particular, she considered further research necessary regarding

- the natural rate of unemployment,
- the role and the measurement of inflation expectations, and
- the specification of the underlying framework for analyzing inflation dynamics (e.g., greater relevance of international output gaps).

"Perhaps most importantly, we need to know more about the manner in which inflation expectations are formed and how monetary policy influences them." (Yellen 2016, p. 11).

In the following, I first discuss different hypotheses on inflationary expectations in Sect. 2 before we engage in some historical reconstruction instead of a rational one in Sect. 3. Finally, in the spirit of the research procedure often followed by John Hicks, who wanted to indicate at the end of many major studies that he had not finished the job, some "ways ahead" are mentioned in the concluding Sect. 4.

2 Different Types of Inflation Expectations

In the following, we discuss different expectation hypotheses by "sophistication" or "rationality," not necessarily historical order:

- Static Expectations,
- Extrapolative Expectations,
- Adaptive Expectations,
- Rational Expectations.

2.1 Static Expectations

Static expectations imply a hypothesis that people expect the value of an economic variable, e.g., the inflation rate, in the next period to be equal to the current value of this variable, e.g., 2%, as many central banks, above all the European Central Bank, have as their inflation target today.

An ongoing error in predicting inflation is possible.

The assumption that individual actors are expecting the same inflation rate for next year as the current one is the simplest hypothesis to make. However, under certain circumstances as, e.g., a higher credibility of the central bank and a realization of its inflation target in the past, it might even be rational to expect the current value (particularly when it is close to the target inflation rate) for the future.

Interestingly a famous economist such as John Hicks made use of the "static expectations" assumption, "that, when the decision to adopt a particular technique for *new* processes is taken, the current wage is expected to remain unchanged"[2], when he analyzed the conditions for the maintenance of full employment when the dominant technique is changing during the traverse. Hicks was aware that the wage rate will vary during the adjustment process to a new technique and that his key assumption of static expectations therefore implies that these expectations turn out to be wrong and the resulting traverse is not an optimum one. Nevertheless, he took this rather unsatisfactory treatment of uncertainty as a good starting point for the analysis of the complexities of Ricardo's machinery problem. He held the view that more satisfactory assumptions on the formation of expectations "may well be best considered as deviations from it" (Ibid.)

Static expectations are an extreme case of adaptive expectations. Slightly less restrictive are

2.2 Extrapolative Expectations

Actual values of inflation in the previous periods determine expected inflation π_t^e for actual or next period with ε as weight is given to later periods → trend

$$\Rightarrow \pi_t^e = \pi_{t-1} + \varepsilon(\pi_{t-1} - \pi_{t-2})$$

However, since the expected inflation rate is a weighted average of the inflation rates of the past, this rather mechanical, entirely backward-looking rule leads to an ongoing error in predicting inflation when inflation is not stable.

2.3 Adaptive Expectations

The hypothesis takes into consideration learning on the basis of errors in predicting inflation in past periods, with δ as learning coefficient. If $\delta = 1$, then only the last period's rate of inflation counts.

$$\Rightarrow \pi_t^e = \pi_{t-1}^e + \delta\left(\pi_{t-1} - \pi_{t-1}^e\right)$$

[2]Hicks (1973, p. 56; see also p. 110). Hicks finished the manuscript of *Capital and Time* exactly at the time when he received the Bank of Sweden Prize in economics in memory of Alfred Nobel.

However, when inflation is not stable, an ongoing error in predicting inflation occurs, and the learning process is not efficient. A downward adjustment of inflation expectations is only possible after a decrease in the real inflation rates. Anti-inflation policies require time.

With adaptive expectations, the equation of the Phillips curve changes from

$$\pi = \pi^e - \beta(u - u^*) + v$$

to

$$\pi = \pi^e_{t-1} + \delta(\pi_{t-1} - \pi^e_{t-1}) - \beta(u - u^*) + v$$

u is the actual unemployment rate,
u^* is Friedman's natural rate of unemployment,
β is a reaction parameter, and
v is a supply shock parameter.

Figure 1 shows the natural rate expectations-augmented Phillips curve which became a common framework for macroeconomic analysis after the contributions by Friedman (1968) and Phelps (1967) until the early 1980s. In contrast to Tobin's (1972) presidential address "Inflation and Unemployment" five years later, who still argues in favor of a genuine long-run trade-off, after a full adjustment of inflation there exists no-trade-off in the long run. In the long-run, the Phillips curve is vertical at the "natural" rate of unemployment u^*. The natural rate of unemployment "separate[s] the real forces from the monetary forces" (Friedman 1968: 9) and is determined by "the actual structural characteristics of the labor and commodity markets" (Ibid: 8). Since according to Friedman any inflation is always and everywhere a monetary phenomenon, the natural rate u^* (which is considered as an equilibrium rate including only frictional unemployment) is not only compatible with price stability but also with any rate of inflation.

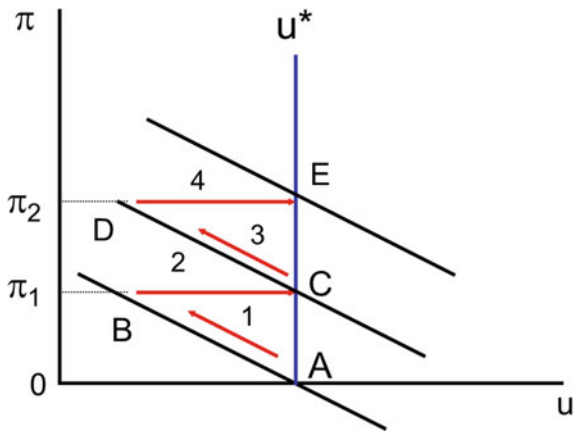

Fig. 1 Natural rate expectations-augmented Phillips curve

A trade-off exists only in the short run.

> [T]here is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off. The temporary trade-off comes not from inflation per se, but from unanticipated inflation, which generally means, from a rising rate of inflation (Ibid: 11).

Since adaptive expectations normally adjust to actual inflation with a time lag, policymakers could keep unemployment below its natural rate but only temporarily. Unexpected inflation, due to money illusion on the labor supply side (never a convincing argument) or the fixity of money wages during the term of the wage contract, would lead to lower real wages and thereby, according to neoclassical theory, reduce unemployment (movement from *A* to *B* in Fig. 1). However, upward correction in expected inflation would lead to higher increases in money wages in the next round and require a further increase in actual inflation to reduce the real wage and unemployment (from *C* to *D*) and so on. The result is the accelerationist proposition stating that only by accelerating the rate of inflation continually unemployment could be kept below the natural rate. However, you can fool some people sometimes but not all people all the time. Since economic agents would not systematically underpredict accelerating inflation and believe in a formula that is contradictory to the process how inflation is actually generated, it would be impossible for monetary policy to "peg the rate of unemployment for more than very limited periods" below the natural rate (Friedman 1968: 5). The Phillips curve becomes vertical in the long run.

Interestingly, in an insightful lecture "Monetary Theory and History—An Attempt at Perspective" given at the University of Western Australia in Perth in February 1967, i.e., ten months before Friedman gave his presidential address to the AEA in New York, John Hicks had not only pointed out the validity (and limitation) of the quantity theory as a long-run equilibrium condition and the fact that an increase in the supply of money could give a real stimulus to the economy in the short run, but also came to the following conclusion:

> Inflation does give a stimulus, but the stimulus is greatest when the inflation starts – when it starts from a condition that has been non-inflationary. If the inflation continues, people get adjusted to it. But when people are adjusted to it, when they *expect* rising prices, the mere occurrence of what has been expected is no longer stimulating. Nor can the fade-out be prevented by accelerating the inflation; for acceleration of inflation can be expected too. (Hicks 1967, pp. 162–3).

As a consequence of adaptive expectations, a successful disinflationary policy, as practiced in the USA after Paul Volcker took office as the Chairman of the Federal Reserve in August 1979, requires to force actual inflation below expected inflation. The achievement of disinflation thereby is a very costly process since the contractionary measures lead to a strong increase in the real interest rate (in a way comparable to the deflation during the Great Depression of the early 1930s), with the consequence of an increase in bankruptcies and unemployment rates much above its "natural" level.

As a consequence of the perceived flaws in approaches based on adaptive expectations and their unreliability in economic policy, the modeling of expectations in

economic theory was increasingly replaced by the hypothesis of rational expectations, originally developed by Muth (1961), which became influential after it was applied by Robert Lucas Jr. in macroeconomics.

2.4 Rational Expectations

Rational expectations depend on all information available for the formation of inflation expectations. Furthermore, it is also assumed that decision-makers know the structure of the economic system. Expectations may be wrong, but there does not exist any ongoing, systematic error in predicting inflation. However, rational expectations do not imply perfect foresight. Although the future certainly is not fully predictable and uncertainty remains, expectations of economic agents are not systematically biased.

$$\prod_t^e = \prod_t$$

only holds as long as there are no exogenous shocks. As a result, not only the long-run Phillips curve would be vertical but also the short-run or expectations-augmented Phillips curve. There would be no room for monetary policies to reduce unemployment. (Lucas 1972). If the central bank would attempt to reduce unemployment below the natural rate through an expansionary monetary policy, economic agents would raise their expectations of future inflation accordingly, thereby counteracting the intended effect of monetary policy. In the presence of this policy-ineffectiveness proposition, a reduction of the unemployment rate would require an error in the predicted inflation rate.

3 Historical Origins

In their recent re-assessment of Milton Friedman's presidential address, "The Role of Monetary Policy," at the occasion of its 50th anniversary, Mankiw and Reis (2018) point out that the two major themes are the focus on expectations and its emphasis on the long run which is characterized by monetary neutrality as in classical economics. With its "aims to undermine the eclectic American Keynesianism of the 1950s and 1960s" (Solow 2018: 421), "for Friedman, expectations were the key to explaining how the economy might appear to face a Phillips curve trade-off and how that trade-off would disappear if we tried to exploit it" (Mankiw and Reis 2018: 84). The introduction of the "natural rate of unemployment", i.e., an equilibrium rate at which the labor market is cleared, along with the long-run vertical Phillips curve and the accelerationist hypothesis, would destroy "Keynesian" ideas of reducing unemployment (beyond full employment in Friedman's view) by creating unanticipated

inflation. Since monetary policy was an inadequate instrument to reduce the unemployment rate in the long run, no policy dilemma between full employment and price stability would exist. The alleged trade-off described by the Phillips curve is only a transitory one, and unemployment must in the end return to its natural rate. Friedman was critical of the earlier debate of cost-push versus demand-pull inflation. For him, any inflation was caused by an excessive increase in the quantity of money. He argued forcefully for rules rather than discretion, i.e., the best way to avoid inflation would be a steady growth in some monetary aggregate.

Friedman's natural rate hypothesis and the expectations-augmented Phillips curve captured the main attention of macroeconomists for quite some time. Although Friedman and Phelps (1967, 1968), whose focus on involuntary unemployment and account of the Phillips curve trade-off is much closer to the pre-Friedmanite "dilemma view"[3], assumed adaptive expectations they contributed to opening doors for the subsequent rational expectations revolution in macroeconomics, which "certainly helped lead macroeconomics to its current state of refined irrelevance" (Solow 2018: 424). The trade-off view being vulnerable to the expectations argument, and attributed as an innovation to Friedman and Phelps, increasingly entered macroeconomics textbooks thereafter, as Forder (2015) has shown in detail. Forder judges this story about Keynesian acceptance of the exploitable Phillips curve in the 1960s as a fictitious story which came up in the 1970s. Phillips himself never reacted to the work of Friedman and Phelps. However, a careful reading of his work shows clearly that he never advocated inflation as a means of reducing unemployment (Schwarzer 2012).

Another fictitious story is the view that there exists a one-way avenue of progress from static via extrapolative and adaptive expectations to rational expectations. This order of sophistication of expectations surely is not identical with the historical order. Phillips (1954) had examined the role of adaptive expectations. However, Lesson (1997: 166) overstates his case when he talks of the "Phillips–Friedman–Phelps critique," since Phillips neither did include inflation expectations as a shift parameter in the analysis of his curve nor did he draw similar policy conclusions as Friedman and Phelps did later at a time of rising inflation rates.

The statistical relation between unemployment and price changes was already addressed by Irving Fisher (1926) at a time when he invented the concept of distributed lags to capture the output, employment, and real interest rate effects of slowly adjusting inflation expectations. Fisher's contribution went largely unnoticed until it was republished in the high years of the Phillips curve debate (Fisher 1973).

Cagan (1956) is widely seen as the influential early modern propagator of the adaptive expectations hypothesis. When studying the behavior of the velocity of money in hyperinflation periods with Milton Friedman in Chicago, Cagan was expressing expected inflation as a weighted average of past inflation rates, with exponentially declining weights, thus converting Phillips' suggestion to relate changes in the expected rate of inflation to the difference between the actual and the expected inflation rate into the analysis of the demand for money. Cagan (2000, p. 22) later

[3]For greater details see Schwarzer (2016, 2018).

explicitly pointed out that Phillips deserves credit to be the originator of the adaptive expectations hypothesis.[4]

Inflation expectations were indeed discussed years before the contributions by Friedman and Phelps. A good example is Bronfenbrenner's (1963) survey of the research papers presented at the Commission on Money and Credit. Here the "Phillips" problem, stated as "How much unemployment…is required to keep money wages from rising more rapidly than labor productivity, and thus imparting to the economy an element of cost inflation" (Bronfenbrenner 1963, p. 115), is discussed intensively. It is clearly pointed out that inflationary policy will lead to the consequence that persistent inflation becomes prominent in the standard expectational pattern of economic agents contributing to the erosion of money illusion.

Erik Lindahl clearly anticipated Friedman's "accelerationist hypothesis," as has been emphasized first by Laidler (1999, p. 60 n. 10) and elaborated by Boianovsky and Trautwein (2006) on the basis of analyzing Lindahl's (un) published writings from the 1920s to the 1950s. Although Lindahl, who in the Wicksellian tradition was concerned with the non-neutrality of monetary policy throughout his academic life, advocated monetary policy rules, he strongly rejected concepts of natural rates, marking thereby a decisive difference to Friedman and Phelps.

It was Friedman who in his presidential address transferred Wicksell's conception of the natural of interest to the labor market and invented the concept of the "natural rate of unemployment." With the Friedman–Phelps approach, the labor market continuously clears and reductions in unemployment can only occur temporarily in response to "fooling" of workers who have money illusion. Inflation expectations now enter the labor supply function, whereas before they were mainly discussed within the demand side of the goods or money market—"flight from cash" (Phillips 1954, Brown 1955), Cagan (1956)—within firms (Scitovsky 1941) or financial markets (Fisher 1926).

Since the natural rate revolution of Friedman and Phelps and the subsequent work by Lucas and Rapping (1969a, b) expectations are particularly related to the aggregate supply function of the labor market, whereas in the "New Keynesian" Phillips curve expectations are mainly considered on the side of firms (Clarida et al. 1999). The Lucas approach had in common with Friedman and Phelps that movements of output and employment away from the natural level required an unanticipated component in the rate of inflation. Rational expectations, as introduced into the Phillips curve debate by Lucas (1972, 1973, 1976), implied that, in the absence of exogenous shocks, there would not exist a trade-off between unemployment and inflation and even the short-run Phillips curve would be vertical. The problem, however, was less the rational expectations hypothesis than the twin assumption of permanent market-clearing, the "macroeconomics of Dr. Pangloss" (Buiter 1980).

[4]For a different view that even Phillips could not claim to be the originator of the expectations critique see Forder (2010).

4 Ways Ahead: Whose Expectations?

Further debates on the rational expectations hypothesis include the following interesting extensions:

- Mankiw (2001): "There is a simple way to reconcile the new Keynesian Phillips curve with the data: adaptive expectations."
  - Problem for rational expectations: data suggest that expectations still are backward-looking → inflation inertia.
- Learning approach (Evans and Honkapohja 2001).
  - Individuals know the model how the economy works,
  - but have to econometrically estimate the parameters and reestimate these, e.g., after policy changes.
- Sticky information (Mankiw et al. 2003).
  - New information is not immediately available but distributed similarly to a Calvo lottery (Calvo 1983).
  - Inertia in inflation due to outdated information.
- Rational inattention (Sims 2003).
  - Processing information is costly.
  - New information only processed if significant reward is possible.
  - Static inflation expectations in the extreme (belief in the central bank target).
  - Close to near-rational approach of Akerlof et al. (2000).

The modified Phillips curve expressing a significant correlation between unemployment and inflation over some time horizons has become an important concept in macroeconomic analysis and policymaking. But today's Phillips curve is not the same as the Phillips curve six decades ago. This is quite natural since the economy, the determinants of inflation and the role of monetary policy as well as our understanding of wage- and price-setting behavior, has evolved significantly over time. For example, Miles et al. diagnosed two conjoined puzzles appearing after the Great Recession, namely the "missing disinflation" and the "missing inflation" puzzle, i.e., "that there was no large fall in inflation after unemployment peaked in 2009–10; and second, that there was no large increase in inflation when the recovery was well under way in 2015–16" (Miles et al. 2017, p. 27). Since Friedman a shift in a rule-based monetary policy has taken place away from a focus on monetary aggregates to Wicksell's concept of a natural rate of interest, due to the works of Taylor and Woodford. The concept of the expectations-augmented Phillips curve allows the relationship between unemployment and inflation to shift due to changes in inflation expectations. As Phelps (2017) has pointed out recently, there is "nothing natural about the natural rate of unemployment." The term "natural" is misleading because the "natural rate" itself is not a constant but can be pushed up or pulled down due to structural changes in the economy. Thus, a higher or lower rate of technical progress

or population growth can move away the "natural rate" from its former level. However, Phelps (2017, p. 1) insists that "the 'natural rate' always exerts its centripetal force."

The rational expectations revolution has changed the idea of an exploitable trade-off. The proposition of ineffectiveness of stabilization policy and the "Lucas critique" are a sequel of rational expectations. According to Lucas rational, optimizing agents will change their behavior when they expect discrete changes in monetary or fiscal policy, thereby changing the relationships defining aggregate behavior in macroeconomic models.

Since Friedman's presidential address, enhanced by the subsequent work of Lucas, a long-run equilibrium framework has become the benchmark for macroeconomic analysis. Although policy announcements and rules, particularly credible ones, will influence the behavior of economic agents, nevertheless stabilization policies in the real world continue to have real effects. Tobin (1980) has shown convincingly that the assumption of continuous market-clearing is the crucial one for the results of new classical macroeconomics. The financial and economic crisis of 2007–9 that started on the financial markets, which usually are supposed to be the most efficient ones because they are dominated by well-informed actors having rational expectations, has demonstrated that a wide gap exists between the "reality" of these types of macroeconomic models and the real world. The reaction speed to policy changes will increase if economic agents have rational expectations. However, prices do not continuously clear all markets. Room therefore remains for quantity-rationing approaches and Keynesian economics featuring disequilibrium dynamics.

Expectations play a crucial role in any modern macroeconomic theory or policy. Anchoring inflation expectations is a major precondition for keeping inflation under control. This is best shown by the case of the European Central Bank, the central bank with the strictest inflation target, which since its foundation came rather close to its target rate of 2% per year.

In his historical reconstruction of the Friedman-Phelps expectations critique Forder concludes that "it was clear from the beginning that there was no practical way to be sure how to model the evolution of expectations" (Forder 2010: 507). Friedman had already indicated that expectations often change rather slowly which would be an argument supporting the adaptive expectations hypothesis. Half-a-century later Janet Yellen pointed out that "historical experience suggests that changing the public's inflation expectations would be neither quick nor easy" (Yellen 2016, p. 12). As inflation expectations were so widely discussed long before Friedman and Phelps, was there a consensus whose expectations are deemed to be important? Is there a common ground whose expectations are driving the economy? Seemingly not, as the following statement by Yellen (Ibid.) indicates.

> Yet another unresolved issue concerns *whose* expectations—those of consumers, firms, or investors—are most relevant for wage and price setting, a point on which theory provides no clear-cut guidance. More generally, the precise manner in which expectations influence inflation deserves further study.

In the Phillips curve retrospective after fifty years, Solow (2009) has argued in a similar way when referring to various interest groups in the economy—bankers, investors, consumers, savers, lenders, borrowers, buyers, sellers, etc.—and the fact that there would be no reason for the economic agents from these various groups to react in a fundamentally same way. So in the real world beyond the fictitious "representative agent," there remains the crucial problem of how to aggregate expectations. There are differences in relevance so that one may attribute different weights to the expectations of the various groups. Whose expectations? is one of the relevant topics for future research, as is the question of the underlying framework for inflation dynamics and the analysis of how the structure of models may be learned by economic agents and how expectations may be formed.

References

Akerlof, G., Dickens, A. W. T., & Perry, G. L. (2000). Near-rational wage and price setting and the long-run Phillips curve. *Brookings Papers on Economic Activity, 31*(1), 1–60.

Boianovsky, M., & Trautwein, H.-M. (2006). Price expectations, capital accumulation and employment: Lindahl's macroeconomics from the 1920s to the 1950s. *Cambridge Journal of Economics, 30*(6), 881–900.

Bronfenbrenner, M. (1963). A sample survey of the commission on money and credit research papers. *The Review of Economics and Statistics, 45*(1), 111–128.

Brown, A. J. (1955). *The great inflation, 1939–1951*. London, New York, Toronto: Oxford University Press.

Buiter, W. (1980). The macroeconomics of Dr. Pangloss. A critical survey of the new classical macroeconomics. *Economic Journal, 90*(357), 34–50.

Cagan, P. (1956). The monetary dynamics of hyperinflation. In M. Friedman (Ed.), *Studies in the quantity theory of money* (pp. 25–117). Chicago: University of Chicago Press.

Cagan, P. (2000). Phillips adaptive expectations formula. In R. Leeson (Ed.), *A. W. H. Phillips: Collected works in contemporary perspective* (p. 22). Cambridge: Cambridge University Press.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics, 12*(3), 383–398.

Clarida, R., Gali, J., & Gertler, M. (1999). The science of monetary policy: A new Keynesian perspective. *Journal of Economic Literature, 37*(2), 1661–1707.

Evans, G. W., & Honkapohja, S. (2001). *Learning and expectations in macroeconomics*. Princeton: Princeton University Press.

Fisher, I. (1926). A statistical relation between unemployment and price changes. *International Labour Review, 13*(6), 785–792.

Fisher, I. (1973). I discovered the Phillips curve. *Journal of Political Economy, 81*(2), 496–502.

Forder, J. (2010). The historical place of the 'Friedman–Phelps' expectations critique. *The European Journal of the History of Economic Thought, 17*(3), 493–511.

Forder, J. (2014). *Macroeconomics and the Phillips curve myth*. Oxford: Oxford University Press.

Forder, J. (2015). Textbooks on the Phillips curve. *History of Political Economy, 47*(2), 207–240.

Friedman, M. (1968). The role of monetary policy. *The American Economic Review, 58*(1), 1–17.

Friedman, M. (1977). Inflation and unemployment. *Journal of Political Economy, 85*(June), 451–472.

Hicks, J. (1967). *Critical essays in monetary theory*. Oxford: OUP.

Hicks, J. (1973). *Capital and time. A neo-Austrian theory*. Oxford: Clarendon Press.

Laidler, D. (1999). *Fabricating the Keynesian revolution studies of the inter-war literature on money, the cycle, and unemployment*. Cambridge: Cambridge University Press.
Leeson, R. (1997). The trade-off interpretation of Phillips's dynamic stabilization exercise. *Economica, New Series, 64*(253), 155–171.
Lucas, R. E. (1969). Real wages, employment, and inflation. *The Journal of Political Economy, 77*(5), 721–754.
Lucas, R. E., Jr. (1972). Expectations and the neutrality of money. *Journal of Economic Theory, 4*(2), 103–124.
Lucas, R. E., Jr. (1973). Some international evidence on output-inflation tradeoffs. *American Economic Review, 63*(2), 326–334.
Lucas, R. E., Jr. (1976). Econometric policy evaluation: a critique. In K. Brunner & A. H. Melzer (Eds.), *The Phillips curve and labor markets, Carnegie-Rochester conference series on public policy* (Vol. 1). Amsterdam: North-Holland.
Lucas, R. E., & Rapping, L. A. (1969). Price expectations and the Phillips curve. *The American Economic Review, 59*(3), 342–350.
Mankiw, N. G. (2001). The inexorable and mysterious tradeoff between inflation and unemployment. *The Economic Journal, 111*(471), C45–C61.
Mankiw, N. G., & Reis, R. (2018). Friedman's presidential address in the evolution of macroeconomic thought. *Journal of Economic Perspectives, 32*(1), 81–96.
Mankiw, N. G., Reis, R., & Wolfers, J. (2003). *Disagreement about inflation expectations*. National Bureau of Economic Research Working Paper Series No. 9796 (June).
Miles, D., Panizza, U., Reis, R., & Ubide, Á. (2017). *And yet it moves: Inflation and the great recession*. Geneva Reports on the World Economy 19. Geneva and London: International Center for Monetary and Banking Studies (ICMB) and Centre for Economic Policy Research (CEPR).
Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29,* 315–335.
Phelps, E. S. (1967). Phillips curves, expectations of inflation and optimal unemployment over time. *Economica, 34*(135), 254–281.
Phelps, E. S. (1968). Money-wage dynamics and labor-market equilibrium. *The Journal of Political Economy, 76*(4), 678–711.
Phelps, E. S. (2017). *Nothing natural about the natural rate of unemployment*. New York: Project Syndicate.
Phillips, A. W. H. (1954). Stabilisation policy in a closed economy. *The Economic Journal, 64*(254), 290–323.
Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wages in the United Kingdom, 1861–1957. *Economica, 25,* 283–299.
Samuelson, P. A., & Solow, R. M. (1960). Analytical aspects of anti-inflation policy. *American Economic Review, 50*(2), 177–194.
Schwarzer, J. A. (2012). A.W. Phillips and his curve: Stabilisation policies, inflation expectations and the 'menu of choice'. *The European Journal of the History of Economic Thought, 19*(6), 976–1003.
Schwarzer, J. A. (2016). Price stability versus full employment: The Phillips curve dilemma reconsidered (PhD thesis). University of Hohenheim, Stuttgart.
Schwarzer, J. A. (2018). Cost push and demand-pull inflation: Milton friedman and the 'Cruel Dilemma'. *Journal of Economic Perspectives, 32*(1), 195–210.
Scitovsky, T. (1941). Capital accumulation, employment and price rigidity. *The Review of Economic Studies, 8*(2), 69–88.
Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics, 50*(3), 665–690.
Solow, R. M. (2018). A theory is a sometime thing. *Review of Keynesian Economics, 6*(4), 421–424.
Solow, R. M., Taylor, J. B., & Mankiw, N. G. (2009). Fifty years of the Phillips curve: A dialog on what we have learned. In J. C Fuhrer, Y. K. Kodrzycki, J. S. Little, & G. P. Olivei (Eds.),

Understanding inflation and the implications for monetary policy: A Phillips curve retrospective. Cambridge, MA and London: The MIT Press.
Tobin, J. (1972). Inflation and unemployment. *American Economic Review, 62*(1), 1–18.
Tobin, J. (1980). *Asset accumulation and economic activity*. Oxford: Basil Blackwell.
Yellen, J. L. (2016). *Macroeconomic research after the crisis*. Speech at "The Elusive 'Great' Recovery: Causes and Implications for Future Business Cycle Dynamics" 60th annual economic conference sponsored by the Federal Reserve Bank of Boston, Boston, Massachusetts, October 14. https://www.federalreserve.gov/newsevents/speech/yellen20161014a.htm.
Yellen, J. L. (2017). *Inflation, uncertainty, and monetary policy*. Speech at the "Prospects for Growth: Reassessing the Fundamentals" 59th Annual Meeting of the National Association for Business Economics, Cleveland, Ohio, September 26. https://www.federalreserve.gov/newsevents/speech/yellen20170926a.htm.

Part III
Expectations and Applications: Economic Models and Case Studies

The War that Bond Markets Did not Perceive as Such

The Prices of South African Bonds During the Second Anglo-Boer War: An Extreme Case of Resilience

Kim Oosterlinck and Marie Van Gansbeke

1 Introduction

War is a well-recognized condition increasing the risk of sovereign default. States may differ in the way they finance the war but debt and money issue usually represent an important proportion in terms of war finance. For France and the USA, during the two world wars, the share of taxation always represented less than 48% with a low of 4% for France during World War I (Occhino et al. 2008). At the end of the nineteenth century, economists were already trying to determine how to optimally finance wars. Printing money was clearly the worst option but "the choice between taxes and borrowing was far from obvious" (Kang and Rockoff 2015, p. 48). Whatever the exact proportion covered by taxation, the increase in money supply linked to the issue of additional bonds generally led to sharp declines in bond prices. Quite logically fixed-income securities were not considered as interesting investments in an inflationary environment. The expected length of the war as well as its outcome also affected bond prices. A longer war was likely to lead to higher costs and losing the war could mean the country would have to pay reparations, a frequent outcome

The authors thank the Rothschild Archive in London as well as J. Cavernelis-Frost and N. Broad for their help in collecting archival records on the 1892 Rothschild Loan and the Gemeente Amsterdam Stadsarchief as well as G. Pravilovic and K. Maarseveen-Helderman for their help in collecting archival records of the Bank Insinger Archives. We also show gratitude to the Stichting Capital Amsterdam and Cherelt Kroeze for granting access to the OPC-Database. We also thank Edward Kerby for his help regarding South African data and Guy Mélard for suggestions.

K. Oosterlinck (✉) · M. Van Gansbeke
SBS-EM, CEB, Université Libre de Bruxelles, 50 av. Roosevelt,
CP 114/03, 1050 Brussels, Belgium
e-mail: koosterl@ulb.ac.be

M. Van Gansbeke
e-mail: marie.vangansbeke@gmail.com

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_9

in the past. Additionally, wars could lead to revolution followed by repudiation as was the case for Russia in 1917 (Oosterlinck and Landon-Lane 2006; Oosterlinck 2016).

Long-run empirical evidence confirms that warfare has indeed a significant and negative impact on sovereign credit risk (Dincecco 2009). The negative effect of wars on sovereign bond prices is thus almost a general phenomenon. This need not mean that yields always reached extremely high values. During the Russo-Japanese war, even though Japan was expected to end up losing the war, the yields on its bonds remained relatively low (Sussman and Yafeh 2000). Military interventions conducted to force reimbursement represent the only systematic exception to this rule (Mitchener and Weidenmier 2005; Oosterlinck 2013; Oosterlinck et al. 2014).[1] Unsurprisingly, bonds issued by the losing party tend to trade at even lower prices and end up sometimes being viewed as almost valueless. Confederate bonds traded at few percentages of par at the end of the US Civil War (Mitchener et al. 2015). The same holds for German and Austrian bonds traded in Amsterdam for which current yields spiked at the end of World War I (Jopp 2016). At the end of World War II, German and Austrian bonds traded at 20% of par in Switzerland (Frey and Kucher 2001). For bonds traded in the enemy's stock exchange, the decline was even more severe. Brown and Burdekin (2002) report that when World War II broke out, the German Dawes and Young bonds immediately fell to more or less 5% of par.

The impact of war-related events has been investigated in a large panel of conflicts including the American Civil War (Willard et al. 1996; Brown and Burdekin 2000; Weidenmier 2002), WWI (Adams 2015; Jopp 2016), WWII (Frey and Kucher 2000, 2001; Frey and Waldenström 2004, 2008; Brown and Burdekin 2002; Hudson and Urquhart 2015), the Russian Revolution (Landon-Lane and Oosterlinck 2006) or the Israeli–Palestinian conflict (Zussman et al. 2008). These empirical studies all suggest important war news could have structural effects on sovereign bonds prices. For instance, Frey and Kucher (2000) find the "official" outbreak of WWII in September 1939 structurally impacted sovereign bond prices on both sides of the conflict. Bond prices also dropped when countries became subject to a foreign occupation. Evidence also suggests war events may affect the investors' anticipations in different ways. Antietam and Gettysburg battles, known as major Union victories during the American Civil War, substantively modified the course of both Union and Confederate greenback and greyback markets (Willard et al. 1996; Weidenmier 2002). However, investors interpreted Antietam as indicative of increased cost of war while Gettysburg was perceived as changing the fortunes of the different parties to the conflict (Weidenmier 2002).

This article deals with a government bond whose evolution contrasts with the general behaviour of sovereign bond markets during periods of war. First, over the whole Second Anglo-Boer War (1899–1902), the prices of the main foreign bond

[1]Exceptions to this rule are indeed hard to find for wars which were not waged to get bonds reimbursed. Vizcarra (2009) provides an analysis of such an exception: yields of Peruvian bonds remained low despite the wars because a third-party guaranteed the payment of the coupons on basis of the sales of Peruvian guano.

issued by the South African Republic—the 1892 SAR bond—remained at remarkably high levels. As a matter of fact, the bond price never fell below 96% of par and this despite the fact that the South African Republic defaulted on its bond well before the war's end. Second, the bond price was almost unaffected by war news. Third, at the end of our sample period, yields on these bonds enjoyed a decreasing trend. How can these observations be reconciled with the literature? We argue that, even though it might seem surprising at first sight, bondholders considered the war outcome as irrelevant. They acted on the belief that should the South African Republic loose the war it would become a British colony and enjoy the guarantee of the British Crown. On the other hand, should the South African republic remain independent, it would pay its dues without problem. Gold discoveries in South Africa meant indeed that its credit risk was perceived as negligible. Furthermore, should any problem arise, investors may have acted on the belief that the underwriter of the bond, the Rothschild, would be inclined to step into preserve its reputation, an element consistent with the findings of Flandreau and Flores (2009). This triple source of potential payers was in all likelihood perceived as a form of insurance by market participants.[2] With the end of the war, it became clear that the South African Republic would become a British colony. Even though bondholders expected Great Britain to take over the South African debt, archival evidence shows this was by no means a foregone conclusion. Prices moved upward at the end of the war when details regarding the treatment of the South African debt became known. In this respect, our analysis highlights the importance of legal discussions (and settlements) when state successions occur highlighting the relevance of law for sovereign debts as suggested by Weidenmaier and Gulati (2016).

To develop our point, the paper proceeds as follows. Section 2 provides the political and economic context. Section 3 outlines the SAR financing history and presents the data. Section 4 discusses the results and reviews the conditions of the 1892 SAR bond transfer. Section 5 provides final conclusions.

2 The South African Republic on the Road to Annexation

The UK firmly established its control over the formerly Dutch ruled Cape in 1806. At the time, its main interest in the territory was its use as a naval base (Pretorius 2009). In the 1820s, Great Britain sponsored a first wave of British settlements and undertook structural reforms to convert the Cape into a high standard English-speaking colony (Meredith 2007). Together with the abolishment of slavery in 1834, these new policies inspired deep indignation amongst the Boer population. A massive exodus of Boer families across the Orange River border started as from 1836 (Meredith 2007). As the Boers progressively extended their territorial claims, the British Empire

[2] Other bonds have experienced relatively high prices because of the same perception. Russian bonds following their repudiation in 1917 provide a striking parallel (Oosterlinck and Ureche-Rangau 2008; Oosterlinck 2016).

stepped in and annexed their newly created Republic of Natalia in 1843 (Pretorius 2009). Most Boer families emigrated again instead of risking falling under British rule (Meredith 2007). The British Empire soon abandoned its interventionist policy beyond its colonies' borders in front of the increasing cost of maintaining peace in the region (Meredith 2007). At the Sand River convention in 1852, the UK recognized the independence of "the Emigrant Farmers" in the territory north of the Vaal River. The Boers named their new state the South African Republic (SAR). In Great Britain, the new state became known as the Transvaal. Two years later, Great Britain acknowledged the independence of the Orange Free State at the Bloemfontein convention (Meredith 2007). Until the 1870s, both Boer Republics remained highly unstable with a disorganized administration, and an almost complete dependence on subsistence farming (Meredith 2007).

Two successive discoveries fundamentally disrupted the patterns of economic development in the region (van Helten 1982). In 1867, the world's richest ores of diamond were discovered in Griqualand, just outside of the Cape colony's borders. This prompted Great Britain to annex the concerned territories (Meredith 2007). This move was followed ten years later by a proclamation declaring the annexation of the South African Republic. By doing so, Great Britain hoped to create a unified South Africa under British rule (Lugan 2010). The escalating Boer resentment led to the First Boer War in 1880–1881 (Lugan 2010). Following an unexpected British defeat, the South African Republic was recognized as a fully independent entity at the London convention in 1884 (Wilson and Thompson 1969).

In 1886, the discovery of the largest gold reef worldwide in the South African Republic disturbed the fragile regional equilibrium (Lugan 2010). As a short-term response, the British government attempted to isolate the Boer Republics to make them dependent on British Cape and Natal colonies (Lugan 2010). But the gold discovery also strongly reignited the British willingness to unify South Africa under British rule (van Helten 1982). Two elements distinguished the SAR gold discovery from the diamond ores ones: its magnitude and importance on one hand and the structure of capital and organization required to exploit it on the other hand (Graham 1996). The gold discovery in the South African Republic coincided with the emergence of gold as the basis for an international monetary system. With international trade expansion, an increase in the world gold supplies was at the time necessary to the enlargement of the overall monetary stock (van Helten 1982). In this context, the South African Republic was taking a crucial importance as the world's largest gold producing area. Exploiting the gold required further digging and treating for extraction (Lugan 2010). Such investment demanded an organized and substantial structure of capital (van Helten 1982), which led to massive inflow of international capital and to the emergence of *Randlords*, the rich English-speaking entrepreneurs controlling the South African Republic mines (van Helten 1979).

Following the gold discovery, many European powers began vying for control of the South African mines. Germany progressively grew as Britain's most serious rival, with German–South African Republic trade growing more than threefold in value

over the 1893–1897 period.[3] British officials also interpreted the increasing German commercial penetration in the South African Republic as closely related to its alleged political ambitions in the region. Diverging economic interests in all likelihood paved the way of the Second Anglo-Boer War (Blainey 1965; van Helten 1978). Modern historians consider the failed Jameson Raid, which took place in December 1895, as the major turning point which would lead to the war outbreak. The Raid was aimed at sponsoring an uprising of the non-Boer white South African Republic inhabitants, known as *uitlanders*. This uprising and its expected repression would then serve as a pretext for military intervention to restore order and take control of the South African Republic (Lugan 2010). Cecil Rhodes, the Prime Minister of the Cape Colony and a prominent *Randlord*, is suspected to have initially framed this plan (Pakenham 1979; Lugan 2010). The Raid ended up in total failure as the Boer administration got informed of the plot before it was launched. The Jameson Raid led the Boer Republics to conclude a military pact in 1897 in a joint desire to oppose British imperialism. Furthermore, it strengthened the pro-Boer sentiment in Germany, materialized by Wilhelm II's explicit support to the South African Republic's president (Lugan 2010). Eventually, it also prompted the South African Republic to invest in modern military equipment (Pakenham 1979).

In the aftermath of the Jameson Raid, tensions continued to escalate. Britain was demanding full voting rights for the South African Republic *uitlanders*. The Boer administration recognized that accommodating such request would result in the loss of ethnic Boer control towards the British citizens because of the massive influx of new *uitlanders* of British origin. Negotiations aimed at finding a compromise failed one after the other. As a last resort, the Boer Republics asked for the help of the USA to act as mediator. Britain's refusal was considered as an ultimate provocation (Lugan 2010). Meanwhile, Great Britain had increased its military presence on the South African Republic borders. On 9 October 1899, the South African Republic issued an ultimatum giving the British troops two days to leave the Boer Republics' borders. Its rejection resulted in the formal start of the Second Anglo-Boer War on 11 October 1899 with a Boer offensive in the British colonies (Pretorius 2009). At the war onset, both British officials and public opinion expected the Southern African campaign to last only a few months. The cost of the war was estimated at £10 million (Meredith 2007). No actor envisioned at the time that overcoming Boer resistance would take almost three years, at a cost of 22,000 British lives and some £217 million (Omissi and Thompson 2002). Or as expressed by Pakenham (1979, xv) that the war proved to be "the longest (two and three-quarter years), the costliest (over £200 million), the bloodiest (at least twenty-two thousand British, twenty-five thousand Boer and twelve thousand African lives) and the most humiliating war for Britain between 1815 and 1914".

Historians generally agree on defining three major phases in the development of the conflict (Lugan 2010). In the first phase, which lasted until the beginning of January 1900, the Boers succeeded in three key offensives, known as the Black

[3]British South African Export Gazette, May I896, April 1898; Customs returns of the Cape, Natal and the Board of Trade. Retrieved from van Helten (1978).

Week, through their mobile and innovative fight method as commandos. With large reinforcements and change in overall command as from mid-January 1900, the British troops managed to turn their misfortune around, thereby initiating the second phase of the conflict. After relieving British colonies' cities one after the other, the British Empire proclaimed the annexation of the Orange Free State in May 1900 and of the Transvaal in October 1900. Although many believed the war to be over, the Boers initiated the third and last phase of the war in November 1900 with an increased reliance on small and mobile military units. In response to this guerrilla, the British command engaged into heavy-handed policies against the Boer population, burning thousands of farms and intensifying civilians' deportation to concentration camps. On 31 May 1902, the war was officially terminated and the conditions of surrender, which included the formal annexation of the Transvaal, were signed by both sides' representatives (Pretorius 2009; Lugan 2010).

3 Data

The South African Republic's financial history is closely intertwined with both the construction of its railways as catalysts for its independence and economic development and the interpersonal power games between influential parties in the region. In January 1876,[4] the South African Republic attempted to place its first large state bond abroad. European markets were reluctant to float any new emerging sovereign bond due to the series of defaults on South American securities. Thomas Burgers, the South African Republic's President, eventually convinced the Insinger Bank in Amsterdam to float a 3.6 million guilders (£300,000) loan. At the beginning of the 1890s, the South African Republic was faced with a rampant liquidity crisis. To deal with this threatening cash shortage, the South African Republic borrowed £650,000 for one year at 6% interest from a syndicate of continental capitalists (Wilburn 1988). This loan was however too small to substantially improve the financial position of the South African Republic. As a result in 1892, it was forced to incur two extra stopgap loans from the National Bank and *Randlord* J. B. Robinson, of, respectively, £50,000 at 6% and £100,000 at 5 3/16%. It was only in July 1892 that the South African Republic managed to issue a substantial loan abroad. This loan, a £2,500,000 long-term bond, carrying a 5% coupon rate and issued at 90% of par, was underwritten by N.M. Rothschild and Sons. This was the first major South African loan issued in London. The implication of the house of Rothschild, which was perceived as the most reputed underwriter, provided a clear signal of quality (Flandreau and Flores 2009). In view of the political situation prevailing at the time, the choice of location to float the bond and the underwriter might seem counterintuitive at first sight. Both Wilburn (1988) and Kerby (2012) provide, however, evidence that this choice was mainly driven by rational economic arguments. In the months preceding the 1892 SAR bond flotation, Lord Rothschild had acquired a substantial stake in deep level

[4]See records of the Insinger Bank Archives in Amsterdam and Katzenellenbogen (1982).

mines located in the South African Republic. Their prospects were deemed to be substantially enhanced by a SAR loan issuance and subsequent railways development. After he had explored other options, Cecil Rhodes also gave support to the Rothschild loan for similar rationales.

The 1892 SAR bond was largely over-subscribed. It was cross-listed on London, Amsterdam and Pretoria. Negotiations between Rothschild and the Boer administration resulted in the loan's security to be defined as "*the entire revenues of the said Republic derived from all sources except so far as the same revenues may be already pledged to existing engagements*".[5] The terms of the contract also prevented any other debt issuance with greater seniority. The funds were to be contractually allocated to the repayment of specified advances and the construction of "*Railways, Roads, Bridges, Buildings, Telegraphs, Telephones, and other Public Works within and not beyond the limits of the said Republic*"[6] unless otherwise pre-agreed for temporary purposes. At the eve of the Second Anglo-Boer War, the public debt of the South African Republic amounted to 2,680,304 British pounds, of which direct liability to the British Crown of £136,644 and the 1892 SAR bond of £2,500,000.[7] Evidence suggests the South African Republic managed to secure another £2 million in 1899, after the beginning of the conflict (van Helten 1978). This amount transited through the Netherlands South African Railway Company (N.Z.A.S.M) which issued a bond in Amsterdam and Berlin thanks to its strong connections in the European banking sphere.[8]

This paper focuses on the largest bond issued by the South African Republic, the 1892 SAR bond. The data series were hand-collected from *De Telegraaf* for the period 2 September 1895 to 31 December 1902 (n = 2295). Daily closing prices for the 1892 SAR bond traded on the Amsterdam Stock Exchange were encoded.[9] This bond was selected because it was the most important SAR bond publicly traded. Relying on data from the Amsterdam market allows avoiding any potential patriotic bias resulting from most bondholders being citizens of one of the warring parties.[10] To the best of our knowledge, there is no data on the volume of trades for this bond. Changes in bond prices indicate that for most of our sample there was a regular trading activity. In some time instances, values are missing. This may either indicate periods of low levels of trading or reporting issues. In general, however, the trading activity appears to be significant for the whole sample and there is no indication that

[5]Rothschild Archive London XIII/230/56. Contract 07/1892 (Clause 2).

[6]Rothschild Archive London XIII/230/56. Contract 07/1892 (Preamble).

[7]The Advertiser (Adelaide), 21 August 1901.

[8]Van Kretschmar to Regeeringscommissaris, 25 April 1899. N.Z.A.S.M. Archief, Amsterdam. Retrieved from van Helten (1978).

[9]Data could then be cross-checked based on digitalized materials provided by Stichting Capital Amsterdam retrieved from the *Nieuw Algemeen Effectenblad* (for weekly prices until 1897) and the *Officiële Prijscourant* (for daily prices from 1898).

[10]As referred to amongst others by Kang and Rockoff (2015).

liquidity was worse than for other emerging market bonds.[11] Note that if liquidity had been an issue, investors would have required a liquidity premium which means that prices should have been even higher than the observed ones in the absence of liquidity issues. This would in fact strengthen our argument.

The 1892 SAR bond bore "interest at 5% per annum, with coupons payable half-yearly, on the 1st January and the 1st July, in London in pounds sterling; and in Amsterdam and Pretoria, at the exchange of the day".[12] The prospectus also defined the initial repayment schedule as to frame the 1892 SAR bond as a British pounds denominated callable bond: "The whole of the Loan will be repaid not later than July, 1942, by a Sinking Fund of equal half-yearly payments to commence in January, 1903, after which date, the Government reserves to itself the right to repay the whole or any portion of the Loan, on giving six months' notice. The Sinking Fund will be applied to the purchase of Bonds when the price is under par, and when at or above par to drawings by lot".[13,14]

Figure 1 shows the quoted daily prices in Amsterdam from 2 September 1895 to 31 December 1902. The bond price series allow identifying three distinct phases, which

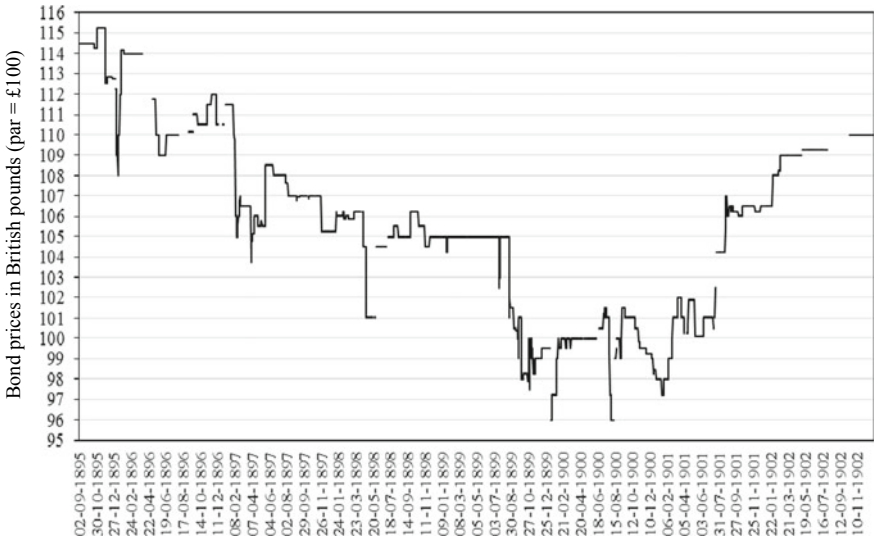

Fig. 1 1892 SAR daily bond prices in Amsterdam over the period September 1895–December 1902

[11] In the absence of bid-ask spreads on the Amsterdam market, it is unfortunately impossible to replicate the methodologies developed by Alquist (2010) or Chavaz and Flandreau (2017) to gauge liquidity.

[12] Rothschild Archive London XIII/230/56. 1892 South African Republic 5% Government Bonds, £2.5 m. Prospectus.

[13] Rothschild Archive London XIII/230/56. 1892 South African Republic 5% Government Bonds, £2.5 m. Prospectus.

[14] See Appendix for a copy of the 1892 SAR bond final certificate.

can be analysed in light of both Dutch and British newspapers as those conveyed public news determining prices. Dutch newspapers often transmitted information inspired from the British press.

The first phase runs from September 1895 to August 1899. Over this first period, bond prices progressively declined from ranges of £112–115 in 1895 to tend towards par at the end of August 1899. This decrease is to be partly attributed to "the growing dearness of money", as underlined by *The Financial Times* in a market analysis released on 21 October 1899. However, the 1892 SAR bond reached its highest value mid-1895 while British consols' prices only started to decline as from the close of 1898 (*The Financial Times*, 21 October 1899). Changes in the conditions influencing high-class investments would therefore be insufficient to fully explain the drop in the 1892 SAR bond prices, the decline being the more significant as it coincided with a rapidly increasing public revenue.[15] No mention of unpaid coupons, which could have distorted the quoted prices, was found over the period. This decrease in prices was therefore supposedly driven by the increasing geopolitical tension in the region together with the growing discontent of *uitlanders* and thus an increase in the risk of war.

The second phase goes from September 1899 to mid-July 1901. Over this period, prices fluctuated in ranges of £96 to £102. With the Boer officials calling up on burghers, prices reached £98 in the week ahead of the Boer ultimatum on 9 October 1899. Over the first phase of the Second Anglo-Boer War, prices evolved slightly below par to reach an absolute low of £96 on 8 January 1900. This is likely to be related to the combination of successive British defeats in December 1899, known as the Black Week, which led to the sending of massive British reinforcements beginning of January 1900 and the subsequent increase in the expected length and cost of war as discussed by *The Economist* on 6 January 1900. As from February 1900, bond prices increased again until end of July 1900. With the most read Dutch newspaper, *De Telegraaf*, suggesting the war would not be over before long on 24 and 25 July 1900, prices decreased up to £96 in the beginning of August 1900. This decrease did not last long as prices had raised again to par at the end of August 1900. The official proclamation of the annexation of the Transvaal on 25 October 1900 had little or no effect and prices decreased again slightly below par from November 1900 to end of February 1901, with a low in January 1901 at £97.25. Prices then slightly fluctuated over par until mid-July 1901, probably driven by mixed effects of uncertainty on war outcome due to ongoing guerrilla and valid expectations the British government would assume the loan and at least part of its past unpaid coupons.

The third and last phase spans the remainder of the time period under consideration. Over this timeframe, prices gradually moved up to top £110 in December 1902. From mid-July to mid-August 1901, prices increased quite substantially under the impulse of both potential leaks and insider trading on settlement negotiations coupled with the news of Lord Kitchener leaving South Africa, which was interpreted as he had "practically accomplished his task", as stated by *The Financial Times* on

[15]From less than £300,000 before the loan issuance to just below £5,000,000 at the eve of the Second Anglo-Boer War. It is mainly related to the development of the gold mining industry.

Table 1 Descriptive statistics—1892 SAR bond prices in Amsterdam in pounds sterling (par = £100) (continuous data over September 1895–June 1902)

1892 SAR bond price series	Daily	Weekly
Mean	105.58	105.57
Median	105.50	105.50
Maximum	115.25	115.25
Minimum	96.00	96.00
Standard deviation	4.57	4.56
Skewness	0.13	0.13
Kurtosis	−0.68	−0.67

23 July 1901. On 20 August 1901, as the British government announced the terms of its guarantee on the 1892 SAR bond, prices increased to £105.5. On 22 August 1901, they rose again to £107. This second jump might have related partly to delay in incorporating assumption news into prices. It also coincided with the SAR President's declaration to a French interviewer, as reported by *The Times,* that "there is still enough money for 'all' requirements". *The Times* commented this statement, declaring the payment of the defaulting coupon on state loan due before annexation had to "obviously rank high amongst these requirements". Over the remaining period of study, prices kept gradually expanding to the close of 1902.

Table 1 presents the main descriptive statistics of the continuous daily and weekly bond price series for the period to which we apply our statistical methods. Bond price series averaged £105.6, with a standard deviation of £4.6 over September 1895–June 1902. Skewness coefficient allows to conclude the distribution is slightly asymmetrical towards higher values (long tail to the right), while the kurtosis coefficient determines our series distribution is slightly too flat to be a normal one.

4 State Succession and Rothschild Guarantee

Our analysis suggests market participants perceived the level of uncertainty to be relatively low. Although they considered the war outbreak somewhat increased the risk, investors likely believed that neither its outcome, nor to a certain extent its timing, would substantially impact the risk profile of the 1892 SAR bond. This argument, in line with the high price levels, can only hold if investors believed there was no substantial risk of non-repayment of the face value of the loan, either on the South African Republic or on the British side. Contemporary newspapers support this view. On 14 October 1899, in the days following the outbreak of the war, *The Economist* declared that the 1892 SAR bond's "comparative steadiness is doubtless due to a belief that whatever may be the outcome of the war the obligation of the Government will be respected". On 21 October 1899, *The Financial Times* reinforced this position by declaring the 1892 SAR bond holders were generally "safeguarded whatever the outcome of the conflict". *The Financial Times* even provided a "buy"

recommendation to investors while the war was in its first phase, arguing that, even in case of British defeat, the Boer administration would not "ruin its credit abroad by repudiating the loan or any part thereon". It seems market participants believed the war only made the payment of the coupons at risk, thereby explaining the increase in yield. We find supportive elements of this. On 21 October 1899, *The Financial Times* nuanced its positive statement as to the 1892 SAR bond low overall risk profile by questioning the next coupon payment, stating "under the extraordinary circumstances it may be difficult, and perhaps not possible, to provide for it punctual payment". Evidence, however, suggests this risk was still perceived as low by investors.

Hence, the decrease in price could be mainly related to the fact that the war outbreak likely reinforced investors' anticipations of an early 1892 SAR bond refinancing. Archival evidence and contemporaneous newspapers suggest market participants anticipated the South African Republic would default on the coupon payments only when it would consider its defeat to be likely. On 3 July 1900, in the days following the first default on coupons' payment, *The Financial Times* reported this first default likely came at a surprise. It declared Messrs Rothschild had been advised by the SAR that "bar gold for the purposes of the coupon had been duly shipped" but that the coupon eventually went by default as "no details were given as to the consignee". *The Financial Times* also suggested the risk of the bond would as from this point in time rest only on the British government and its decision to assume the loan as the Boer officials were "only too aware that they will never be looked to for the payment of any other coupon". The absence of any substantial decrease in price at the time thus appears to further support that investors firmly believed the British government would assume the 1892 SAR bond. This would have transformed a sovereign bond into a colonial one. As a result the bonds' riskiness would have been very close to the riskiness of a British consol, as it would have benefited by an implicit British guarantee (Accominotti et al. 2011). Chavaz and Flandreau (2017) have indeed shown that, for colonial bonds, liquidity was one of the main drivers of any difference in yield with the consols.

The legal literature on state succession allows drawing the legal context in which market participants likely framed their anticipations as to the fate of the 1892 SAR bond. In theory the literature on state successions stresses the continuity of obligations even though political and economic considerations may play an important role in its application (Weidemaier and Gulati 2015). Over the nineteenth century, Britain developed a new doctrine vis-à-vis successor state liability in line with its imperialist ambitions. This theory provided that a successor state which assumed the debts of a predecessor did so *ex gratia* rather than *ex lege*. This practice was highly similar to that established by the USA in pursuing its "manifest destiny". Britain consistently applied its new doctrine in its annexation of the Fiji Islands in 1874 (Feilchenfeld 1931) and Upper Burma in 1886 (Keith 1907). Hoeflich (1982) argues Britain also illustrated the consistency of its position in Chile's annexation of the guano rich Peruvian region of Tarapacà in the aftermath of the War of the Pacific (1879–1883) by refuting any legal obligation laid upon Chile. According to Hoeflich (1982) the

only major test of Britain's "grace" theory occurred in our case study, with the question of the transfer of the Boer Republics' public debts.[16]

To the best of our knowledge, the economic literature on state succession only provides two empirical studies of full state annexation that were analysed per se from a market-based perspective. These consist in the US annexations of Texas in 1845 and of the Hawaiian Kingdom in 1898. The US annexation of Texas was the first illustration of USA's rejection of the "continental theory and practice of accepting there was a strict legal obligation to assume debts" of predecessor states (Hoeflich 1982). The US annexation of Texas led to a decade of uncertainty for its bondholders. The final settlement, provided by the 1855 Payment Act, did not satisfy all bondholders in full and materialized a substantial modification in terms. The price of the bonds issued by Texas remained well below par, with a minimum at 10%, over the whole period of uncertainty (Burdekin 2006). Furthermore, both actual legislation and rumours of pending action did structurally impact Texas bond prices. To the contrary, the Hawaiian Kingdom bond price levels remained high, with a minimum at around 90% of par, at the eve of its annexation. The context surrounding the public debt transfer of the Hawaiian Kingdom is substantively different from our case of study. Firstly, the Hawaiian Kingdom annexation did not require a war on the part of the USA.[17] Secondly, Burdekin and Laney (2008) associate the absorption of Hawaii's foreign sovereign debt with the US new expansionist policy which culminated in Roosevelt Corollary's 1904 announcement.[18] Sicotte and al. (2010) analyse the impact of a partial annexation on the yield spreads of both the predecessor and successor states. The authors suggest a relationship between resource abundance and interest of foreign investors.[19] Market-based studies have also been performed on other forms of state succession, which the legal literature treats as a single and integrated topic. For instance, Collet (2013) demonstrates the existence of a risk premium attached to the "odious character" of debt in the context of the Spanish–American war (1898) and the subsequent Cuban independence. Collet and Oosterlinck (2019) also find the existence of a risk premium for the 1906 Russian loan.

The British Empire formally proclaimed Transvaal's annexation as early as on 25 October 1900, in the aftermath of Pretoria's seizure (Pakenham 1979). The question whether Britain was legally bound to assume the public debt of the Boer Republics

[16] We can note the 1892 SAR bond holders might also have had in mind the Cuban debt controversy as to the "odious" character of Cuban bonds which settled in 1900 although the contexts involved differed (Collet 2013).

[17] The US annexation would have even been actively supported by the Hawaiian provisional government after the overthrown of the Hawaiian Kingdom in 1893.

[18] Under this new policy, the USA would stand ready to intervene and ensure debt repayment by otherwise sovereign nations in the Caribbean, Central America and northern South America. Not assuming legitimate public debt of an annexed territory would have hence sent a contradicting signal to financial markets.

[19] Other market-based studies indirectly provide information on the fate of public debt in case of state annexation. As an example, in their study of the Confederates' odds of winning the Civil War, Oosterlinck and Weidenmier (2007) base their calculations on the assumption market participants anticipated their bonds would not be repaid in case of South defeat.

was therefore on the agenda well in advance of the final peace treaty. Legal advice was sought by various stakeholders. The views expressed differed substantially. According to Johannes Wessels,[20] legal adviser to Lord Kitchener and Lord Roberts, "the only logical solution is to adopt the principle that the conqueror succeeds to all the assets (corpora) and all the rights (jura) of the conquered, but that he is not compelled as of right to acknowledge the obligations of the defunct state",[21] ending his analysis by "a complete refusal to be bound by the obligations of the South African Republic is the stronger and in my opinion the correct position".[22] This position is supported by Keith (1907): "In 1900, His Majesty's Government declined to recognize any strictly legal liability for the debts of the conquered territories". O'Connell (1956) specifies, based on a confidential paper of the British Foreign Office,[23] dated 30 November 1900, that certain British officials desired to take "extreme positions" towards the general public debt of the Boer Republics. British officials wanted to make clear there was no legal liability for such debt and suggested Britain could modify the terms if it chose *ex gratia* to assume the debt given the increased security provided by the debtor's substitution. O'Connell (1956) provides a precision on the eventual restrictive nature of the potential transfer of debt, stating "In the negotiations for peace in South Africa, the British Government in general repudiated liability for debts contracted during the war". In this prospect, the British position towards the Boer Republics debt appeared to be similar to the US stand in both Texan and Cuban debt controversies (Hoeflich 1982). This view was however not shared by all. In a confidential memorandum on the Public Debts of the late South African Republics dated from June 1901, Mr. Bertie asserted that "By international law, *alias* custom, it is incumbent upon Her Majesty's Government to take over the public debts of the South African Republics incurred before the war". The press did not give a much clearer picture. On 30 June 1900, *The Financial Times* declared the British government "did not intend to take over the liabilities of the Orange River Colony". In the absence of a firm legal commitment, one may wonder whether investors were relying on other binding mechanisms.

A feature of the 1892 SAR bond may help explaining the firmly held belief that it would be repaid. The 1892 SAR bond was floated by The House of Rothschild which was considered as the most prominent banking house of the time.[24] The literature on the quality of underwriter hence potentially defines a route for rationalizing the South African Republic debt price resilience. Over the nineteenth century, foreign lending to governments occurred through banking houses located in the contemporary leading financial markets (Flandreau et al. 2009). The originating houses played a different role from today. At the time, underwriting acted as a signal of creditworthiness. Renowned underwriters substituted for the borrowers' reputation, enabling to

[20]The document is not dated but it was forwarded in September 1900. National Archives, DO119-662.

[21]National Archives, DO119-662.

[22]National Archives, DO119-662.

[23]F.O. Confidential Paper (7516), No. 22A, reprinted in O'Connell (1956).

[24]It led both London and Paris market for foreign government debt issues (Flandreau et al 2009).

solve their pre-commitment problem. In turn, these banking institutions could monitor the issuers. Prestige was driven by safety. Premier underwriters hence worried about excess volatility and frequently engaged in market operations to support the government bonds they had issued or served as lenders of last resort (Flandreau et al. 2009a, b, 2011). Empirical evidence of the importance of Rothschild as an underwriter is provided in extreme contexts (Collet 2012). Weller (2010) demonstrates that Rothschild provided for the bail-out of the Brazilian public debt in 1914 in part to preserve its status as a premier underwriter.

Archival evidence supports this view. The positive British settlement appeared to have been pre-committed as from the issuance of the 1892 SAR bond. Kerby (2012) highlights the preamble to the 1892 draft bond certificate included a clause which stipulated "… the current government in the Transvaal or any government whether invading or annexing the Transvaal would be responsible for the repayment of the loan". According to Kerby (2012), this condition suggests the British government agreed in advance to assume the loan repayment in case of annexation of the Transvaal or was at least aware of the ex ante commitments taken by the South African Republic to ensure the successful flotation of the bond. We can indeed assume the British government could not have been out of the loop of such agreement given its close link with the House of Rothschild.

This "pre-commitment" argument provides the most credible rationale to explain the price evolution of the South African Republic. At the time of the loan issuance, Rothschild was likely aware the Second Anglo-Boer War and the subsequent annexation of the South African Republic were deemed or at least likely to happen over the lifetime of the bond. The private correspondence between Rothschild and the South African Republic indicates none of the parties to the flotation of the bond objected to insert a clause in the final bond certificate dated 12 December 1892,[25] which implicitly revealed the perceived likelihood of an Anglo-Boer conflict, by stating: "*The payment of the interest on this Loan and the redemption thereof shall be effected in time of war as well as in time of peace without distinction, whether the holders of the Bonds belong to a friendly or a hostile nation*". We can assume Rothschild, knowing the SAR likely annexation, would not have put its reputation at risk by floating this loan without the implicit consent of the British Crown.

The high prices observed for the 1892 SAR bond would, therefore, lie in the fact that Rothschild, as its underwriter, credibly signalled to investors a pre-commitment existed on the part of the British Crown as to the assumption of the 1892 SAR bond. In political terms, investors might indeed at first have wondered why a leading underwriter based in London would agree to float the bond of one of Great Britain's potential future enemy. How could investor reconcile the fact that the "banker to the British Crown" was floating a bond for the "Boer enemy"? Indeed, investors might have thought such association would not have been considered without the agreement or even the impulse of the British government. The investors could have therefore anticipated Rothschild floated the 1892 SAR bond as to make the Transvaal colony more valuable to Britain. Indeed, the loan proceeds allowed for the timely building

[25] Rothschild Archive London 000/73/119.

of roads, railways and other public works which would have been in working order by the time the Transvaal would be annexed.

If we believe the British Empire was committed to assume the 1892 SAR bond, it would probably have aimed at reaching a settlement and refinancing the bond as soon as it could. As long as a final settlement was not reached, the British Empire had to pay coupons at 5% rate per annum on the 1892 SAR bond which was well above what its own credit risk would have required.[26] Archival records appear in line with that argument. Negotiations as to the 1892 SAR bond were launched even before the proclamation of Transvaal's annexation. On 23 October 1900, a confidential letter addressed to The House of Rothschild by the Colonial Office attested to the early discussions, by stating that "the question of the position assumed by Her Majesty's Government in connection with the Transvaal 5% Loan is under consideration and that Mr. Chamberlain hopes before long to be able to give a definite answer". Going forward, the 1892 SAR bond appears to be the first SAR security for which a settlement was agreed upon.[27]

The House of Rothschild contributed to create an environment in which all interests were aligned towards a positive and rapid final settlement. Likely, the speed of actions was also driven by Britain long-term ambitions. The British Empire's interest was to float a single new Transvaal loan at the end of the war, which would account for all Transvaal financing needs,[28] as to minimize costs. The sooner an agreement was reached on all matters it related to, the sooner it could be issued.[29] This observation may lead to a reinterpretation of the British "grace" doctrine as to successor state liability. The *ex gratia* argument might have been mainly required as a vehicle to fasten tough and potentially harmful negotiations as to the state of international law. Such assertion is supported by archival evidence from the British Foreign Office. A confidential memorandum, dated 3 February 1901, recommended "it would be politic for His Majesty's Government to determine what would be an equitable settlement, and to announce it publicly without previous negotiation with

[26] As testified by the yields on the British consols which at the time of the SAR annexation amounted to around 2.5% (Klovland 1994).

[27] New archival evidence from the Insinger Bank in Amsterdam reveals Messrs Insinger & Co. communicated on 29 May 1902 that 150 bonds of 1000 guilders of the 1876 SAR bond were still outstanding, indicating the repayment of these claims had not been settled several months after the 1892 SAR bond assumption terms were announced.

[28] These included the war compensation payments, the reconstruction and the acquisition of railways and the development of the gold mining industry.

[29] Archival records support this interpretation. At the issuance of the prospectus of the new 3% Transvaal guaranteed Loan in May 1903, Mr. Chamberlain made it clear in his statement to the House of Commons how crucial and intertwined were the final settlements on all issues to the flotation of this necessary loan, declaring on 7 May 1903, that beyond the other purposes included in the loan, "£2,500,000 would be required for the conversion of the old Five per Cent debt". A few days after the successful flotation of the new Transvaal Loan, on 15 May 1903, the final 1892 SAR bond repayment date was notified in *The London Gazette* to be 15 August 1903, "from which date all interest on the Bonds will cease and determine".

foreign Governments or bondholders".[30] The terms offered to the holders of the 1892 SAR bond were as follows. Not only did the British government repay the bond at par but it also assumed the repayment of all coupons due after it proclaimed the Transvaal annexation. Britain did so even before the final peace treaty was signed in May 1902. The Colonial Office only required bondholders to forgo the semi-annual coupon due in July 1900.[31] As from 20 August 1901, the bonds were also made callable at par on three months' notice, failure to surrender their bonds and all future coupons attached on a defined later date left the bondholders "dependent for the payment of both principal and interest on the revenues of the Transvaal only, on which the advances made by His Majesty's Government for the settlement of this debt, and for other purposes, would be a prior charge".[32] While this call clause might appear a substantial modification of the terms to jurists, it only slightly modified the content of the initial contract. *The Economist* gave support to this interpretation in an article published on 24 August 1901, by saying "The terms upon which our Government has offered to assume responsibility for the Transvaal loan of £2,5000,000 issued in 1892 are sufficiently generous".

5 Conclusion

Wars are generally known to have an economically significant impact on the states they involve. Their harmful effect illustrates itself in the behaviour of their government bonds. This article provides insight as to which mechanisms could prevent wars from impacting sovereign bond markets. Over the whole Second Anglo-Boer War (1899–1902), the South African Republic main foreign sovereign bond—the 1892 SAR bond—traded at remarkably high and stable levels. The war outcome was not expected to substantially impact future cash flows. A Boer success would in all likelihood guarantee repayment but from the onset of the war this was generally viewed as extremely unlikely. Investors thus had to believe that even in the case of a British victory they would be repaid. In other words, that Britain would not only accept to recognize a liability for this loan but also to offer to reimburse it on generous terms.

At the time of the SAR annexation, neither the British doctrine nor the advocated legitimate character of the bond could have justified such strong assumption. Bondholders may have believed Britain would take over the 1892 SAR bond but the high price at which it traded meant that they were almost sure this would be the case and without substantial loss. We attribute investors' confidence to an informal

[30]The National Archives, Kew. FO 881/7528—AFRICA: Memo. Public Debts of late South African Republics. (Mr. F. Bertie).

[31]The Times, 21 August 1901; The Rothschild Archive, London. Box XI/III/119.

[32]The Times, 21 August 1901; The Rothschild Archive, London. Box XI/III/119.

pre-commitment to take over the debt. Archival evidence indicates indeed that a pre-commitment had likely been framed on the part of the British government as to the 1892 SAR bond assumption at the time of its issuance. Rothschild, as underwriter to the bond, is deemed to have signalled market participants such pre-commitment existed. Knowing the SAR annexation to be likely, Rothschild would not have put its premier underwriter reputation at risk by underwriting a loan without some kind of guarantee. This argument is further supported by the apparent puzzling association this bond established and the likely interest the British Empire had in the flotation of the 1892 SAR bond. In a certain way, the 1892 SAR bond could hence credibly be viewed as a future colonial bond as from its issuance. This evidence would suggest the signalling role of a premier underwriter over the nineteenth century could hold in the most extreme circumstances. Indeed, the underwriter did not only monitor the issuer state but also its most likely successors. Timing of the settlement was short. This observation could possibly allow for a reinterpretation of the British "grace" theory as a vehicle intended to bypass controversies as to the state of international law. The case of the SAR 1892 bond brings thus a new perspective to the existing economic literature on state succession in general, and state annexation in particular.

Appendix—The 1892 SAR Bond Certificate

See Fig. 2.

Fig. 2 1892 South African Republic 5% government bonds, £2.5 m. Original bond certificate. Scrip for £1000. *Source* Rothschild Archive London 000/337/1/17

References

Archival Sources

The Rothschild Archive, London.
The National Archives, Kew.
The Insinger Bank Archive, Gemeente Amsterdam Stadsarchief, Amsterdam.

Secondary Sources

Accominotti, O., Flandreau, M., & Rezzik, R. (2011). The spread of empire: Clio and the measurement of colonial borrowing costs. *Economic History Review, 64*(2), 385–407.
Adams, D. S. (2015). Contemporary perceptions of the First World War reflected in the capital markets. *Scandinavian Economic History Review, 63*(1), 1–23.
Alquist, R. (2010). How important is liquidity risk for sovereign bond risk premia? Evidence from the London stock exchange. *Journal of International Economics, 82*(2), 219–229.
Blainey, G. (1965). Lost causes of the Jameson Raid. *Economic History Review, 18*(2), 350–366.
Brown, W. O., & Burdekin, R. C. K. (2000). Turning points in the U.S. civil war: A British perspective. *Journal of Economic History, 60*(1), 216–231.
Brown, W. O., & Burdekin, R. C. K. (2002). German debt traded in London during the Second World War: A British perspective on Hitler. *Economica, 69*(276), 655–669.
Burdekin, R. C. K. (2006). Bondholders gains from the annexation of Texas and implications of the US Bailout. *Explorations in Economic History, 43*(4), 646–666.
Burdekin, R. C. K., & Laney, L. L. (2008). Financial market reactions to the overthrow and annexation of the Hawaiian Kingdom: Evidence from London, Honolulu and New York. *Cliometrica, 2*(2), 119–141.
Chavaz, M., & Flandreau, M. (2017). High and dry': The liquidity and credit of colonial and foreign government debt in the London Stock exchange (1880–1910). *Journal of Economic History, 77*(3), 653–691.
Collet, S. (2012). Sovereign bonds: Odious debts and state succession. Ph.D. thesis. Solvay Brussels School of Economics and Management, Université Libre de Bruxelles.
Collet, S. (2013). How big is the financial penalty for 'unfair' debt? The case of Cuban bonds at the time of independence. *European Review of Economic History, 17*(3), 364–387.
Collet, S., & Oosterlinck, K. (2019). Denouncing odious debts. *Journal of Business Ethics, 160*(1), 205–223.
Dincecco, M. (2009). Political regimes and sovereign credit risk in Europe, 1750–1913. *European Review of Economic History, 13*(1), 31–63.
Feilchenfeld, E. (1931). *Public debts and state succession*. New York: Macmillan.
Flandreau, M., & Flores, J. H. (2009). Bonds and brands: Foundations of sovereign debt markets, 1820–1830. *Journal of Economic History, 69*(3), 646–684.
Flandreau, M., Flores, J. H., Gaillard, N., & Nieto-Parra, S. (2009). The end of gate keeping: Underwriters and the quality of sovereign bond markets, 1815–2007. In *NBER working papers* 15128.
Flandreau, M., Flores, J. H., Gaillard, N., & Nieto-Parra, S. (2011). The changing role of global financial brands in the underwriting of foreign government debt (1815–2010). In *IHEID working papers.* Economics Section, The Graduate Institute of International Studies.
Frey, B. S., & Kucher, M. (2000). History as reflected in capital markets. The case of World War II. *Journal of Economic History, 60*(3), 468–496.

Frey, B. S., & Kucher, M. (2001). Wars and markets. How bond values reflect the Second World War. *Economica, 68*(271), 317–333.
Frey, B. S., & Waldenström, D. (2004). Markets work in wars: The Zurich and Stockholm markets during World War II. *Financial History Review, 11*(2), 51–67.
Graham, W. (1996). The Randlord's bubble 1894–6/South African gold mines and stock market manipulation. In *Discussion papers in economic and social history*, n. 10. University of Oxford.
Hoeflich, M. H. (1982). *Through a glass darkly: Reflections upon the history of the international law of public debt in connection with state succession* (pp. 39–70). University of Illinois Law Review.
Hudson, R., & Urquhart, A. (2015). War and stock markets: The effect of World War Two on the British stock market. *International Review of Financial Analysis, 40,* 166–177.
Jopp, T. A. (2016). Contemporaries' opinions of the Allied and Central Powers' performance during the First World War: Measuring turning points in perception with sovereign debt prices. *European Review of Economic History, 20*(2), 242–273.
Kang, S. W., & Rockoff, H. (2015). Capitalizing patriotism: The liberty loans of World War I. *Financial History Review, 22*(1), 45–78.
Katzenellenbogen, S. E. (1982). *South Africa and Southern Mozambique: Labour, railways and trade in the making of a relationship*. Atlantic Highlands, NJ: Humanities Press Inc.
Keith, A. B. (1907). *The theory of state succession with special reference to English and colonial law*. London: Waterlow and Sons Limited.
Kerby, E. (2012). *Financing the foe: A case study on the financing of an independent railway to the Zuid-Afrikaansche Republiek between 1882 and 1902*. Paper prepared for the African Economic History Network meeting. Stellenbosch, July 2012.
Klovland, J. T. (1994). Pitfalls in the estimation of the yield on British Consols, 1850–1914. *The Journal of Economic History, 54*(1), 164–187.
Landon-Lane, J., & Oosterlinck, K. (2006). Hope Springs eternal: French Bondholders and the Soviet repudiation (1915–1919). *Review of Finance, 10*(4), 507–535.
Lugan, B. (2010). *Histoire de l'Afrique du Sud: des origines à nos jours*. Paris: Ellipses.
Meredith, M. (2007). *Diamonds, gold and war*. The Making of South Africa, London: Simon & Schuster UK Ltd.
Mitchener, K., & Weidenmier, M. (2005). Empire, public goods, and the Roosevelt corollary. *Journal of Economic History, 65*(3), 658–692.
Mitchener, K., Oosterlinck, K., Weidenmier, M., & Haber, S. (2015). Victory or repudiation? Predicting winners in civil wars using international financial markets. *Journal of Banking & Finance, 60,* 310–319.
Occhino, F., Oosterlinck, K., & White, E. N. (2008). How much can a victor force the vanquished to pay? France under the Nazi boot. *The Journal of Economic History, 68*(1), 1–45.
O'Connell, D. P. (1956). *The law of state succession*. New York: Cambridge University Press.
Omissi, D., & Thompson, A. (2002). *The impact of the South African war*. New York: Palgrave Macmillan.
Oosterlinck, K. (2013). Sovereign debt defaults: Insights from History. *Oxford Review of Economic Policy, 29*(4), 697–714.
Oosterlinck, K. (2016). *Hope Springs Eternal. French Bondholders and the Repudiation of Russian Sovereign Debt* (p. 264), Yale University Press.
Oosterlinck, K., & Ureche-Rangau, L. (2008). Multiple potential payers and sovereign bond prices. *Finance, 29*(1), 31–52.
Oosterlinck, K., Ureche-Rangau, L., & Vaslin, J. M. (2014). Baring, Wellington and the resurrection of French public finances following Waterloo. *The Journal of Economic History, 74*(4), 1072–1102.
Oosterlinck, K., & Weidenmier, M. D. (2007). *Victory or repudiation?: The probability of the Southern confederacy winning the Civil war*. National Bureau of Economic Research.
Pakenham, T. (1979). *The Boer war*. London: Weidenfeld and Nicolson.
Pretorius, F. (2009). *Historical dictionary of the Anglo-Boer War*. Lanham, MD: Scarecrow Press.

Sicotte, R., Vizcarra, C., & Wandschneider, K. (2010). Military conquest and sovereign debt: Chile, Peru and the London bond market, 1876–1890. *Cliometrica, 4*(3), 293–319.
Sussman, N., & Yafeh, Y. (2000). Institutions, reforms, and country risk: Lessons from Japanese government debt in the Meiji era. *The Journal of Economic History, 60*(2), 442–467.
van Helten, J.-J. (1978). German capital, the Netherlands Railway Company and the political economy of the transvaal 1886–1900. *The Journal of African History, 19*(3), 369–390.
van Helten, J.-J. (1979). British capital, the British state and economic investment in South African 1886-1914. *Collected Seminar Papers, Institute of Commonwealth Studies, 24,* 1–17.
van Helten, J.-J. (1982). Empire and high finance: South Africa and the international gold standard 1890–1914. *The Journal of African History, 23*(4), 529–548.
Vizcarra, C. (2009). Guano, credible commitments, and sovereign debt repayment in nineteenth-century Peru. *The Journal of Economic History, 69*(2), 358–387.
Waldenström, D., & Frey, B. S. (2008). Did nordic countries recognize the gathering storm of World War II? Evidence from the bond markets. *Explorations in Economic History, 45*(2), 107–126.
Weidemaier, M. C., & Gulati, G. M. (2015). The relevance of law to sovereign debt. *Annual Review of Law and Social Science, 11,* 395–408.
Weidenmier, M. D. (2002). Turning points in the U.S. civil war: Views from the Grayback market. *Southern Economic Journal, 68*(4), 875–890.
Weidemaier, M. C., & Gulati, G. M. (2016, February 28). *Differing perceptions? Market practice and the evolution of foreign sovereign immunity* (UNC Legal Studies Research Paper No. 2739423).
Weller, L. (2010). What is worst, trade crisis or revolution? Foreign creditors, sovereign debt and Bailouts in Brazil and Mexico, 1913–14. In *Working paper*.
Wilburn, K. (1988). The nature of the Rothschild loan: International capital and South African railway diplomacy, politics, and construction 1891–1892. *South African Journal of Economic History, 3*(1), 4–19.
Willard, K., Guinnane, T., & Rosen, H. (1996). Turning points in the civil war: Views from the Greenback market. *The American Economic Review, 86*(4), 1001–1018.
Wilson, M. H., & Thompson, L. M. (1969). *The Oxford history of South Africa*. Oxford: Clarendon Press.
Zussman, A., Zussman, N., & Nielsen, M. Ø. (2008). Asset market perspectives on the Israeli-Palestinian conflict. *Economica, 75*(297), 84–115.

Expectations and Planning at the FED: 1939–1941

Rob Roy McGregor and Warren Young

1 Introduction

The years 1939–41 were crucial ones for the Federal Reserve System (FRS).[1] The transition from an economy emerging from depression—albeit at peace—to one facing the economic impact of war in Europe and the prospect of becoming engaged in that war, with resulting massive increases in defense expenditure, brought about significant changes in the expectations and plans of the BG over the period. This paper focuses on the ***expectations-planning nexus*** at the BG over the period 1939–41, which, in our view, has not received the attention it deserves. Our focus on this period emanates from what we think is a ***lacuna*** in Meltzer's magisterial *History of the Federal Reserve* (2003, Vol. 1).

Now, while he devoted a significant section of Chap. 6 of his first volume to a discussion of "**Policy and War Preparations, 1939–41,**" Meltzer's ***main*** concern was upon "the search for a policy guide" under wartime conditions with reference to possible open market operations advocated by the BG. This is evident from his focus on a memorandum dealing with this issue ("Despres to Goldenweiser, April

[1]The Federal Reserve Act (1913) created the Federal Reserve Board (FRB). It also created the Federal Advisory Committee (FAC) (1913 Act, Sec 12). The Banking Act of 1935 replaced the FRB with the Board of Governors (BG) and finalized the structure of the Federal Open Market Committee (FOMC). Between 1939 and 41, the Fed was not "independent," only regaining it after the Treasury-Fed Accord of 1951. See Meltzer (2003, Chap. 7), Bordo (2010, 2–3, 2016, 230–31) and Hetzel and Leach (2001).

R. R. McGregor
UNC Charlotte, Charlotte, USA
e-mail: rrmcgreg@uncc.edu

W. Young (✉)
Department of Economics, Bar Ilan University, Ramat Gan, Israel
e-mail: youngwprof@yahoo.com

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_10

29, 1940"), and on the BGs discussion of "the dangers of continuance" of its "easy money" policy, as the FAC's critique called it, and which the BG had to address.[2]

As will be seen, however, despite Meltzer's not having recalled their efforts, Emile Despres, along with Alvin Hansen, Charles Kindleberger, among other less well-known economists, such as Walter Gardner and Martin Krost, authored *numerous* BG memos on the possible impact of the war on the US economy, the wartime economy of the USA, the implications of a possible German victory or negotiated peace on the US economy, the "gold problem," and other issues, while serving in various capacities at the FRS, over the period 1939–1941.[3]

To comprehend the scope of our study, we must first define what is meant here by the term ***expectations-planning nexus*** *over the period 1939–41*. Regarding ***expectations,*** we take them to have been ***qualitative and exogenous***, either ***deduced*** from the structure of the international political system ***extant*** or ***inferred*** from ***previous*** events and patterns, ***as perceived at the time***.[4]

As for ***planning***, we distinguish between ***internal*** and ***external planes***. The internal plane deals with available resources and how to deploy them within an organization. The external plane takes account of economic, technological, social, and political factors when assessing and then setting organizational objectives, that is to say, taking the environment, or state of the world in which the organization operates and is planned to operate, into account. Here, we focus on the ***external plane*** of the war planning process undertaken by the BG and its economists as manifest in the "**Defense Program**" in which it took part.

---

[2]BG Minutes, May 21, 1940. Although ***not cited*** by Meltzer, also see *Minutes of Meeting* of the FAC, May 19–21, 1940, for the background to the debate between the FAC and BG over "easy money."

[3]Despres was brought to the BG from the NY Fed in 1938 by Lauchlin Currie, then Assistant Director of the BG's Division of Research and Statistics (DRS). Hansen was appointed Special Economic Adviser attached to the DRS and Kindleberger an Associate Economist in the DRS, both in July 1940. Gardner was a senior economist, and Krost, Currie's former student, was an economist in the DRS. The DRS was headed by Emmanuel Goldenweiser. It should be noted here that Jones (1972) in his study of "The Role of Keynesians in Wartime Policy and Postwar Planning. 1940–1946" also does not mention the work of the "Keynesian"economists, such as Hansen, ***at the BG*** over the period 1940–1941.

[4]The exogenous state of expectations directly and indirectly influencing the BG's economists is derived from four sources: (i) British War Cabinet Documents; (ii) Minutes of the US State Department's Advisory Committee on Foreign Relations; (iii) Directives of War Department's Joint Planning Committee, and (iv) Memos of the US Army's War Plans Division. It should be recalled here that the ***expectations-planning nexus*** was also manifest at the State Dept. over the period 1940–41, in the form of the "Advisory Committee on Problems of Foreign Relations," set up by Secretary Hull in January 1940. It was chaired by Undersecretary Welles and composed of State Dept. planners such as Notter. It also held joint meetings with personnel from other Government Departments such as Treasury, Commerce, and Agriculture. Its initial goal was exploring the possibility of peace and studying post-war recovery "with primary reference to the best interests of the USA." See for example, Notter (1949) and Welles (1951). But, as will be seen below, ***expectations*** regarding what constituted "peace" and "post-war" shifted over the period 1939–41. The ***nexus*** also characterized the approach of the War Dept. and US Army, with the focus on "war" rather than "post-war" planning, albeit, again, shifting over the period.

We will be presenting *documentary* evidence to deal with the ***nexus***. The material is compiled, in the main, from FRASER and other *primary* sources. In other words, we have followed the wise advice of Rolnick, who wrote, over three decades ago (1985, 195) that in order to understand what was going on at the FRS and its constituent elements over their history, and it was necessary "to dig deeper and more carefully into" their *documentary* history and "behavior to uncover evidence."

Up to now, studies of the FRS in terms of its expectations and planning activities in the WWII period have focused, in the main, on factors such as price policy and price expectations, and open market operations in wartime, while its planning activities have been studied for the most part in terms of "post-war," that is to say, ***post Allied victory*** economic stabilization plans, such as those of White and Keynes; preparations for and implementation of the Bretton Woods agreements; and designing post-war US economic policy so as to generate "full employment," ***from about late 1942 to early 1943 onwards***.

But, as mentioned above, the BGs economists ***earlier*** dealt with war-related issues in the context of the expectations-planning nexus, on the ***external*** plane. ***Three periods*** and their associated memos written by BG economists can be identified. The ***first*** involved the expectations and planning regarding the possible impact of the war that had broken out in Europe on the US economy, that is to say, from ***September 1939 to April 1940***. The ***second*** comprised expectations of, and planning for the possibility of a "post-war" situation of the German ***victory***, or a "negotiated peace between Britain and Germany, over the period ***April 1940–September 1940***. The ***third,*** from ***September 1940–December 1941***, involved expectations of, and planning for a number of contingencies, ranging from the continuation of hostilities, with possible of eventual US involvement; a possible complete British defeat; and the possibility of an imposed "peace" in Europe. Moreover, there was also consideration of "***post-defense"*** economics, ***given*** the possibility of an "allied" victory over Germany, as the result of US involvement; these memos also appeared over the period September 1940–December 1941, ***and this, even before the entry of the US into war against both Germany and Japan in December 1941***.

On the ***internal*** plane, the FRS was also concerned with research functions and their organization ***within*** the system, that is, expectations of who ***should*** control them—the BG or District Banks (DBs)—and planning for research activities, from 1943 onwards, but this has been, and will be dealt with in further detail, elsewhere.[5]

[5]Meltzer ***does not deal*** with the debates and controversy between the BG and DBs in the form of the Chairman of the BG and DB Presidents, regarding who was to set the direction of research from 1943 onwards, and into the post-war period. In a previous paper, we surveyed the debates regarding this; see Young and McGregor, "Fed vs. Fed: some essential tensions," presented at the Guggenheim Conference, Geneva, 2015. On the role of DB Presidents, see McGregor and Young (2013), "Federal Reserve Presidents as Publc Intellectuals" *History of Political Economy* 45: 166–190.

2 Expectations and Planning for War: September 1939–April 1940

Laughlin Currie, formerly at the FRS, was appointed as Economic Assistant to Roosevelt in July 1939. On September 7, 1939, a week after the start of the war in Europe, he sent a memo to Eccles regarding his own role in "dealing with the economic problems raised by the European war" and the role he envisioned for his former colleagues at the FRS. Harvard-educated Currie had been recruited from the Treasury to the BG in 1934, by the incoming Chair, Eccles, to assist him with the drafting of the Banking Act of 1935, which revised the BG–DB relationship. Currie had recommended Harvard-educated Emile Despres, who was serving at the NY Fed, for a position at the BG in late 1938. Despres was initially seconded to the Research Division, and then took up a more permanent position in February 1939 (for a detailed account of Currie's work at the FRS, see Sandilands 1990). In his memo to Eccles, Currie wrote (September 7, 1939; EP 46/8/3):

> The function has been assigned to me of allocating among the various departments and agencies the numerous special studies now urgently required to provide a basis for dealing with the economic problems raised by the European war. I am most eager that as much of this work as possible be assigned to the Board's Division of Research and Statistics. For this purpose an expansion of personnel, particularly in Mr. Despres' section, seems to be required at once.

Despres, who as noted that been recruited from the New York Fed to the BG by Currie in 1938, served in the DRS led by Goldenweiser. In this capacity, he compiled economic assessments for Eccles and BG from 1939 until his departure 1941 for the Joint Intelligence Staff in the War Department and OSS.[6]

Despres' first memo for Eccles after the outbreak of war in Europe was entitled "Tentative Appraisal of the Economic Effects of European War upon the United States," dated September 30, 1939. On the cover sheet to this memo (EP 49/10/1), Despres wrote:

> In accordance with your request, a study of the economic effects of European war on the United States has been undertaken, and the first installment of a memorandum on the subject is attached. The memorandum will contain three main sections and two appendices as follows:
>
> 1. The present situation
> 2. Short-term and longer-term prospects …
> 3. Implications for policy.
>
> Appendices …

In order to understand this memo, we must ***first*** determine its contextual state of expectations and assumptions ***regarding the war*** itself. Fortunately, Despres was quite clear regarding this when he wrote (1939, EP 49/10/1; 4, 13):

---

[6]Despres later became State Dept. adviser on the German economy in 1944 and a member of the US delegation to Potsdam in 1945. His academic career included Chairs at Williams College, 1946–61, and Stanford, from 1961 until his death in 1973.

> Appraisal of the short-term and long-term effects of the European war upon the American economy will be based upon the assumptions that the war will continue for several years and that the scale of the actual hostilities will grow; that the United States will not enter the war; and that our embargo on arms exports to belligerents will be repealed…. It was announced on September 9 that the British War Cabinet "decided to base their policy on the assumption that the war will last three years"[7]

Despres outlined his view of the state of "business expectations" in the USA at the start of hostilities in Europe as follows (1939, EP 14/10/1; 2-4):

> Business expectations at the outbreak of war in 1939 were based directly upon memories of the powerful inflationary stimulus exerted by the war of 1914–18. In forming business judgments, little weight was given to the possible effects upon our economy of the disruption of peacetime trade in European countries…it seems appropriate to suggest that business and financial expectations at the outbreak of war in 1939, founded upon a simple projection of our last previous wartime experience, may turn out to have been as misguided in many respects as the expectations which prevailed in 1914. It would surely be dangerous to assume that the behavior of our commodity and security markets following the outbreak of war reflected a careful appraisal of the probable impacts of war upon our economy.

But, he continued (EP 14/10/1;4):

> The assumption that the United States will not enter the war provides a fairly safe basis for analyzing the short-term effects of European war upon our country, but this assumption becomes less certain in appraising the longer term impacts of war. If the United States eventually participates as a belligerent, much of the analysis here presented of the longer term economic effects of the European war upon the United States will require modification.

In other words, for Despres, expectations and planning for war were ***contingent.*** He thought that ***nexus*** would change according to alternate scenarios he outlined. And indeed, as will be seen below, this was the case starting in April 1940, and after December 7, 1941, when the "European war" became World War. But as early as September 1939, Despres took this into account when he wrote (EP 14/10/1; 13) that: "Great Britain and France base their prospects of victory upon their economic staying power and the gradual deterioration of popular morale in Germany, rather than quick military successes". How then, according to Despres, would this affect the US economy? In his view (EP 14/10/1; 22) the main "impact of war" would be on US "foreign trade." He divided this into "short" and "longer" term influences. He wrote (EP 14/10/1; 22-24):

> In view of the British and French determination to draw sparingly upon their gold exchange reserves, which constitute an important element in economic staying power, their net imports

---

[7]The source for Despres statement and ***expectation*** was the British War Cabinet meeting held on September 9, 1939 (National Archives, CAB/65/1/9, War Cabinet 9(39), Minute 15, 68):

The War Cabinet agreed:

(a) That a declaration should be issued to the Press the same day stating in solemn form that His Majesty's Government were mobilizing the entire resources of the country with a view to the successful prosecution of the war, and that to this end ***were making all their plans on the assumption that the duration of the war would be at least three years… .***

> from the United States will increase only moderately during the early months of the war as increased importation of goods related directly to war needs will be offset by forced curtailment of imports for civilian consumption. The increase in net imports from the United States will later become increasingly substantial as a widening of the scale of hostilities, consumption of war material, loss of life, and impairment of productive facilities through enemy bombing and ordinary depreciation make necessary larger dependence on overseas sources of supply… it would appear that our total exports will show only a moderate expansion over the next nine months, that over, say, the next three months the increase in imports associated with inventory accumulation in this country may cause a temporary contraction in our export surplus, and that thereafter our excess of the exports will tend at first to increase only gradually. Over the longer term, however, our net exports will expand at an accelerating rate as the war proceeds. The growth in exports, both at once and over the longer term, will occur primarily in such products as munitions, chemicals, airplanes, machinery, steel, petroleum products, motor trucks, and passenger automobiles, while American agricultural products and light consumer goods, such as textiles, will be comparatively little affected by increased foreign demands.

Despres sent Eccles a memo dated October 12, 1939, entitled "Comments on the Lee Bill for forced loans in wartime" (EP 100/4/3). The provisions of the bill were that individuals would be required to buy Government bonds bearing only 1%, with a 50 year maturity, collected from them based upon their wealth, so as to cover the Government borrowing requirement in case of war or national emergency. Despres expressed strong objection to "this proposed method of war borrowing." According to him, there was the difficulty of gathering data on individuals' wealth and using such data as a measure of "capacity to purchase Government securities"; it was based upon a "confusion between wealth and money"; it was questionable whether such a scheme "provided a desirable instrument for war finance."[8]

In his "Note on Gold," sent to Eccles on October 19, 1939, Despres wrote (EP 49/10/2, 1-2):

> …it seems to me unlikely that the President would give serious consideration at this time to any proposed measure which would substantially impair the ability of Great Britain and France to obtain American goods… Your objective of eventually achieving a rational gold policy would best be served, in my judgement, by merely proposing at this time that we quietly and informally ask the British and French governments, that as far as possible, they finance their purchases of goods from us through the sale in the United States of income-producing assets, and that they to keep their sales of gold to us within reasonable limits…. If these countries, which hold nearly half of the monetary gold outside the U.S., dispose of the bulk of their holdings during the course of a prolonged war, the prospects of return to some form of a international gold standard will be slight.

In a memo to Eccles dated October 20, 1939, and entitled "The War Boom and Some Implications for Policy," Despres wrote (EP 100/5/1; 1,4):

> The influence of the European war upon the American economy will be exerted through its effects (1) upon our exports to foreign countries, both belligerent and neutral, and (2) upon our own equipment outlays, and other public expenditures and receipts. What happens to domestic capital expenditure, consumer outlays, total employment, and national income, depends largely upon the behavior of these two factors… the cumulative stimulus to total

[8]The method of financing WWI in the US was not via imposed war bonds or compulsory loans. See Rockoff (2005).

> activity and employment provided by temporary and abnormal war demands will be much smaller than the stimulus which an equivalent volume of diffused, peacetime expenditure would furnish.

On October 25, 1939, Despres sent Eccles "a first rough draft" of the speech Eccles was scheduled to deliver at the upcoming meeting of the National Industrial Conference Board in New York, the next day. In the speech, he wrote for Eccles, Despres mentioned (EP 84/19/1; 1, 14-15) "the extraordinary rapidity of change resulting from the outbreak of European war"; as the "long-range prospect" the US could achieve "whether or not the war continues"; and " adequate utilization" of "manpower…materials…and…equipment, provided businessmen do not fall into the easy fallacy that a war boom will sooner or later solve all our domestic economic difficulties, leaving no problems for them or for the Government." About a week later, in a memo to Eccles entitled "Outlook for Income-Generating Expenditure in Early Months of 1940," dated October 30, 1939, which according to Despres was "prepared at Currie's request," Despres talked about (EP 100/5/3;1) "the future state of business expectations," their impact on inventory policy, and, in turn, on "income-generating expenditure," depending on "the prospect of an early peace," or "little prospect of early peace."

A fortnight before the new and fateful year for European war, in a memo dated December 20, 1939, Despres dealt with "The problem of the approaching debt limit". He wrote (EP 100/6/3; 5):

> For the duration of the war we can scarcely experience withdrawals of foreign capital not related to goods purchases, and any outflow of capital associated with the restoration of peace and stable political conditions in Europe is likely to be used in the process of reconstruction. We will almost surely return Europe's capital in the form of goods rather than in the form of gold.

Over the period January–April 1940, expectations regarding the possibility of "negotiated peace" as against continuation of the war were difficult to assess. Roosevelt sent Undersecretary of State Welles to Europe in order to ascertain the situation. Interpretations of the object of Welles' extended mission—over February–March 1940 differ (Hilton 1971; Rofe 2007). But perhaps, most important for our purpose here is the letter of March 4, 1940, from Roosevelt to Eccles (EP 2/16/5;/4/13). Roosevelt wrote:

> I have had to decide on refunding $738 million of notes—because honestly I am fearful of the international situation and, confidentially, the news from Welles' visit in Berlin does not make me happier. I think it best to get a little more than half of the refunding out of the way rather than less than half. We can always use a portion of the stabilization fund for debt retirement, later on.

In an untitled memo to Eccles dated March 18, 1940, Despres wrote (EP 49/11/1):

> I have received a letter from Charles Kindleberger, a former assistant of mine now with the Bank for International Settlements, in which he sets forth a private proposal of his own regarding American gold policy. Although there may not be much chance of implementing his proposal at the moment, the suggestion seems to me better conceived and more practical than any of those which I have seen for doing something about gold… Under the system

> which Kindleberger proposes dollar exchange for the purchase of goods whose exports we want to encourage would be more readily available to foreigners, and therefore cheaper, than exchange for the purchase of war goods. Since the demand for war goods is insistent and inelastic, our exports of such goods probably would not be significantly curtailed but we would receive in exchange assets other than gold; our exports to foreign countries now feeling the adverse effects of discriminatory measures abroad would be stimulated. In my judgement, the plan would reduce the gold inflow without depriving us of the stimulus which foreign spending in our markets provides. Moreover, it would induce a more balanced and diversified export pattern.[9]

3 Expectations and Planning for War: April–September 1940[10]

In his account of the events of May 1940 at the FRS cited in the introduction above, Meltzer referred to the BG Minutes of May 21, 1940, regarding the issue of "easy money." *However, there is a crucial missing element in his narrative of the early "wartime" FRS, relating to the specific limitations of the wartime Fed, as recognized by none other than the BG's Chair, Eccles*. This is evident in Eccles' statement, as recorded in the minutes of the joint BG-FAC meeting that took place the *previous day*, May 20, 1940. As this meeting, Eccles set out the constraints on the FRS in the context of "the present critical period… [and] the… unsettled situation resulting from… ***total war in Europe***…and the "urgent problems before the Board" (Minutes, May 20, 1940, 1-2, 4-5). As he put it "***the principal problems created by the war situation were not within the field of responsibility***" of the FRS (Minutes, May 20, 1940, 5) [our emphasis]. He went on to say that "such matters as the control of foreign property and funds of belligerent nations and their nationals, and foreign-exchange, were being handled by the Treasury… that the credit situation was one in which the system was without adequate powers to control reserve requirements and member banks were no longer borrowing from Federal Reserve banks, so that the discount rate was ineffective… and that ***the primary responsibility of the System at the moment was one of exercising a stabilizing influence in the Government securities market"*** (Minutes, May 20, 1940, 5) [our emphases].

[9] Kindleberger served as Despres' assistant at the NY Fed 1936–39, having previously worked at the Treasury. He was at the BIS 939–40 and returned to the USA to work with Despres at the BG/DRS, 1940–42.

[10] Dating this period from April 1940 may seem strange to some. But in terms of setting expectations for the subsequent period up to September 1940, it is more than valid. Indeed, as early as April 9–10, 1940, the Directives of the Joint Planning Committee (JPC) at the War Department's Joint Army-Navy Board (JB) ***assumed defeat of France and Great Britain*** (JB 325; 642, 642–1).

In terms of hostilities, April–September 1940 marked the turning point from "phoney" to "***total***" war in Europe. Ongoing German military success shifted expectations regarding both the duration and possible outcome of the conflict at the BG and elsewhere.[11]

In a confidential memo dated May 22, 1940, entitled "Essentials of American Policy in Event of German Victory" (EP 100/8/3), Despres analyzed possible economic scenarios and economic policy options in this context. Some two weeks later, on June 6, 1940, Walter Gardener also sent a memo to Goldenweiser, Director of the BG/DRS, entitled "Projected Work on US Adjustment to a Possible German Victory" (US National Archives, FRS, RG 82, Box 186).

Because of its importance for our narrative, we cite from Despres' memo at length below. In his introduction to the memo, Despres wrote (EP 10/8/3; 1-4):

> The present note presents certain observations on, rather than a comprehensive analysis of, what American policy should be in the event of a total German victory in the near future. By total victory is meant the destruction or elimination of Allied resistance on the western front in the next few weeks or months… American policy will depend in the last analysis on the details of the peace to be imposed by Germany… Judging by the speed of the movement of the German forces and the comparatively small material damage inflicted on industrial plant and communications, it would appear that the reorganization of the war-torn areas and the rest of the territories acquired will be relatively easy and quick. In other words, practically the whole industrial, commercial, and raw wealth of the regions from Narvik in the Artic Circle to the Cape of Good Hope will remain intact and will serve such purposes as the new masters may agree upon… The dangers confronting the United States are, therefore, not direct… American policy should have a twofold objective, the prevention of infiltration and disintegration through economic pressures and propaganda, and the building up of an armed strength sufficient at least to discourage hopes of a successful invasion. Needless to say, the action called for will be stupendous, but certainly not outside the capacities of this country. Actual will also have to be immediate; it would have to assure the first objective almost at once and achieve the second within eighteen months or two years.

---

[11] Documentary evidence of the shift from optimistic to pessimistic expectations include the fact that the US Army's War Plans Division (WPD) assumed in its Memo to George Marshall, Army Chief of Staff (COS), dated May 22, 1940, a decisive Allied defeat in Europe, followed by German aggression against the Western Hemisphere (WPD/COS, 4175-1). On May 25, 1940, the British Chiefs of Staff advised the War Cabinet that without full economic and financial support from the USA, they did "not think" Britain "could continue the war with any chance of success…it is impossible to say whether or not the UK could hold out in all circumstances" (National Archives, WP (40) 168; 1,10). The US State Department's "Advisory Committee," Chaired by Welles, met on May 27, 1940, with representatives of the US Treasury, Commerce, and Agriculture Departments to consider "post-war" international economic policy, in terms of a "post-war" Europe ***dominated by Germany***. Ex-post, perhaps the best known example of the shift in expectations, was reported by Chuchill himself, who told Roosevelt, as early as May 15, of the "desparate situation…Great Britain might give way…." See Churchill (1949), 23–25. In terms of planning, both American military planners and Intelligence officers expected that Germany would defeat Britain, and thus recommended to US Army COS Marshall that additional arms not be sent to Britain, and rather that he implement rearmament of US forces, see "National Defense Policy," June 17, 1940 (WPD 4250-3, RG 165, US National Archives). By late June, the Army-Navy JB and Army WPD, in a joint planning paper presented to the US Army COS and Navy's Chief of Naval Operations (CNO counterpart of COS) wrote that it was "doubtful that Great Britain…will continue to be an active combatant by Fall/Winter 1940." See Report JPC/JB to COS and CNO, WPD 4250-3, June 26, 1940.

Despres went on to outline "Components of an American Policy". (EP 10/8/3; 4-6); these were: (A) "Diplomatic"; and (B) "Financial and Economic Measures". Regarding "Diplomatic Measures," he suggested "purchase, lease or acquisition" of Anglo-French possessions in both the Atlantic and Pacific; setting up of a "super-confederation" of Western Hemisphere nations; and "the conclusion of a far-reaching agreement with the Japanese concerning Pacific possessions, economic and military matters…even if it meant the cession of the East Indies and Philippines to the Japanese." Depres continued, "If necessary, such a transaction might not be too great a price to pay for covering the Pacific flank of this hemisphere, and releasing the American fleet for service in the Atlantic…Though the Japanese might become independent in petroleum supplies by acquiring the East Indies, they will still require the raw materials and industrial products imported from this hemisphere."

Despres then turned to "Financial and Economic Measures." He outlined five policy prescriptions (EP 10/8/3; 6-10). The first involved reorganization of foreign trade in the Western Hemisphere, that is to say, both North and South America, so as to prevent economic pressure and preserve US foreign trade, and alteration in the composition of trade according to the needs of US foreign policy, military, or defense strategies. The second was to "render various parts of the hemisphere less competitive and more complementary." The third policy change dealt with "the suspension of gold purchases" by the USA and ways for implementing the policy. The fourth was "the impounding of all foreign assets, other than belonging to the countries" of the Western Hemisphere. The fifth policy change involved "heavy expenditures on rearmament."

A week later, in a memo dated May 29, 1940, entitled "The National Defense Tax Program," Despres dealt with the impact of defense expenditure and the possible impact of termination of the European war. He wrote (EP 4/11/3;1-2):

The rate of expansion of our national defense program over the next few years is still uncertain, and it seems likely that it will be limited by the necessity for organizing the production of the specific items needed rather than by the size of appropriations. There is every reason to believe that with the most strenuous efforts to accelerate defense activity, the increase in output over the next twelve months will be small in relation to the idle capacity of our industrial structure a whole.

The increase in defense expenditures, therefore, cannot be relied on to bring about anything like full employment in that period. Meanwhile, domestic business may be subject to two serious shocks. One of these is the possible termination of hostilities in Europe with all that it will mean in terms of cancellation of Allied orders and dislocation of our export trade. The second is a catastrophic decline in stock prices which is certain to be the stock market's reaction to peace or rumors of peace.

Linking expectations and fiscal planning, Despres then wrote:

In light of these possibilities it would be extremely unwise to tighten consumers' belts now… The size of the immediate increase in defense expenditures will not be such as to require hasty enactment of a superficial and ill-considered tax measure. Financial preparedness for a defense program of the scale which can already be foreseen makes it essential, however, to plan a comprehensive revision of our whole

tax structure adapted to our total economic situation as it develops under the gradually mounting impetus of our defense outlays.

With the evacuation of Dunkirk, and the start of the final German offensive to take Paris, Despres sent Goldenweiser, Director of the DRS, a memo dated June 7, 1940, entitled "Preparation of General Economic Program for US in the Event of German Victory" (EP/46/9/5; 1-5) In this, he outlined a number of measures. He first made "suggestions" regarding "additions to staff in connection with the preparation of a general economic program to meet the problems which would be presented by German victory" and recommended Hansen and Kindleberger in this regard. He went on to make suggestions regarding the program itself in the outline form. He divided this into "Internal" and "External Economic Policy" options. Regarding "internal economic policy" planning, he suggested focusing upon the following areas: (i) national defense; (ii) expanding employment programs to include both "unemployed" and "redundant" labor and their retraining for the Defense Program; (iii) developing methods of financing the Defense Program; and (iv) designing methods for preventing "war profiteering." With regard to "external economic policy" options, he first presented his assessment of "Germany's principal advantage in the sphere of foreign trade," which in his view emanated from "her achievement of full employment," which brought about the situation in which "she has not been obliged to seek exports as a means of creating employment and income at home, but rather as a means of obtaining needed imports."

Despres then turned to assess the situation that would face the US and wrote (46/9/5;5):

The impact effect of a German victory upon our foreign trade would doubtless be an abrupt curtailment in our exports. If Germany's succeeds in consolidating her victory and carrying out her program, there is reason to believe, however, that a German-controlled Europe, mobilized on a full employment and full production basis, would be eager to a to expand its imports from the United States and from the whole non-German world. Thus Germany would be ready rather quickly to purchase the output of American industry and agriculture in exchange for the gold and dollar assets of conquered European countries. If we agreed to furnish a market in this country for products whose exportation Germany was disposed to encourage, she would further expand her purchases here. So long as we sought export markets as a means of providing employment and income at home, while Germany's attention was focused, primarily on obtaining imports, our position would be extremely weak. Cessation of gold purchases or the blocking of European dollar assets would arouse the intense opposition of exporting groups in this country, while if we sought to encourage exports to German-controlled Europe as a means of sustaining employment and incomes at home, we should give to Berlin an important means of influencing our own level of internal economic activity.

Given the assumption that the U.S. economy "will have solved" the problem of its "underutilization of resources," Despres then made the following policy suggestions (EP 46/9/5; 5-6): (i) discontinuation of gold purchases; (ii) blocking of foreign assets; (iii) developing alternate sources of supply for products from problematic areas, such

as Asia and Africa; (iv) development of a balanced policy toward Latin America regarding exports and export markets.

A week later, on June 14, 1940, Despres forwarded to Eccles a detailed memo by Krost, who worked in the section of the DRS headed by Despres. The memo by Krost was dated June 3 and entitled "Financing National Defense" (EP 24/11/9). Krost's memo, and its proposals regarding tax changes, started a month-long series of DRS memos to Eccles from Despres and Hansen, which outlined their views on the matter of a national defense tax program. These included memos from Despres dated the day before, June 13, entitled "National Defense Tax Bill" (EP 4/11/8), and his memo of the following month, dated July 22 (EP 24/12/1), to which was attached Hansen's "Notes on Proposals for Increased Taxes" (EP 24/12/3). In his memo of July 22, after suggesting implementation of "an excess profits tax" and "removal of tax *exemption* for future issues of Federal, State, and local securities," Despres went on to provide some estimates of what he saw as the limited impact effect of defense expenditures on economic expansion "leaving aside the possibility" of the USA "becoming a belligerent." He concluded (24/12/1; 2) that "Under these circumstances it seems appropriate to postpone drastic increases even in nonconsumption taxes until a substantial amount of general economic expansion has been achieved." Hansen, in his "Notes" for Eccles dated July 22 wrote (24/12/3; 1-2):

We should be concerned lest, under cover of financing the new defense program, we balance the entire Federal budget too rapidly… The defense expenditure in fiscal 1941 will probably be relatively small…and the relatively small expansionist effect may largely be offset by unfavorable trends in foreign trade and inventories and by a sharp increase in taxes… At the moment we ought to rely heavily on borrowing from the public. Later, when incomes have risen considerably and we have approached more nearly full employment, we ought to stiffen the excess profit tax, the estate tax, the corporate income tax, and the individual income tax—all designed to tap the savings stream. Finally as it becomes necessary to divert production resources from consumption to defense expenditure, we shall have to impose consumption taxes.

4 Expectations and Planning for War: September 1940–December 1941

After Dunkirk and the fall of France, the "Battle of Britain" began, as Churchill called it in his speech to Parliament on June 18, 1940. But by mid-September 1940, the nature of this "Battle" had changed; the "Blitz" had commenced, and the serious military situation ***directly*** threatening Britain had more or less stabilized. And, with the failure of the German effort to destroy Britain's air defense system, even the proposed invasion of Britain was called off. But the war continued in other areas of hostilities. Italy had declared war on France and Britain on June 10, 1940; Italian and German forces fought the British in the Mediterranean and North Africa, with Italian forces invading Egypt in mid-June 1940.

At the end of September and beginning of October 1940, a formal "Defense Program" at the BG was underway. By this time, Hansen himself had produced reports on defense expenditure in relation to the ongoing war which, as he put it, impacted "upon the American economy with special reference to problems of concern" to the BG and FRS (EP 100/9/1). In his covering document dated Oct 3, 1940, Hansen outlined his reports, all dated September 25, 1940, and attached a list of other Defense Program memoranda emanating from the BG/FRS, in addition to those produced by other Government Departments and agencies (EP 100/9/1; 2-5). Among those produced by economists at the BG/FRS were "Immediate Prospects for Trade in War or Peace" and "Gold, Blocked Balances and Exchange" (Gardner); "Direct Methods of Price Control" and "Gold, Blocked Balances and Export Controls during the War" (Despres); "Gold Policy" and "Future Gold and Trade Policy" (Hansen); "German Armament and War Financing (Kindleberger); "Financing National Defense" (Krost); and interestingly enough, two memos by the Vice President for Research at the Minneapolis Fed, Arthur Upgen entitled "The Future Position of Germany and the US in World Trade" and "The Resources of Germany and the US."

Hansen (EP 100/9/1; 5-9) also listed and described the conferences and meetings between BG/FRS officials and other Government economists "dealing with various aspects of the Defense Program and its impact on the economy." He outlined five major areas of research in this context: (i) taxation and fiscal policy; (ii) impact of the Defense Program on national income and employment; (iii) the Defense Program and labor supply; (iv) gold and foreign trade policy; and (v) the Defense Program and prices.

In his "Preliminary Report on the Defense Program and the National Economy" dated September 25, 1940, Hansen "discussed the general situation confronting" the US economy "at the point when the Defense Program begins; the magnitude and timing of the Defense Outlays, together with the probable effect of these outlays on national income, output and employment; …the appropriate measures of financing defense expenditure…; the danger… [of] the defense program… running into general inflation… and the various means of meeting the situation; post-war slump and how to deal with it; and finally, long-range fiscal policy" (EP 100/9/1; covering letter October 3, 1940, 1). In the body of his report (EP 100/9/1; (6) 16) Hansen wrote: "The international situation might, of course, easily become so menacing that we should be compelled to push on as rapidly as possible to a full defense effort …This, however, would involve a drastic increase in the defense program as now contemplated."

Hansen then presented three alternate scenarios and their possible economic impact on the US "Defense Program": (i) "negotiated peace"; (ii) German conquest of Britain; (iii) Britain withstanding attack and an indefinite continuation of hostilities.[12] With regard to the first scenario, Hansen wrote (EP 100/9/1; (8) 18):

[12]Hansen's "peace" alternative probably emanated from Hitler's Reichstag "Last Appeal to Britain" speech of July 19, 1940, at the start of the "Battle of Britain," and the subsequent speech of his Economics Minister, Funk, to the Foreign Press a week later, on July 25, 1940, in which he proposed "the economic reorganization" of Europe. Indeed, Keynes was asked to prepare a reply to Funk's proposals, which he wrote in late November 1940. This was circulated in December 1940, with a revised version in January 1941. It was used in a speech by British Foreign Secretary Eden in

"In the event… of a negotiated peace—say next spring—in which England emerged as the undisputed sea power, we should probably feel relatively safe even though Germany were left in command of the European continent. The present defense program might, therefore, appear adequate, while, on the other side, the situation would not be sufficiently safe to warrant its abandonment or curtailment." He went on to say (EP 100/9/1; (8) 18-(9) 19): "Two other eventualities are however equally, or perhaps even more, probable. On the one side Germany may conquer England. In this event it would appear certain that we shall step up our defense expenditures with utmost speed… this would definitely put us on a war basis so far as expenditures are concerned." Regarding the third possibility, he said: "On the other side, England may withstand the attack and continue the war indefinitely. This outcome presents two alternatives for us. Either we shall speedily enter the war, or we shall continue to give maximum support to England short of war, meanwhile rapidly increasing our own military strength."

Hansen concluded this section of his report by saying: "Whatever the various alternative outcomes (excluding entrance into the war) the [defense] expenditures in fiscal 1941 are not likely to fall… If we actually enter the war these upper limits will of course be exceeded by a wide margin."[13]

By mid-May 1941, Hansen, now working at the BG and advising the National Resources Planning Board, had drafted a study dealing with the problem of, as he put it "Post-Defense Full Employment" (EP 62/2/2; May 14, 1941). At a meeting of the BG on July 24, 1941, Hansen presented the issues surrounding economic coordination between the USA and Canada with respect to defense production, "reducing the probable post-war economic dislocation consequent upon the changes" which both economies were "presently undergoing" and preparation of "reports concerning long-range studies looking to the post-war period." Kindleberger served as Secretary of the Joint US-Canadian Committee set up by Rossevelt and Canadian Prime Minister King in April 1941 to deal with these issues (Minutes, July 4, 1941, 6-7). Indeed, the term "post-defense" characterized the title of a number of studies, including one jointly authored by Currie, at the White House, and Krost, at the BG. In this study, dated October 9, 1941, and entitled "Fiscal Policy in the Post-Defense Period" (EP 100/10/13; 1), Currie and Krost outlined what they called "a post-defense fiscal program" which was, in their view, "a presentation of the type of fiscal structure" which they thought the USA was "likely to have at the end of the war" (EP 100/10/13; 1).

May 1941. See Moggridge (1992, 653–54). Earlier, however, in August 1940, Bernard Baruch had written a reply to Funk which he sent to Roosevelt. On August 20, 1940, Roosevelt's aide, "Pa" Watson, sent Baruch's reply to Eccles for his comments (EP 7/7/4). For a vivid first-hand account of the speeches of Hitler and Funk, and the impact on German public opinion, and that of the German leadership, including Hitler, of the British reaction to Hitler's diatribe in the form of the reply by British Foreign Secretary Halifax, see Shirer (1941) diary entries for July 19 through July 25, 1940, inclusive.

[13] In his report of September 25, 1940, Hansen actually provided ***detailed estimates*** of expenditures under the "Defense Program" ***based upon these alternative scenarios***, ***with projections up to Fiscal Year 1950***. See Hansen (EP 100/9/1; (6) 16-(9) 19), especially table "Defense Program" at the top of page (8) 18.

5 Summary and Conclusion

To sum up, over the period 1939–1941, activities at the BG/FRS were not only limited to debates over "easy money" and the problem of inflation. Rather, operating in an expectations-planning ***nexus*** that underwent significant shifts from the start of European war in September 1939 to the entry of the USA into World War December 1941, economists at the BG/FRS wrote numerous policy memos on the impact of the European war on the US economy and conducted research on a wide range of war-related issues in the context of the "Defense Program" as outlined above. The scope and tone of these memos reflected the shifts in expectations as to the duration of the war; the possibility of a negotiated peace; the possibility of a German victory; the possibility of an imposed peace; and the possibility of the conflict continuing. In this context, the ***meaning*** of the term "post-war" shifted a number of times, from the situation ***following*** a possible negotiated peace, a possible German victory, or the possibility of an imposed peace on Britain accordingly.

Once the USA entered the war in December 1941, the nature of the expectations-planning ***nexus*** at the BG/FRS ***once again*** shifted, so that by late 1942 and early 1943, it refocused on "post-war" and "post-defense" issues such as international stabilization, reconstruction, and US employment in the post-war period, as did other Government bodies such as the Treasury, Commerce, and Labor departments.

References[14]

Baruch, B. (1940, August). Reply to Funk (EP 7/7/4).
Board of Governors, Minutes, May 21, 1940.
Board of Governors, Minutes, July 24, 1941.
Bordo, M. (2010, March). *The Federal Reserve: Independence gained, independence lost*.... Shadow Open Market Committee.
Bordo, M. (2016). Some historical reflections on the governance of The Federal Reserve. In J. Cochrane & J. Taylor (Eds.), *Central Bank governance and oversight reform* (pp. 221–254). Stanford: Hoover Institution.
Currie, L., & Krost, M. (1941). *Fiscal policy in the post-defense period* (EP 100/10/13), October 9, 1941.
Despres, E. (1939a). *Tentative appraisal of the economic effects of European war upon the United States*, (EP 49/10/1), September 30, 1939.
Despres, E. (1939b). *Comments on the Lee Bill for forced loans in wartime*, (EP 100/4/3), October 12, 1939.
Despres, E. (1939c). *Note on gold*, (EP 49/10/2, 1-2), October 19, 1939.
Despres, E. (1939d). *The war boom and some implications for policy*, (EP 100/5/1), October 20, 1939.
Despres, E. (1939e). Speech for Eccles. IN *National Industrial Conference Board*, (EP 84/19/1), October 25, 1939.
Despres, E. (1939f). *Outlook for income-generating expenditure in early months of 1940*, (EP 100/5/3), October 30, 1939.

[14]Note: EP refers to Eccles Papers available on the FRASER Web site.

Despres, E. (1939g). *The problem of the approaching debt limit*, (EP 100/6/3; 5), December 20, 1939.
Despres, E. (1940a). *Memo to Eccles*, (EP 49/11/1), March 18, 1940.
Despres, E. (1940b). *Essentials of American policy in event of German victory*, (EP 100/8/3), May 22, 1940.
Despres, E. (1940c). *The National Defense Tax program*, (EP 4/11/3), May 29, 1940.
Despres, E. (1940d). *Preparation of general economic program for US in the event of German victory*, (EP/46/9/5), June 7, 1940.
Despres, E. (1940e). *National Defense Tax bill*, (EP 4/11/8), June 13, 1940.
Despres, E. (1940f). *Memo*, (EP 24/12/1), July 22, 1940.
Eccles, M. (1966 [1951]). *Beckoning frontiers* (2nd ed.). Knopf: New York.
Federal Advisory Committee, Minutes, May 1-21, 1940.
Gardner, W. (1940). *Projected work on US adjustment to a possible German Victory*, June 6, 1940 (US National Archives, FRS, RG 82, Box 186).
Hansen, A. (1940a). *Notes on proposals for increased taxes* (EP 24/12/3), July 22, 1940.
Hansen, A. (1940b). *Preliminary report on the defense program and the national economy* (EP 100/9/1), September 25, 1940.
Hansen, A. (1940c). *Covering letter to "Report of Defense Program"* (EP 100/9/1), October 3, 1940.
Hansen, A. (1941). *Post-defense full employment* (EP 62/2/2), May 14, 1941.
Hetzl, R., & Leach, R. (2001). The treasury-fed accord: a new narrative account. *Federal Reserve Bank of Richmond Economic Quarterly, 87,* 33–55.
Hilton, S. (1971). The Welles mission to Europe, February–March 1940: Illusion or realism? *Journal of American History, 58,* 93–120.
Jones, B. (1972). The role of Keynesians in wartime planning and postwar planning, 1940–1946. *American Economic Review, 62,* 125–133.
Krost, M. (1940). *Financing national defense*, (EP 24/11/9), June 3, 1940.
McGregor, R., & Young, W. (2013). Federal Reserve Bank Presidents as public intellectuals. *History of Political Economy, 45,* 166–190.
Meltzer, A. (2003). *A history of the Federal Reserve, Volume 1, 1913–1951*. Chicago: University of Chicago Press.
Moggridge, D. (1992). *Maynard Keynes: An economist's bibography*. London: Routledge.
National Archives. (1939). National Archives, CAB/65/1/9, War Cabinet 9(39), Minute 15, 68, September 9, 1939.
Notter, H. (1949). *Postwar foreign policy preparation, 1939–1945*. US Department of State, Office of Public Affairs.
Rockoff, H. (2005). Until it's over there: The U.S. economy and WWI. In S. Broadberry & M. Harrison (Eds.), *The economics of WWI*. Cambridge: Cambridge University Press.
Rofe, J. (2007). *Franklin Roosevelt's foreign policy and the Welles mission*. London: Palgrave-Macmillan.
Rolnick, A. (1985). Research activities and budget allocations among Federal Reserve Banks: Comment. *Public Choice, 45,* 193–195.
Roosevelt, F. D. (1940). *Letter to Eccles* (EP 2/16/5), March 4, 1940.
Sandilands, R. (1990). *The life and political economy of Lauchlin Currie*. Durham, NC: Duke University Press.
Shirer, W. (1941). *Berlin diary*. New York: Knopf.
Welles, S. (1951). *Seven decisions that shaped history*. New York: Harper.
Young, W., & McGregor, R. (2015). Fed vs. Fed: Some essential tensions. In *Presented at Guggenhein Conference*, Geneva, 2015.

Israel's Struggle Toward Macroeconomic Stability: Role of Inflationary Expectations

Assaf Razin

1 Introduction

The State of Israel has registered remarkable economic achievements over its lifetime. It emerged in 1948 as a rather weak and impoverished agricultural economy. Over the past seven decades, though Israel thoroughly transformed itself into a strong and wealthy industrial economy, one has become a world leader in many areas of high technology, ranging from computers to medicine, as attested to by its membership in the Organisation for Economic Co-operation and Development (OECD). Thanks in large part to its steadily advancing integration into the global economy, Israel has moved firmly out of the developing world and into the developed world.[1] Israel's strong growth since it stabilized inflation in 1985 owes much to an international economy in which capital, labor and ideas are mobile and in which trade and investment flow readily across far-flung international borders.

The essay begins with an inquiry into the causes and consequences of the hyperinflation that rocked the Israeli economy in the mid-1980s, as well as into the stabilization measures instituted by Israeli policymakers that eventually after a decade and a half tamed the problem. The reduction in inflation, coupled with the mass migration of highly skilled immigrants from the former Soviet Union throughout the 1990s, allowed the Israeli economy to gather a head of steam. The worldwide financial crisis that began in 2008, which was followed by the "Great Recession," had only a minor and fleeting effect on the Israeli economy in comparison with many other advanced economies; in large part because of the financial, fiscal and monetary reforms Israeli policymakers had put in place to contain the hyperinflation of the mid-1980s.

[1] See Razin (2018a, b).

Based on Razin (2019).

A. Razin (✉)
Tel Aviv University, Tel Aviv, Israel
e-mail: razin@post.tau.ac.il

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_11

The essay offers an economic history perspective of the long struggle with inflation. It covers the early acceleration to three-digit levels, lasting 8 years; the stabilization program, based on political backing triggered sharp fall in inflationary expectation, and consequently to sharp inflation reduction to two-digit levels; the convergence to the advanced countries' levels during the "Great Moderation"; and Israel' resistance to the deflation–depression forces that the 2008 crisis created. The emphasis is on the forces of globalization and the building of institutions, political, regulatory, financial, budget design and monetary, which helped stabilize prices and output. Analysis identifies the crucial role played by inflation expectations in constraining policymakers regulating inflation pressures.

Globalization, the integration of markets in goods, services and capital, whose pace accelerated in the 1990s with the fall of communism, is currently under attack. Globalization and new technology forces accelerated the decline in low-tech manufacturing industries, the rise of the financial and the surge of immigration. Brexit may have been the first wave of anti-globalization and rising populism that gushes over most advanced nations. Then came the 2017 change of guards in the USA. Meanwhile, European countries, straightjacketed inside the confines of the single currency area like Germany, France, Greece, the Netherland, Spain, Poland and others, witness that the anti-EU forces gather strength. Israel in many ways provides a counterexample. Globalization technology forces, like the technology surge, the new markets in East Asia and immigration waves, have been a boon. Israel's remarkable developments provide an historical counterexample.[2]

Israel has had a remarkable development, emerging from a low-income high-inflation developing economy in the 1970s, to a medium- to high-income advanced economy in the 2000s; at all stages of its development, globalization played a key role.

The paper is organized as follows. Section 2 outlines the inflation crises. Section 3 discusses the political–economy underpinning of the inflation crisis. Section 4 discusses the end to the political deadlock. Section 5 analyzes the interactions between inflation and fiscal revenues. Section 6 analyzes balance-of-payment crises. Section 7 emphasizes the role of globalization in taming domestic inflation. Section 8 describes the climbdown from two-digit to one-digit inflation rates. Section 9 observes the convergence of Israel's inflation rates to those of the industrial world economy. Section 10 analyzes the resistance in Israel to the depression–deflation forces coming from the "Great Recession." Section 11 concludes.

[2]Israel's fast development although unique is not unknown elsewhere. Ireland somewhat parallels Israel in awesomely benefitting from globalization. Ireland entered the 1950s as a very poor post-colonial society. However, it realized major successes by the integration into the EU and reaching an elite hi-tech status. Ireland was able to attract from the rest of the world (other than the EU) massive FDI, thanks to it being a tax-sheltered gate to the EU massive markets. Ireland, however, had relatively little regulation of its banking sector: This allowed the credit bubble to flourish in the wake of the 2008 global financial crash. It is overexposed banking sector collapsed during the financial crisis. Ireland has continued to be burdened by the eurozone's nearly secular stagnation. Israel's robust performance during the crisis is partly attributable to not being a member of a single currency area.

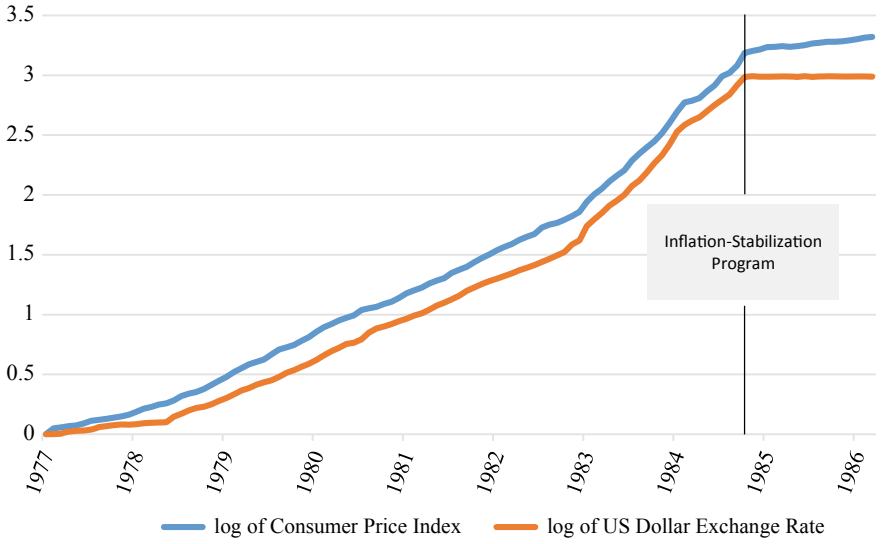

Fig. 1 Price level and exchange rate, 1977–1986. *Source* Bank of Israel

2 Inflation Crisis

Israel's inflation accelerated in the 1970s, rising steadily from 13% in 1971 to 111% in 1979. Some of this higher inflation was "imported" from the world economy, instigated by extreme oil price rises in 1973 and 1979. Inflation kept gathering pace. From 133% in 1980, it leaped to 191% in 1983 and then to 445% in 1984, threatening to become a four-digit figure within a year or two. After several failed efforts, the successful phase of the stabilization of the Israeli economy began with the heterodox program introduced in July 1985. The initial success of the stabilization program included a decrease in inflation, from 445% in 1984 to 185% in 1985 and 20% in 1986. There was also an increase in real economic activity, with the annual growth in business sector product per capita rising from 0.4% in 1984 to 4.3% in 1985 and 3.6% in 1986. However, in the second half of 1987, the economy slid into recession, an after-shock event. Inflation, however, did not converge to advanced countries' inflation.[3]

Figure 1 describes the price level and the exchange rate paths for the inflation-rising period in the wake of the hyperinflation crisis and the aftermath of the 1985 stabilization program.

[3]Calvo and Vegh (2001) observe that in many high-inflation stabilization programs around the world inflation failed to converge to world averages. Real economic activity expanded in the early years of the stabilization program. Later in the program, a recession set in. Unlike Israel stabilization program, in many developing economies the program later collapsed.

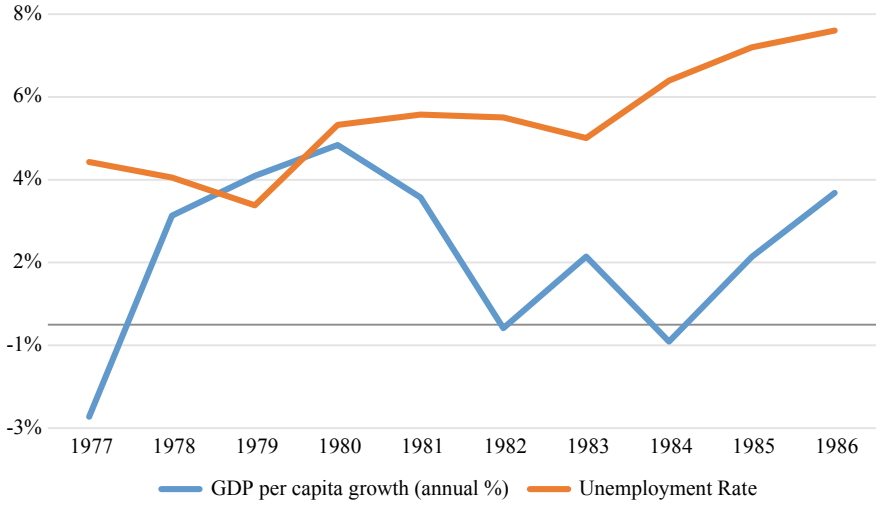

Fig. 2 Output growth and unemployment, 1977–1986. *Source* Israel CBS, World Bank

The figure demonstrates the accelerated path of inflation and the lagging path of exchange rate depreciations in the 1980s. It highlights the sharp flattening of inflation that took place immediately after the implementation of the 1985 stabilization policy package. All along, the depreciations fell short of inflation; therefore, the real exchange has been markedly appreciated throughout the period. The real exchange rate appreciation naturally corresponds also to the rise in unemployment and output growth decline. Figure 2 describes the path of major output-and-employment indicators. They point out to severe slackness in economic activity during the hyperinflation crisis leading to unprecedented unemployment.

Economic activity was impacted severely by the swelling credit frictions because the inflation crisis undermined the well functioning of credit institutions. Banks and financial market regulation also failed during the 1980s. At the time, bank stocks accounted for more than 90% of all issued stocks in the stock market. Their monopoly power in the stock market allowed the large banks to build up low-cost loan portfolio and give it out to borrowers with poor selection and poor monitoring. Central bank oversight of commercial banks was almost nonexistent. Israel's Securities and Exchange Commission was powerless, legally and administratively. Massive stock issues allowed banks to increase their available capital as a source of investments, loans, etc. To get market participants to continue investing in the large bank's stock, it began buying back its own stock. On October 6, 1983, known as the "Black Thursday," an onslaught of banks' stock sales brought down the stock market. Largest banks became state-owned through a swift bailout.

3 Political–Economy Policy Shift

The economic crisis started to develop when the opposition "*Gahal*" (now "*Likud*") party gained power for the first time since independence. The political upheaval in 1977, the so-called *Maapach*, was a game changer for economic policy in Israel. The newly elected government, adopting macroeconomic populistic policy, abruptly switched away from a long-running economic regime, which had been able to maintain fiscal discipline in the presence of strong external shocks (the Yom Kippur War and the first oil crisis).[4] Monetary policy was moderately accommodative, underpinned by a fixed exchange rate regime and shielded from capital flights by capital controls. Notwithstanding the oil price shock, inflation was low double digit.

A useful way to understand the framework within which the economic policy was conducted prior to the political regime switch, and afterward, is to think about the basic trilemma in international finance.[5] In international finance, the trilemma stems from the fact that, in almost every country, economic policymakers would like to achieve the following goals: First, make the country's economy open to international capital flows, because by doing so, policymakers of a country enable foreign investors to diversify their portfolios overseas and achieve risk sharing. The country also benefits from the expertise brought to the country by foreign investors. Second, use monetary policy as a tool to help stabilize inflation, output and the financial sector in the economy. This is achieved as the central bank can increase the money supply, reduce interest rates when the economy is depressed, reduce money growth and raise interest rates when it is overheated. Moreover, the central bank can serve as a lender of last resort in case of financial panic. Third, maintain stability in the exchange rate. This is because a volatile exchange rate, at times driven by speculation, can be a source of broader financial volatility and makes it harder for households and businesses to trade in the world economy and for investors to be able to plan.

---

[4]Dornbusch and Edwards (1989) address macroeconomic populism in Latin America, which they roughly defined as policies that are favoured by a substantial part of the voting population, but which ultimately harm the majority of the population. They found that populism surfaces when the economy has endured a period of external shocks and domestic upheavals, and "a highly uneven income distribution usually presents a serious political and economic problem, providing the appeal for a radically different economic program." In the first phase after their policies are enacted, the populists are vindicated. Growth and wages indeed rise as a combination of profligate spending and intrusive government control does expand the economy. The surging government spending and mandated wage hikes tend to produce a temporary "sugar high," followed by a crash. Beneath the surface, however, the country's economic potential is deteriorating. Financial disorders appear. Rather than making the hard choice of returning to principled economic oversight, the populist leader recommits to harmful policies and steers the country toward decline, capital flight and sometimes debt crises. In all cases, write Dornbusch and Edwards, "There were disastrous effects for those groups who were supposed to be the beneficiaries of the policies."

[5]The trilemma as a situation in which someone faces a choice among three options, each of which comes with some inevitable problems. In international finance, it is cast in terms of economic regime choices. The international finance trilemma goes back to the classical works of Flemming (1962) and Mundell (1963). See Mankiw for blog interpretation (2010). For a balance-of-payment crisis model in the trilemma regime-switch framework.

The problem, however, is that a country can only achieve two of these three goals. In order to maintain a fixed exchange rate and capital mobility, the central bank loses its ability to control the interest rate or equivalently the monetary base—its policy instruments—as the interest rate is anchored to the world interest rate by the interest rate parity, and the monetary base is automatically adjusted. This is the case of individual members of the European Monetary Union. In order to keep control over the interest rate or equivalently the money supply, the central bank has to let the exchange rate float freely, as in the case of the USA. If the central bank wishes to maintain both exchange rate stability and control over the monetary policy, the only way to do it is by imposing capital controls, as in the case of China.

Following the 1977 political change, the economic regime switched from pegged exchange rate, capital controls and fiscal discipline to loosely managed exchange rate, relaxed controls on outgoing capital flows and fiscal laxness. Right from the beginning, the new government lifted some capital controls without putting safeguards in place; that is, no prudent financial and banking regulatory measures existed. Intensive shifts in demand and supply for foreign exchange followed almost instantly. Key to the steady increase in inflation, the new populistic government also embarked on an uncontrolled path of fiscal expansion accommodated by monetary expansion. Exchange rate and capital flow fluctuations called for the Bank of Israel to intervene occasionally, at first, and significantly later, in the foreign exchange market on a day-to-day basis to smooth out these fluctuations. A massive wave of capital flight caused over a few years a fast depletion of the stock of international reserves, which weakened the ability of Bank of Israel to intervene in the foreign exchange market. Therefore, Israel lost control over inflation. Lax safeguards brought stock market crashes. The lesson learned from the first-generation currency crisis literature is that such inconsistent set of policies become quickly unsustainable, leading to massive speculative attacks on international reserves and followed rounds of financial and stock market crashes.

4 Populism and Seigniorage Finance

Israel's high inflation posed sharp challenges to both political and economic institutions. Failing economic governance made it essential for the government to raise revenue through money expansion. At the time when the newly elected government was catering to populistic demands, the printing press was used to finance the fast-expanding government spending and transfers.[6]

[6]The temptation to inflate during the 1977–1985 high inflation period was moderated by the fact that, due to a preexisting stock of nominal loans by government to the private sector and the Olivera-Tanzi effect, government revenue from seignorage was much smaller than would appear to be the case at first blush, and at times, perhaps even negative. The jump to a high inflation plateau was due to a series of policy actions or inactions. Once inflationary expectations adjusted upward, this process became a persistent feature of the economy making it costly to stabilize due to the expected

Dividends from seigniorage (the profit made by a government by issuing currency) are derived from the exclusive ability of the central bank to issue banknotes. In addition, central bank can hold required reserves from commercial bank deposits, which pay no interest. Central banks can also inflate the non-indexed portion of the public debt and raise the real revenue intake with progressive tax schedule. However, how much can the central bank lower the consolidated government fiscal burden depends not just on how actual inflation is consistent with expected inflation.[7]

A central lesson from Friedman (1971) is that steady-state seigniorage from revenue-maximizing central bank is small. However, Israel, as well as previous historical episodes, offers a counterexample. Inflation spikes can be a significant source for government revenue. Time inconsistency on the part of the central bank in producing these spikes is due to harmful incentives. They lead policymakers to implement inflation levels that they may eventually come to regret. These incentives are no rarity; they are very common in economies that do not have the instruments to reach a first-best equilibrium. Moreover, these incentives cannot be ruled out, even under rational expectations in such a time-inconsistent setup.

One crucial issue about inflation is to identify whether existing inflation is temporary in nature, reflecting short-term spikes, or whether it is useful to analyze it as if it is a steady-state phenomenon. The steady-state seigniorage curve in Fig. 3, which shows two distinct (steady state) inflation levels for a given amount of seigniorage, led some economist to attribute the cause of the high inflation to just expectation-based phenomenon that can be rectified by synchronizing wages, prices and exchange rates alone.[8]

On this issue, Calvo (2016) writes: "Repeated use of surprise inflation is unlikely to be successful in increasing seigniorage, because the public will start to expect a rate of inflation larger than the one that optimizes steady-state revenue from inflation. Thus, eventually the economy may land on the excessive-inflation territory highlighted in Friedman (1971). However, this is not due to an elementary economics error on the side of the central bank, as Friedman's results might lead us to conclude. An

reduction in economic activity associated with stabilization. The traditional analysis of steady-state seignorage appears therefore of secondary importance, and I would consider dropping it or drastically reducing it.

[7] Indeed, Karni (1983) made rough calculations and found significant seignorage revenue that Israel's hyperinflation generated. In a related context, Cukierman brings evidence for a significant share of revenue attributable to seigniorage in the 1920s' German hyperinflation.

[8] Liviatan (1984) offered a heuristic explanation for the nature of Israel's hyperinflation based on "inflation inertia." Inflation inertia, he argued, is caused due to the government's periodic attempts to boost exports, the indexation of wages to the cost of living and the adjustment of public expectations to the vicious cycles. Each time the government devalued the currency to support exporters, prices rose and wages followed. Liviatan suggested using the US dollar as an anchor by fixing the Israeli shekel exchange rate to it: This "will lower inflation to the level of the inflation in the USA," and it would not require the use of administrative and coercive powers which "undermine the foundation of the liberal regime." Bruno and Fischer (1984) argue that contrary to the orthodox economic view, for a similar level of public spending, there are multiple levels of inflation due to the "metastable equilibrium" that is caused by indexation, the financial structure and the exchange rate system.

See Krampf (2018).

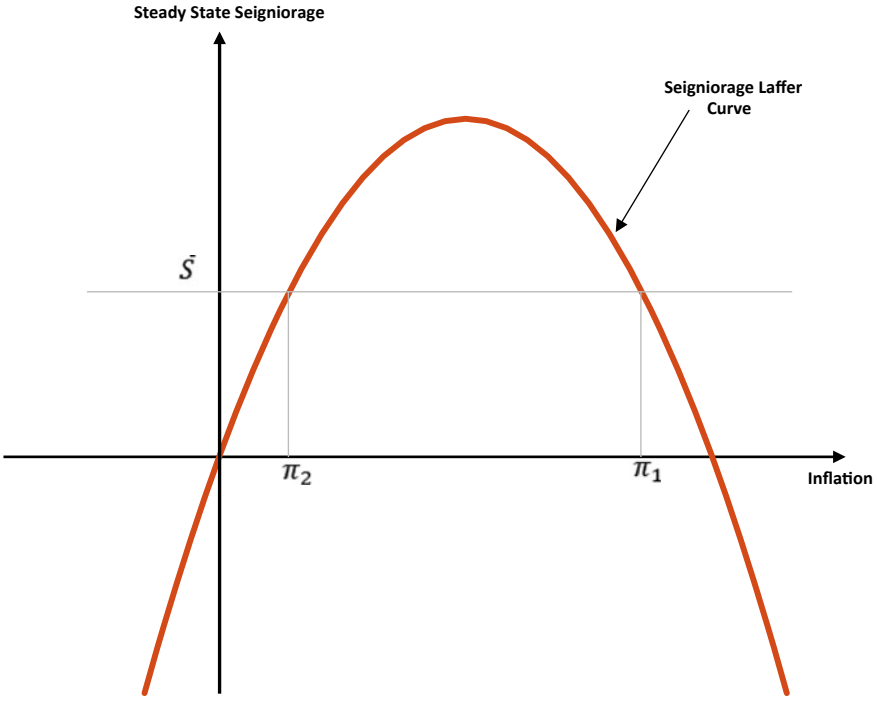

Fig. 3 Steady-state seigniorage and inflation

inflation spike is, in the short run, one of the cheapest and most expeditious manners for securing additional fiscal revenue. Moreover, this "carrot" is always there. As noted, though, a problem arises if the government repeatedly reaches out for the carrot. However, even in this case, the evidence presented in Friedman (1971) does not prove that authorities were making an error. To assess that, one needs information of how quickly the public catches up with the inflation-spike strategy."

Even in the time inconsistency paradigm,[9] however, there is room for policy. One could try to neutralize the harmful incentives if the central bank banned from extending loans to the fiscal authority. Following almost 8 years of the hyperinflation economic chaos, from 1977 to 1985, the Israeli voters brought about some major political rebalancing toward the political center. The newly established unity government ("Likud" plus "Avoda") implemented successfully key stabilization measures; all of them required political consensus.[10] Following 8 years in which the

[9] See Kydland and Prescott (1977) and Calvo (1978).

[10] Schneider and Tornell (2004) provide a model of boom-bust episodes in middle-income countries which may explain the logic of the boom-bust episodes which followed the stabilization program. It is based on sectoral differences in corporate finance: The non-tradable sector (e.g., real estate and financial services) is special in that it faces a contract enforceability problem and enjoys bailout guarantees (e.g., bailing out mortgages). As a result, currency mismatch in the balance sheet arises endogenously in that sector. This sectoral asymmetry allows the model to replicate the main features

seigniorage served as a means of financing the deficit, a new legislation ("*Khok Hahesderim*") allowed the government to exercise tighter control over its spending and taxation. A new law forbade the central bank to monetize the budget deficit ("*Khok Iee Hadpassa*") and ended the accommodating monetary policy. A tri-party agreement between the government, the Federation of Labor ("*Histadrut*") and the association of private-sector employers stabilized the wage-price dynamics and enabled a sharp nominal devaluation that ended in a competitiveness-boosting real devaluation. The exchange rate depreciation had not passed through to wages and prices; in high likelihood because the entire macro-regime has changed; and as in the modern expectation-based macroeconomic setup. Because of the *credibility* of the policy measures, backed by the tri-party agreement between the unified government, the Federation of Labor and the industry employer organization, and the greater independence of the central bank, inflation expectations adjusted rapidly to the policy steps.[11]

5 The Distributive Effects of Inflation Stabilization

Sargent (1999) argues that high inflation can be stopped quickly and at a low cost. His argument is that inflationary expectations are quick to adjust when the economic regime shifts considerably. However, he ignores the fiscal burden and the income distribution that follow.

What are the fiscal implications of deep-rooted inflation expectations, before hyperinflation is stopped?[12]

Stopping hyperinflation has major distributive implications. This explains why a cross-party government, where each party represents different economic interests, is often needed to enact credible sustainable policies. To understand the essentials of these matters, imagine a simple economy where there is a stock of public debt denominated in domestic currency, D.[13] We denote one-period nominal interest rate by i. Then, the next-period full service of the government debt (i.e., principal plus interest) will be $(1+i)D$. We choose the units of measurement so that the present

---

of observed boom-bust episodes. In particular, episodes begin with a lending boom and a real appreciation, peak in a self-fulfilling crisis during which a real depreciation coincides with widespread bankruptcies and end in a recession and credit crunch. Israel's economy fully recovered in the late 1980s and early 1990s, a time when there was a new wave of immigrants from the former Soviet Union (Chap. 2). For detailed account of the stabilization policy, see Razin and Sadka (1993).

[11] This expectation-changing episode is akin to Volcker policy effect on inflationary expectations in the USA; see Sargent (1999).

[12] With the benefit of hindsight, it can be concluded that the 1985 cold turkey stabilization produced a large permanent drop in the rate of inflation. However, at the time of the stabilization, there was substantial uncertainty about the extent to which this dramatic drop would persist. The uncertainty was induced by wide gyrations in inflation and several failed attempts to stabilize prior to the 1985 successful stabilization. See Cukierman et al. (2018).

[13] See Calvo (1992).

price level is equal to one, and assume that the real interest rate is equal to zero. We also denote the one-period expected inflation rate, π^e, so that inclusive of the inflation premium, the nominal (gross) interest rate is $i = 1 + \pi^e$, and the next-period price level is equal to $1 + \pi^e$. If the government surprise market participants by setting the actual inflation rate equal to zero, so that the actual bond return–gross return is equal to one, the actual *real* burden of servicing the next-period debt is equal to:

$$\left(1 + \pi^e\right) D.$$

On the other hand, if the government fulfills the private sector entrenched inflationary expectations and set the actual inflation equal to expected inflation, the *real* burden of the debt is just D.

Thus, a temptation **not** to stop inflation in its tracks may be irresistible.

Similarly, if the government surprise market participants by abrupt stopping of hyperinflation in the presence of entrenched inflation expectations, the fiscal burden of public sector wage bill and subsidies to basic food must rise. Therefore, the government may hesitate to do so.

To overcome this difficulty, there must be a full-fledged social agreement between the government, savers (who hold government bonds), public sector wage earners and recipients of food subsidies. To fix the inflated outlays on debt service, wage bill and subsidies, some major redistribution of income must accompany the inflation-halting step. This is in essence the lesson from Israel's inflation stabilization policy.

6 A Balance-of-Payment Crisis

Inflation crises are often intertwined with balance-of-payment crises. Budget deficits were the root cause of the balance-of-payment-cum-inflation crisis. The high inflation period (1977–1985) comprised with prolonged balance-of-payment crisis. Large budget deficits make the inflation–employment trade-off acute, under the regime of pegged exchange rate and liberalized international capital flows, the pre-stabilization regime in Israel. In order to maintain a pegged exchange rate and liberalized capital mobility, the central bank lost its ability to control the interest rate. Both inflation and unemployment ensued. The stabilization package resulted in a regime switch; the government effectively shifted the regime from the first goal of the trilemma to the second goal, while being able to sharply reduce budget deficits. Balance-of-payment crises occur when a country lifts restrictions on capital mobility (in Israel it begun in 1977) without the consolidation of its fiscal stance and regulatory institutions, especially those overseeing the financial intermediaries. If under these conditions the country is trying also to maintain a fixed exchange rate regime, it then unavoidably faces conflicting policy needs (such as fiscal imbalances or a fragile financial sector) that need to be resolved by independent monetary policy.

Governments try to maintain certain financial and monetary arrangements, most notably a fixed exchange rate regime. Their goal is to stabilize the economy. At times,

these arrangements become unstable and collapse leading to financial crises. This strand of the literature analyzes currency crises characterized by a speculative attack on a fixed exchange rate regime.

Currency crises occur when the country is trying to maintain a fixed exchange rate regime with capital mobility, but faces conflicting policy needs, such as fiscal imbalances or fragile financial sector, that need to be resolved by independent monetary policy and effectively shift the regime from the first solution of the trilemma to the second solution.[14]

Krugman (1979) describes a government attempting to maintain a fixed exchange rate regime. But, it is subjected to a constant loss of reserves, due to the need to monetize persistent government budget deficits. These two features of the policy are inconsistent with each other and lead to an eventual attack on the international reserves of the central bank that culminate in the collapse of the fixed exchange rate regime.

In what follows, we provide a simple description of this model. Recall that the asset side of the central bank's balance sheet at time t is composed of domestic assets $B_{H,t}$ and the domestic currency value of foreign assets $S_t B_{F,t}$, where S_t denotes the exchange rate, i.e., the value of foreign currency in terms of domestic currency. The total assets have to equal the total liabilities of the central bank, which are, by definition, the monetary base, denoted by M_t.

Due to fiscal imbalances, the central bank domestic assets grow at a fixed and exogenous rate:

$$\frac{B_{H,t} - B_{H,t-1}}{B_{H,t-1}} = \mu.$$

Because of perfect capital mobility, the domestic interest rate is determined through the interest rate parity, as follows:

$$1 + i_t = \left(1 + i_t^*\right) \frac{S_{t+1}}{S_t}$$

where i_t denotes the domestic interest rate at time t and i_t^* denotes the foreign interest rate at time t. Finally, the supply of money, i.e., the monetary base, has to be equal to the demand for money, which is denoted as $L(i_t)$, a decreasing function of the domestic interest rate.

The inconsistency between a fixed exchange rate regime, $S_t = S_{t+1} = \bar{S}$, with capital mobility and the fiscal imbalances comes due to the fact that domestic assets of the central bank keep growing, but total central bank assets cannot change since the monetary base is pinned down by the public at large demand for money, $L\left(i_t^*\right)$, which is anchored by the foreign interest rate. Hence, the obligation of the central bank to keep financing the fiscal needs puts a downward pressure on the domestic interest rate, which, in turn, puts an upward pressure on the exchange rate. In order

[14] See a formal exposition in Piersanti.

to prevent depreciation, the central bank has to intervene by reducing the inventory of foreign reserves. Overall, $\bar{S}B_{F,t}$ decreases by the same amount as $B_{H,t}$ increases, so the monetary base remains the same.

The problem is that this process cannot continue forever, since the reserves of foreign currency must have a lower bound. Eventually, the central bank will have to abandon the solution of the trilemma through a fixed exchange rate regime and perfect capital mobility to a solution for the trilemma through flexible exchange rate, with stabilizing monetary policy (i.e., flexible monetary base or equivalently domestic interest rate) and perfect capital mobility.

The question is this. What is the critical level of domestic assets $B_{H,T}$ and the corresponding period of time T, at which the fixed exchange rate regime collapses? As pointed out by Flood and Garber (1984), this happens when the shadow exchange rate, defined as the flexible exchange rate under the assumption that the central bank's foreign reserves reached their lower bound while the central bank keeps increasing the domestic assets to accommodate the fiscal needs, is equal to the pegged exchange rate.

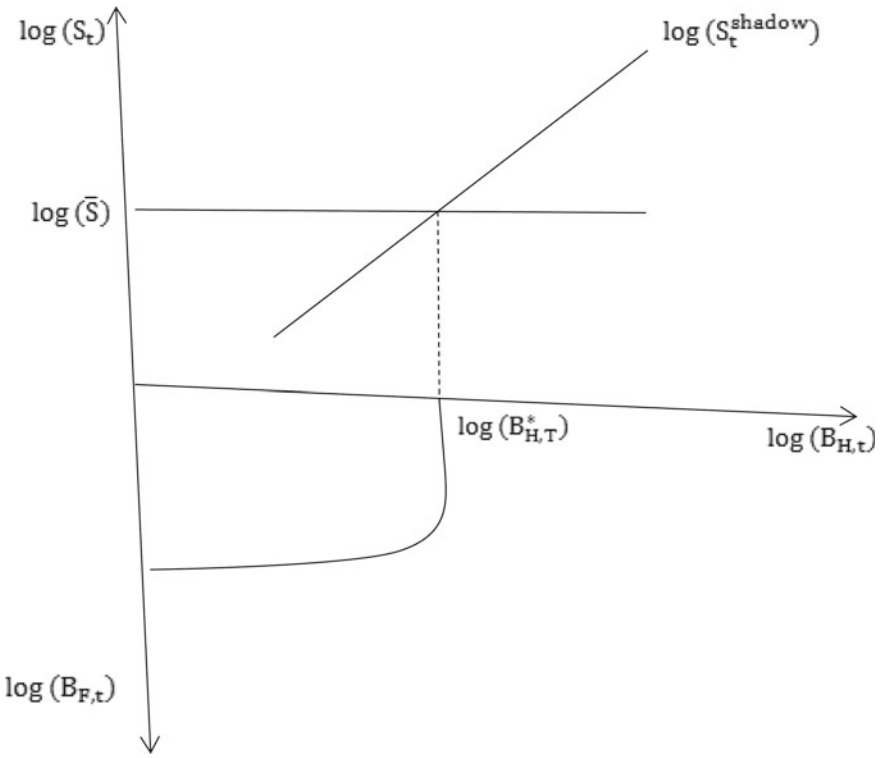

Fig. 4 Exchange rate and international reserves

Figure 4 describes the critical value of central bank domestic assets where foreign assets are suddenly depleted and a switch to fully flexible exchange rate regime occurs.

Some of the macroeconomic institutional changes, brought about by the inflation stabilization, have lasted until these very days. The hyperinflation *cum* financial collapse episode has not reoccurred. Thanks to more disciplined monetary and fiscal policies, and well-regulated banks, the inflation rate converged to low rates, enjoyed by the advanced economies during the Great Moderation era.

In contrast, inflation stabilization programs adopted by other developing countries, especially in Latin America, proved not to have similar long-term durability. Argentine's stabilization program, relying on a rigid currency board setup as its major pillar, was different. A lack of adequate budget discipline and importantly inadequate bank regulations were some of the major weaknesses of the program. With a sovereign debt crisis and international capital flow reversal, "all hell broke loose." The abruptly collapsed currency board and the run on the banks created a severe liquidity shortage. Sovereign debt default ensued. The world had cut Argentina from the international capital market. More than 10 years later, prices are not stabile. The country was able only recently to have better access to the international capital markets. Chile's stabilization program, however, had long-lasting outcomes, similarly to the Israeli program.

In contrast to the crisis management experience in Latin America, the Asian crisis has been a game-changing event that put the Asian economies (particularly South Korea and Indonesia) on a durable growth track. To a large measure, the post-crisis Asian financial and monetary institutions restructuring enabled the entire region to escape the 2008 global crisis.

7 Disinflation and Globalization

The globalization wave has swept emerging markets in Latin America, European transition economies, East Asian emerging economies and Israel, over the last decades. The 1992 single-market reform in Europe and the formation of the eurozone were watersheds of globalization. Emerging markets, including China and India, likewise became significantly more open. Wynne and Erasmus (2007) note that in the 1970s more than three quarters of industrial countries had restrictions of some sort on international financial transactions. By the 2000s, none did. Likewise, restrictions on these transactions among emerging markets fell from 78% in the 1970s to 58% in the 2000s. Israel was exposed intensively in the globalization forces and was able to exploit them to climb down from three-digit inflation rates in the early 1980s and the double-digit rates in the late 1980s and early 1990s.

The "Great Moderation" refers to the significant business cycle volatility starting in the mid-1980s, believed at that time to be permanent, in developed nations in the later part of the twentieth century. Sometimes during the mid-1980s, major economic variables such as real GDP growth, industrial production, monthly payroll and the

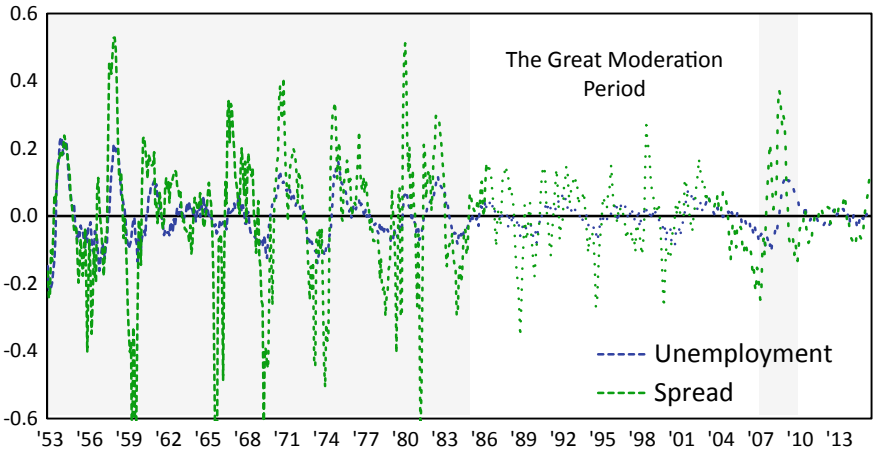

Fig. 5 HP filter de-trended unemployment rate and 5-year bond-yield (corporate/treasury) spread, USA, 1953–2014. *Source* FRED, BLS, FRED, BLS, an extension to Eckstein et al. *Notes* De-trended unemployment rate obtains through HP filter, in SD. Bond-yield spread is defined as the difference between two things: 5-Year Treasury Constant Maturity Rate and Moody's Seasoned Baa Corporate Bond Yield, HP-filtered, in SD

unemployment rate began to decline in volatility. These reductions are primarily due to greater independence of the central banks from political and financial influences which has allowed them to follow macroeconomic stabilization.

Figure 5 describes deviations from trend of the unemployment rate and the bond-yield corporate–treasury spread, for the period 1953–2014. The figure highlights the significant reduction in the fluctuations of the unemployment rate and the bond-yield spread between Baa corporation rates and the treasury rate.

Global inflation declined from 30 to 4% between 1993 and 2003.[15] Rogoff (2003, 2004) conjectures that globalization—interacting with deregulation and privatization—has played a strong supporting role in the past decade's disinflation. An important feature of openness relates to international labor flows. International migrants constituted 2.9% of the world population in the 2000s, up from 2.1% in 1975. In some countries, changes have been more dramatic. In Israel in the 1990s, there was a surge of immigrants of up to 17% of the population, and the central bank achieved a sizable decline of inflation. It is possible that the two events are related. In Spain in 1995, the percentages of foreigners in the population and in the labor force were below 1% and below 0.5%. At the end of 2006, these rates were around 9 and 14%.

[15] Kenneth Rogoff's paper was prepared for the Federal Reserve Bank of Kansas City Conference on "Monetary Policy and Uncertainty: Adapting to a Changing Economy," Jackson Hole, WY, August 29, 2003.

By easing labor bottlenecks, migrants help to keep down prices of goods and services. Pass-through of world's low inflation, and low interest rates, to domestic prices and interest rates, the effects of migration on wages, is to be addressed by the standard Phillips curve analysis.[16]

8 The Flattening of the Phillips Curve

The core mechanism in the New Keynesian paradigm depends on the Phillips curve, that is, the trade-off between surprise inflation and the level of economic activity. The reason why the New Keynesian framework is capable of generating such a trade-off between inflation and economic activity is that producer-desired prices (once prices are adjusted) rise with the economy's output, when marginal costs slope upward due to diminishing returns to scale. Furthermore, when the labor supply increases, workers tend to experience increasing marginal disutility of labor efforts. The resulting increased real wage demands must rise. Increased wage demands put an upward pressure on the marginal cost and consequently on the producer-desired price setting.

Bean (2006) succinctly summarizes the effect of globalization on the Phillips curve in the era of globalization:

> One of the most notable developments of the past decade (that is, the 1990s) has been the apparent flattening of the short-run trade-off between inflation and activity. The seventies were characterized by an almost vertical relationship in the United Kingdom, in which attempt to hold unemployment below its natural rate resulted in rising inflation. In the eighties, the downward sloping relationship reappears, as inflation was squeezed out of the system by the slack of the economy. However, since the early nineties, the relationship looks to have been rather flat. Three factors—increased specialization; the intensification of product market competition; and the impact of that intensified competition and migration on the behavior of wages—should all work to flatten the short-run trade-off between inflation and domestic activity."[17]

Independence of central banks is a way to overcome dynamic inconsistency: Expected inflation leads to output, employment and financial market distortions; surprise inflation is employment and output boosting (through the Phillips curve mechanism). In the absence of central bank independence, the non-commitment equilibrium is one of the high-expected inflations. Central bank independence is a necessary condition for overcoming the dynamic inconsistency and consequently weakening

[16] Bentolila et al. (2007) have addressed the impact of the Spanish immigration boom on the Phillips curve.

[17] Similarly, Mishkin (2007a, b) writes about the U.S. inflation-output trade-off: "The finding that inflation is less responsive to the unemployment gap, suggests that fluctuations in resource utilization will have smaller implications for inflation than used to be the case. From the point of view of policy makers, this development is a two-edged sword: On the plus side, it implies that an overheating economy will tend to generate a smaller increase in inflation. On the negative side, however, a flatter Phillips curve also implies that a given increase in inflation will be more costly to wring out of the system.

the inflation bias. Accordingly, Rogoff (2003, 2004) attributes the moderation in world inflation to a broad-based move toward having them run by conservative anti-inflation-oriented central bankers; similar developments happened also in Israel. The increased competitiveness was a result of the interplay of globalization, deregulation and a decreased role for governments in many economies. Given this diagnosis, he foresaw continued disinflation and even deflationary pressures (which came into a stark relief in the Great Recession) arguing that the most important factor supporting worldwide disinflation has been the mutually reinforcing mix of goods market and financial deregulation and globalization, and the consequent significant reduction in monopoly pricing power. These developments increased competitiveness; diminishing the gains, a central bank can reap via unanticipated inflation, because it reduces the gap between the economy's monopolistically competitive equilibrium and the more socially desirable competitive equilibrium. In addition, both theory and empirics suggest that economies that are more competitive have more flexible nominal prices, making that smaller the Barro–Gordon-type output gain the central bank can achieve by inflating and making them more ephemeral. In a standard, stylized political–economy model, Rogoff shows that it is easier to credibly sustain low inflation in a competitive than in a highly monopolistic economy.[18]

Evidence of the effect of globalization on the Phillips curve is provided by Loungani et al. (2001), Loungani and Razin (2001), and Clarida (2008). Previously, Romer (1993, 1998) and Lane showed that inflation and trade liberalization are negatively (significantly) correlated among the large (flexible exchange rate) OECD economies.

Evidently, changes in the foreign price pass through into domestic inflation in the open-economy case even if the exchange rate depreciation trend does not change. If, in addition, the exchange rate depreciation tapers down, and once the foreign exporters to the home country are also given a chance to adjust prices in response to the moderation in the exchange rate depreciation, the home country import price inflation moderates as well. In the world of the Great Moderation, the home country inflation abates.

Opening up of the economy to capital, goods and labor mobility also flattens the Phillips curve. In the New Keynesian framework, Binyamini and Razin (2008) show how increased volume of trade in goods, greater financial openness and labor migration affects the trade-off between output and inflation by flattening the Phillips curve.[19] Minimizing the (utility-based) loss function implies moderate inflation, akin to the Great Moderation. They demonstrate analytically how the opening up of the domestic economy to trade in goods, international borrowing and lending and migration flattens the Phillips curve (see Appendix). Every successive round of the opening up of the economy contributes to flatten the aggregate supply curve. The intuition is

[18]Rogoff's prediction has proven correct. Global inflation moved sideways also after 2003 and then fell sharply asymptotically, approaching zero after 2008, despite massive monetary and credit expansion in the USA and the European Union.

[19]See Appendix.

that when an economy opens up to trade in goods, it tends to specialize in production but to diversify in consumption. This means the number of domestically produced goods is less than the number of domestically consumed goods. Consequently, the commodity composition of the consumption and output baskets, which are identical if the trade account is closed, are different when trade in goods is possible. As a result, the correlation between fluctuations in output and in consumption (which is equal to unity in the case of a closed trade account) is less than unity if the economy opens up to international trade in goods. The decomposition of utility-based Phillips curve to the various forces of migration, output gap and real exchange rate is shown in an Appendix. In words, these globalization forces work analytically as follows.

When the capital account is open, then the correlation between fluctuations in consumption and domestic output is further weakened, and this is because with open capital accounts the representative household can smooth consumption through international borrowing and lending and thereby separate current consumption from current output. The inflation effects of shocks to the marginal cost are therefore reduced, because the fluctuations in labor supply are also smoothed, because of the consumption smoothing.

When the labor market is internationally closed to outward migration, wage demands faced by domestic producers are upward sloping, both under in-migration and under a completely closed labor market. However, when the labor market is open to in-migration, domestic producers face an expanded labor supply: Additional to the skilled native-born labor supply (with upward sloping wage demand), they also face a complementary unskilled foreign labor supply (with exogenously determined wage demand). That means that in-migration acts on the Phillips curve essentially like a domestic productivity shock.

There has been some evidence of greater restraints on domestic prices and wage growth in sectors more exposed to international competition, such as textiles and electronics. Chen et al. (2004) analyzed disaggregated data for EU manufacturing over the period 1988–2000. They find that increased openness lowers prices by reducing markups and by raising productivity. This finding implies a downward shift of the Philips curve. In response to an increase in openness, markups show a steep short-run decline, which partly reverses later, while productivity rises in a manner that increases over time. If globalization reduces the markup, our model predicts that this effect, by itself, leads to a more forceful anti-inflation policy and lessens the attention given by the policymaker to the fluctuations in economic activity. One can conjecture that more frequent price updating steepens the trade-off between inflation and activity; however, to our knowledge, neither theory nor empirical evidence exists in support of any systematic relationship between globalization and frequency of price updating. Notably, Gopinath and Rigobon report that the time frequency of price adjustment of US imported goods trended downward, on average, during the Great Moderation. Gopinath and Itskhoky exploit the open-economy environment, which provides a well-identified cost shock, namely sizeable exchange rate shocks. They use this identification method to test the effects of price-adjustment frequencies

and pass-through. They demonstrate that high-frequency adjusters have a long-run pass-through that is at least twice as high as low-frequency adjusters in the data are. Borio and Filardo (2007) present cross-country evidence in support of their contention that global factors have recently become empirically more relevant to domestic inflation determination.[20]

9 Convergence of Inflation Rates

Globalization—interacting with deregulation and privatization—has played a strong supporting role in Israel's disinflation. The moderation is due to a large extent to the increasing independence of the Bank of Israel, conducting effective anti-inflation policies in the presence of worldwide disinflation.[21]

Figure 6 shows the convergence of Israel inflation rate to USA, Germany and OECD rates. Inflation fall started after the 1985 inflation stabilization policy but converged to the low one-digit rates of advanced economies in the 1990s.[22]

Full international financial integration requires that in the long run (when prices adjust to various shocks and markets clear), the following arbitrage equation holds.

$$1 + r_t^{US} = \left(1 + r_t^i\right) \frac{q_{i/US,t+1}}{q_{i/US,t}},$$

where i stands for Israel, Canada, Germany and the UK and q stands for the real exchange rate *vis-a-vis* the US dollar[23]:

---

[20] See Binyamini and Razin (2008). See also Gali (2008) for a comprehensive treatise of the open-economy New Keynesian model. Borio and Filardo (2007) present cross-country evidence in support of their contention that global factors have recently become empirically more relevant to domestic inflation determination. But Ihrig et al. have shown that their result is very specific to the econometric method used. Based on cross-country analysis, Badinger (2007) finds that globalization is also correlated with more aggressive policy toward inflation. Tetlow and Ironside (2007), although not dealing with globalization, find that for the USA, the slope of the Phillips curve has—largely and continuously—lessened during recent years. However, Ihrig et al. have shown that results are very specific to the econometric method used.

[21] Globalization affected also conduct of central banks. Inflation targeting was born in New Zealand in 1990. Admired for its transparency and accountability, it achieved success there and soon in Canada, Australia, the UK, Sweden and Israel. It subsequently became popular as well in Latin America (Brazil, Chile, Mexico, Colombia and Peru) and in other developing countries (South Africa, South Korea, Indonesia, Thailand and Turkey, among others).

[22] Leiderman (1999) comprehensively analyzes Israel's disinflation with a focus on monetary policies related to inflation and disinflation in Israel. He especially focuses on inflation targeting as an instrument of disinflation.

[23] Recall that by the Fisher equation:
$1 + r_t^{US} = \left(1 + i_{US}^t\right) \frac{P_{US,t}}{P_{US,t+1}}$, that is, $\left(1 + r_t^i\right) \frac{q_{i/US,t+1}}{q_{i/US,t}} = \left(1 + i_i^t\right) \frac{P_{i,t}}{P_{i,t+1}} \frac{q_{i/US,t+1}}{q_{i/US,t}}$.

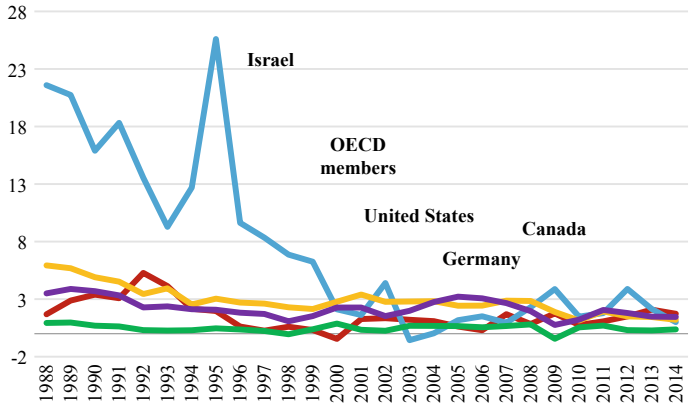

Fig. 6 Inflation rates (annual GDP deflator, percentage). *Source* The World Bank

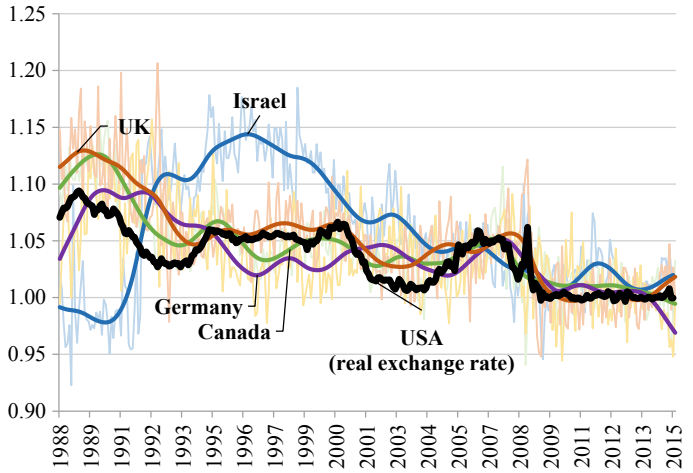

Fig. 7 Gross real interest rate adjusted for real exchange rate changes (US benchmark). *Note* Series are HP-filtered. Monthly data are shown in the background. *Source* Stats Bureau, FERD, World Bank, real exchange rate adjusted, yields on three-month government bonds for Israel, Canada, Germany and the UK, and the yields on three-month US government bonds

$$q^t_{i/US,t} = E_{i/US,t} \frac{P_{US,t}}{P_{i,t}},$$

In addition, E stands for the nominal exchange rate, vis-a-vis the US dollar, and P stands for the price level.

Figure 7 plots the graphs of the real interest rate, adjusted for real exchange rate changes, the yields on three-month government bonds for Israel, Canada, Germany

and the UK, and the yields on three-month US government bonds. International financial integration generates more synchronized country-specific yields. Time series are filtered to wash out short-run idiosyncratic fluctuations. Figure 7 demonstrates strikingly that in the 1990s Israel integrates sufficiently into the world capital market, while convergence occurred at the beginning of the 2000s. demonstrates strikingly that in the 1990s Israel integrate significantly into the world financial markets.

The absence of constraining rules on actions of the Bank of Israel and on Israel's fiscal authority has induced strongly accommodative monetary policies and uncontrolled inflation. With improper financial sector regulation (e.g., the so-called *Visut Menayot*), banks were on the verge of collapsing in the 1984 crisis. They were able to recapitalize making their investment portfolios less risky over the next two decades, thanks to more rigorous bank regulations.

10 Depression–Deflation Resistance

Israel's resilience to the external financial shock during the global crisis is rooted in (a) the absence of credit boom in the wake of the crisis and (b) the relatively small commercial banks' exposure in terms of toxic assets that for the European countries played a major role. For analysis of financial crisis see Diamond and Dybvig (1983), De Grauwe (2011).

The newly emerging macroeconomic paradigm spans the gamut from an analytical framework that features full capital market arbitrage, smooth credit, Ricardian-equivalence properties, representative agents and efficient monetary management, to a framework with multiple agents, incorporating debt frictions, liquidity traps and relatively ineffective monetary management, and provides a role for fiscal policy in aggregate demand management. The analytical framework based on the friction-less paradigm captures well the role of globalization forces and the reduction in inflation in the 1990s Great Moderation era. The multiple-agent, market-friction-revised analytical framework captures some key features of the Great Recession that occurred in the aftermath of the 2008 global financial crisis. It gives insight about the macroeconomic effects of debt overhang on economic activity and inflation, when the monetary policy rate reaches its lower bound.

The concern at the time was that Israel, being well integrated into the world markets and the world finance, might suffer contagion that will be long lasting. At the end of the day, Israel suffered only a temporary trade shock because of the decline in world demand.

As shown in Fig. 8, Israel did not have a significant credit boom in the wake of the 2008 crisis. The USA and the UK, in contrast, were vulnerable to a gigantic credit expansion (Germany, as if Israel escaped such credit bubbles).

Nevertheless, GDP growth has averaged 4% over the 2005–2010 period years, compared with 0.7% on average for OECD countries. The overall living standards continue to improve gradually, with per capita real GDP growing more rapidly than in other OECD countries. The economy's resilience has been underpinned by solid

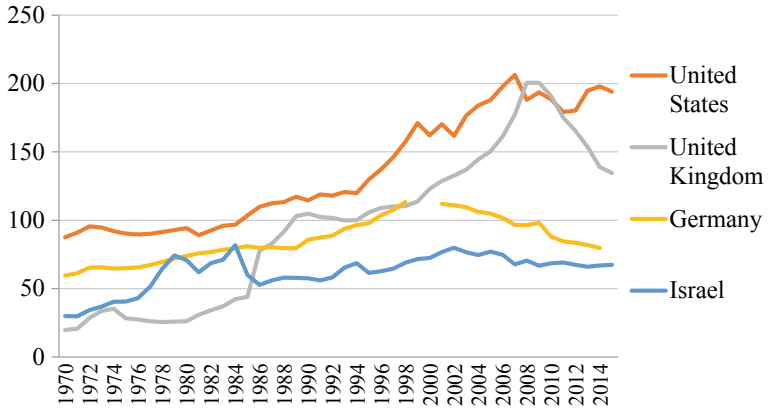

Fig. 8 Domestic credit to private non-financial sector (% of GDP). *Source* International Monetary Fund, International Financial Statistics, data files, and World Bank and OECD GDP estimates. *Note* Domestic credit provided by the financial sector includes all credit to various sectors on a gross basis, with the exception of credit to the central government, which is net. The financial sector includes monetary authorities and deposits money banks, as well as other financial corporations where data are available (including corporations that do not accept transferable deposits but do incur liabilities such as time and savings deposits). Examples of other financial corporations are finance and leasing companies, moneylenders, insurance corporations, pension funds and foreign exchange companies

economic fundamentals, including large foreign reserves, a dynamic high-tech export sector and the absence of economy-wide deleveraging pressures leading to the downfall in economic activity. Because, Israel did not have a credit bubble in the years preceding the global financial crash, like the other major advanced economies, which burst during the financial crisis.

Israel's growth performance depicted in Fig. 9, during and after the global crisis, however, was not unique. Figure 3.3 shows that among similar small open economies

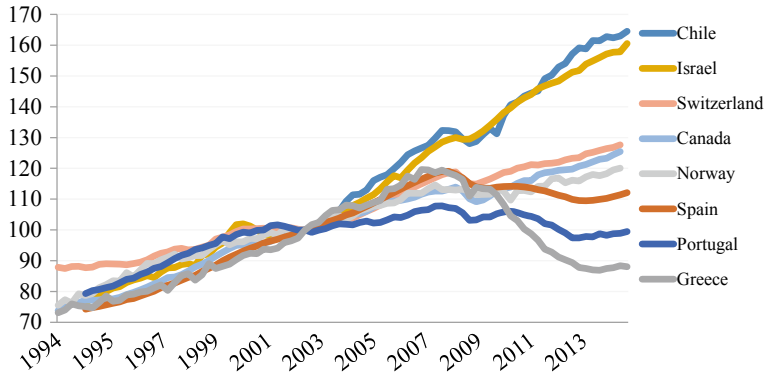

Fig. 9 Real GDP, Israel and selected countries (January 2003 = 100). *Source* FRED

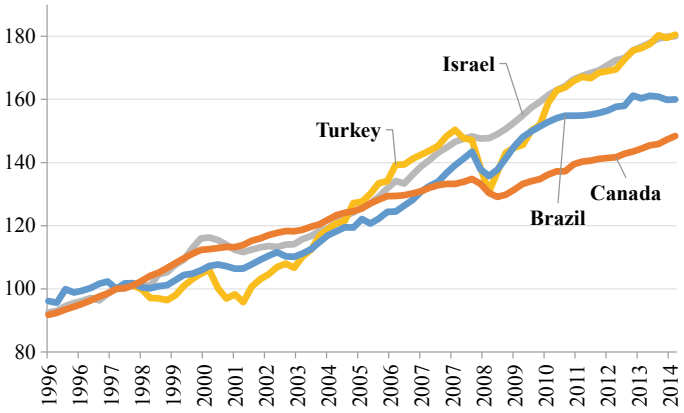

Fig. 10 Real GNP, countries with no financial crisis (January 1998 = 100). *Source* FRED

Israel's GDP grew over recent 20 years, including the 2008–2010 period at a similar cumulative rate as Chile, but at a much higher rate than Greece, Spain and Portugal;, which had a financial sector crash.

Furthermore, Fig. 10 depicts GNP levels for Israel, Turkey, Brazil and Canada, economies, which spared financial, sector crash. Israel exhibits a more moderate drop of output than all these countries.

Capital flows provide another measure of the resilience of the Israeli economy to the shocks. In the aftermath of the global financial crisis, expansionary monetary policy in advanced economies conventional or unconventional that were conducted to boost up the economy has affected emerging market economies and others, such as Israel, through four channels: capital inflows; exchange rate appreciation; reduced exports; and effects of capital inflows on the domestic financial system.

A number of studies have found an effect of monetary policy on specific gross flows. Bruno and Shin (2015), for example, using a VAR methodology over the pre-crisis period (1995:4 to 2007:4), find an effect of the federal fund rate on cross-border bank to bank flows; the effect is however barely significant. Fratzscher et al. (2013), using daily data on portfolio equity and bond flows, find significant effects of different monetary policy announcements and actions since the beginning of the crisis. Their results however point to the complexity of the effects of apparently largely similar monetary measures. For example, they find QE1 announcements decreased bond flows to EMs, while QE2 announcements increased them. In terms of the equations above, this suggests that, in each case, monetary policy worked partly through its effects on the risk premium. These studies cannot settle the further issue of whether or not total gross inflows increase with advanced economy monetary expansions: The increase in the inflows the researchers have identified may be offset by a decrease in other inflows[24]. However, studies of total inflows, or of the set of inflows adding

[24]See Blanchard (2016) who surveyed the literature about post-2008 crisis in advanced economies and emerging economies that were hit to different degrees by the global financial crisis. To a large

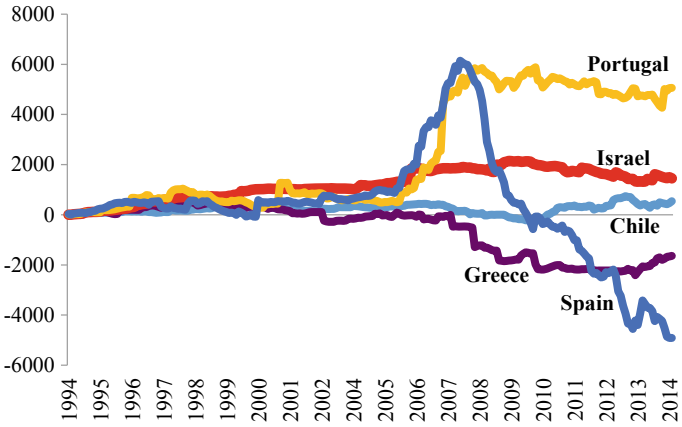

Fig. 11 Portfolio flows, crisis economies (index, December 1994 = 100). *Source* Anusha Chari

up to total inflows, yield some mixed conclusions. A representative and careful paper, by Cerutti et al. (2015), using quarterly flows over 2001:2 to 2013:2, suggests two main conclusions. The most significant observable variable in explaining flows into emerging markets (EMs) is the VIX index[25]: An increase in the VIX leads to a decrease in inflows to EMs. The coefficients on the monetary policy variables, namely the expected change in the policy rate and the slope of the yield curve, typically have the expected sign. Several studies found that movements in the VIX are strongly associated with global capital flows.[26]

It is worth looking now at capital inflows to EMs and Israel from the USA, the epicenter of the global financial crisis, and the country, which adapted with virtually no lag a brief expansionary fiscal policy and a persistent expansionary monetary policy.

Figure 11 describes the portfolio capital outflows from the USA to selected countries. Israel is in the middle of the pack of countries that enjoy inflow of portfolio capital investments in the aftermath of the 2008 global financial crisis. These inflows put appreciation pressures on the exchange rates. Some central banks, including BOI, conducted a policy of a massive purchase of foreign currency-denominated assets, to protect against the declining competitiveness in the world trade.

Figure 11 describes the nominal exchange rate of various countries that engaged in the "currency war" period: Israel, Sweden, Switzerland, Brazil and Indonesia. Israel seems to have undervalued its currency the least among these countries; possibly because of its international financial integration and almost no capital controls, the effectiveness of sterilized intervention was weak.

---

extent, the emerging markets escaped the brunt of the crisis. Israel evidently belongs to the second group.

[25]The VIX is the Chicago Board Options Exchange Market Volatility Index. It is a measure of the implied volatility of S7P 500 index options, the VIX.

[26]See Rey (2015).

How did the Israeli policymakers react to the 2008 world financial depression and global trade-diminishing shocks? Policymakers' concern was threefold: first, bank exposures to toxic assets such as mortgage-based securities and foreigners' debt obligations. Partly because Israel skipped the credit bubble, and bank regulations were relatively tight, Israel showed a sound resilience to the global financial shock. Second, Israel export markets softened and demand conditions deteriorated. Third, Israel domestic currency was strengthened. Bank of Israel addressed the last two issues by a massive foreign exchange market intervention to weaken the value of the domestic currency and stimulate exports.

In the aftermath of the global financial crisis, expansionary monetary policy in advanced economies' conventional or unconventional that were conducted to boost up the economy has appreciated the currencies of the emerging market economies, including Israel. The question for these economies was whether an expansionary monetary policy, which tends to depreciate the currency and boost exports, requires a direct foreign exchange market intervention or whether the latter can succeed without the former.[27]

Israel monetary authorities were concerned about the "Great Recession" downward pressures on the demand for Israel's exports and the strengthening of the Israeli currency as capital inflows rose. They engaged in an intensive (sterilized) intervention in the foreign exchange market to prevent the appreciation of the currency. However, there are evident limits to how much such policy can stimulate the demand for Israel's output.

Sterilized intervention is ineffective when there is high private capital mobility to the extent that domestic and foreign securities viewed by a large group of investors are close substitutes. Conditions under which sterilized intervention is effective happen to exist for a crisis economy, however, when there is a probability of capital flow reversal, liquidity shortage or major real trade shock, leading to financial intermediary collapse. Under conditions where foreign and domestic assets are close substitutes, sterilized intervention is ineffective. Through a central bank sale of domestic government debt assets, following a purchase of foreign currency in the foreign exchange market, the money supply fully adjusts to bring back the pre-intervention expected rates of return on domestic and foreign currency bonds into equilibrium (the standard interest parity). Sterilized foreign exchange market intervention, by the monetary authorities, where the domestic money supply is unchanged, is incapable of pushing the exchange rate up or down. However, the proposition may change in the presence of imperfect asset substitutability, where domestic and foreign bonds command a different liquidity premium and risk premium. Changing the composition of central bank assets, between foreign and domestic assets (the case of sterilized foreign

[27] In the presence of trending capital exports, the central bank cannot persistently appreciate the domestic currency by selling foreign exchange-denominated assets in the foreign exchange market, because depleted international reserves could quickly reach their lower bound. However, in the presence of trending financial capital imports, the central bank can persistently depreciate the domestic currency by purchasing foreign exchange-denominated assets in the foreign exchange market through money issue. This is why the foreign exchange market intervention is referred to as a "half instrument" in the hands of the central bank.

exchange rate market interventions), can then have real economic effects in the presence of credit market spreads and frictions. In this case, sterilized foreign exchange market intervention could effectively change the value of the foreign currency in terms of domestic currency. A sterilized purchase of foreign assets may change the liquidity premium that domestic bonds command, relative to foreign bonds, even though the money supply is left unchanged. A similar outcome may transpire when foreign exchange intervention changes market views of future foreign exchange market interventions. Similarly, liquidity-based imperfect asset substitution between domestic government and domestic private-sector bonds during liquidity crises can be exploited by the central bank.[28] Israel's foreign exchange market half-decade episode started when credit frictions were relatively intensive following the Lehman moment in the USA in October 2008.

Recall that the most significant observable variable in explaining short-term flows into Emerging Markets (EMs) is the VIX index: An increase in the VIX leads to a decrease in inflows to EMs. The VIX index is directly related to the risk-adjusted return on domestic government bonds in the periphery countries, like Israel. Sterilized foreign exchange market purchase of US government bonds by the central bank is then capable of blocking exchange rate appreciation. This was the rationale for the Bank of Israel policy in the aftermath of the global financial crisis. However, the effectiveness of such policy is short lived. Once the VIX index falls, sterilized foreign exchange market intervention becomes ineffective. Excessively high foreign reserves also have fiscal medium-term costs.[29]

11 Conclusion

Historical patterns of booms and busts typically exhibit frequent small recessions interrupted by rare but deep and long recessions. Traditional macroeconomic models, used often by central banks and many other policy-making institutions, do not capture the full features of crises: frequent small recessions punctuated by rare depressions. They do not illuminate how small open economies, like Israel, which are substantially integrated into the world economy, perform when a global financial shock takes place, leading to recession as deep and persistent as the Great Recession. We discussed the relatively robust performance of Israel (as well as some other advanced economies

[28] See Krugman et al. (2015).

[29] Sorezcky (2015) provides evidence on the effectiveness of the 2008–2009 strong intervention period, and Ribon (2017) provides a broader overview of forex intervention over the entire global financial crisis. Cukierman (forthcoming) compares the methods of interventions in Israel and Switzerland and their implications for forex reserve accumulation. Another reason for the tameness of inflation despite massive base expansion in the USA since 2008 is that this expansion was in response to a huge increase in the demand for liquidity on the part of banks and the entire financial sector rather than to a governmental craving for seignorage revenues. Cukierman (2017) demonstrates and argues that this is very different than the base expansion in post-WWI Germany in which the main motive for base expansion was seignorage for government.

(e.g., Canada) and major Emerging Markets in the aftermath of the 2008 global financial crisis. Factors contributing to this robustness are the absence of credit and real estate bubbles, and banks' tight regulation in the wake of the crisis, which precluded the deleveraging process following the financial crisis.

The state of Israel, founded in 1948, benefited immensely from the post-1945 globalization wave. Steadily reforming its financial and commercial institutions, and becoming increasingly globalized in trade, labor market and finance, Israel became a member of the OECD; the accession took place in 2010. Currently, Israel's is a thriving economy, integrated tightly into the world economy, and Israel also features a remarkable technological prowess. The Israeli economy is a remarkable development success story. A middle-income economy in the midst of a hyperinflation in the early 1980s, Israel grew into one of the most thriving economies in the world: this despite the ongoing security challenges that are most certainly a drain on its resources.

Fiscal policy has been recently given greater emphasis in the post-2008 crisis that nominal interest rates appear to be persistently low—and below the annual growth rate of nominal GDP. This recent phenomenon generated renewed debate on the role of "printing money" in financing government deficits. Modern Monetary Theory, or MMT, argues that a country borrowing in its own currency can finance fiscal stimulus by printing money. That is, governments able to issue fiat money cannot go bankrupt, regardless of whether investors are willing to buy their bonds. By extension, MMT would allow the government to control inflation through tax policy. Instead of asking the Fed to stabilize prices through monetary policy, the government could raise taxes when prices get too high and cut taxes when prices get too low. However, basic macroeconomics wisdom suggests that deficit finance by money issue will not leave banks sitting idle on their newly acquired reserves; they will convert them into currency, which they lend to individuals. So, the government indeed ends up financing itself by printing money, getting the private sector to accept pieces of paper in return for goods and services, and this would lead to inflation. Recall that a deficit financed by money issue is more inflationary than a deficit financed by bond issue. When the central bank purchases a government bond in the open market in exchange for commercial bank reserves, all it does is substitute a very short-term liability (reserves have zero maturity) for a longer-term liability. That is, a central bank purchase of government bonds simply alters the maturity structure of the consolidated government's liabilities, thereby pushing up inflationary pressures.

Recently, there has been extensive analysis of changes in the Phillips curve in Israel. Elkayam and Ilek (2016) gave evidence that between 2003 and 2013 there was a substantial reduction in the Israeli natural (or NAIRU) rate of unemployment. Such a development likely contributed to the maintenance of the two percent inflation target in spite of substantial decreases in unemployment. Another important development that most likely shifted the Israeli Phillips curve to the left is the fiscal reforms of the early twenty-first century that mandated a long-term gradual reduction in national debt and the imposition of a 3% ceiling on public deficits. As argued by Braude and Flug (2012) and others, those policies created a fiscal policy space that could be used later to moderate the deflationary impact of the GFC on the Israeli economy.

Appendix: Globalization and the Phillips Curve

Binyamini and Razin (2008) show how trade in goods, financial openness and labor in- and out-migration affect the trade-off between output and inflation by successively flattening the Phillips curve. Let the range of the mass of domestically produced goods, n, be (0, 1) and that $\omega > \omega_p$.

In the case of *perfect mobility of labor, capital and goods*, the log-linear approximate aggregate supply curve (Phillips curve) is given by[30]:

$$\widehat{\pi}_t = \kappa \cdot \left[\frac{\omega_p \cdot n}{1+\omega_p\theta} \cdot x_t + \frac{\omega_p \cdot (1-n)}{1+\omega_p\theta} \cdot \left(\widehat{Y}_t^F - \widehat{Y}_t^N \right) + \frac{1}{1+\omega_p\theta} \cdot \widehat{w}_t^W + \frac{(1-n)}{n} \cdot \widehat{q}_t \right] + \frac{(1-n)}{n} \cdot \left(\widehat{q}_t - \widehat{q}_{t-1} \right) + \beta \cdot E_t \left[\widehat{\pi}_{t+1} - \frac{(1-n)}{n} \left(\widehat{q}_{t+1} - \widehat{q}_t \right) \right],$$

where $\widehat{\pi}_t$ is the deviation of CPI inflation from its target; $x_t \equiv \left(\widehat{Y}_t^H - \widehat{Y}_t^N \right)$ is the domestic output gap; $\left(\widehat{Y}_t^F - \widehat{Y}_t^N \right)$ is the difference between foreign output and domestic natural output; the parameter ω_p is the elasticity of the marginal cost with respect to producer's output; θ is the intra-industry elasticity of substitution; σ stands for the intertemporal elasticity of substitution; $\beta\cdot$ denotes the subjective discount factor. The term n denotes the mass (number) of domestically produced goods, w is domestic wage, and superscripts F, N and W denote foreign, natural and world variable, respectively.

The term $\kappa = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}$ captures the degree of price flexibility; $(1-\alpha)$ is the probability of receiving a price-updating signal. The variable $\widehat{q}_t$ is the real exchange rate, formally defined as:

$$\widehat{q}_t = \widehat{\varepsilon}_t + \widehat{P}_{F,t} - \widehat{P}_t,$$

where $\widehat{P}_{F,t}$ denotes the foreign consumer price index.

Denote the slope of the Phillips equation by ψ; for the open-economy expression slope of the aggregate supply, equation is $\psi_1 \equiv \frac{\kappa n \omega_p}{1+\omega_p\theta}$.

Let us turn to the case of no labor mobility and no capital mobility. If the domestic economy is not integrated to the international financial market, then there is no possibility of consumption smoothing, and we have that the value of aggregate current spending equals the value of aggregate domestic output:

$$\hat{P}_{C,t}\hat{C}_t = \hat{P}_{Y,t}\hat{Y}_t \quad ; \quad \hat{P}_{C,t}\hat{C}_t^N = \hat{P}_{Y,t}\hat{Y}_t^N,$$

[30] See also Razin (2014), Chap. 11.

where $\hat{P}_{C,t}$ the CPI-based price is level and $\hat{P}_{Y,t}$ is the GDP deflator. In this case, the aggregate supply curve is

$$\widehat{\pi}_t = \kappa \cdot \left[\frac{(\omega \cdot n + \sigma)}{1 + \omega\theta} \cdot x_t + \frac{\omega \cdot (1-n)}{1 + \omega\theta} \cdot \left(\widehat{Y}_t^{\mathrm{F}} - \widehat{Y}_t^{\mathrm{N}} \right) + \frac{(1-n)}{n} \cdot \widehat{q}_t \right] + \frac{(1-n)}{n} \cdot \left(\widehat{q}_t - \widehat{q}_{t-1} \right) + \beta \cdot E_t \left[\widehat{\pi}_{t+1} - \frac{(1-n)}{n} \left(\widehat{q}_{t+1} - \widehat{q}_t \right) \right].$$

The Phillips curve slope is:

$$\psi 2 \equiv \frac{\kappa(\omega n + \sigma)}{1 + \omega\theta}.$$

In the *closed economy* case, the aggregate supply equation (Phillips curve) reduces to

$$\widehat{\pi}_t = \frac{\kappa}{1 + \omega\theta} \cdot (\omega + \sigma) \cdot x_t + \beta E_t \widehat{\pi}_{t+1},$$

In the case of the closed economy, the Phillips curve slope is:

$$\psi_3 \equiv \frac{\kappa(\omega + \sigma)}{1 + \omega\theta} \geq \psi 2 \equiv \frac{\kappa(\omega n + \sigma)}{1 + \omega\theta} \geq \psi_1 \equiv \frac{\kappa n \omega_p}{1 + \omega_p \theta}.$$

The Phillips curve is steeper in the closed economy case, compared to the open-trade case with no labor and no capital. The latter is steeper than the slope with perfect mobility of labor, capital and goods. The model features a moderating impact of in-migration on wages.

Furthermore, changes in the foreign price pass through into domestic inflation in the open-economy case, but these effects are absent in the closed economy case. This observation validates the proposition the globalization in the world of great moderation exert inflation moderating influences.

References

Badinger, H. (2007). *"Globalization, Taylor rules, and inflation," mimeo*. Vienna, Austria: Wirtschaftsuniversität Wien.

Bean, C. (2006). Globalization and inflation. *Bank of England Quarterly Bulletin, Q4,* 468–475.

Bentolila, S., Dolado, J., & Jimeno, J. (2007). Does immigration affect the Phillips curve? Some evidence for Spain. *European Economic Review, 52*(8), 1398–1423.

Binyamini, A., & Razin, A. (2008). Inflation-output tradeoffs as equilibrium outcomes of globalization. *Israel Economic Review, 6*(1), 109–134.

Blanchard, O. (2016). Currency wars, coordination, and capital controls". NBER working paper 22388.

Borio, C., & Andrew, F. (2007). *Globalization and inflation: New cross-country evidence on the global determinants of domestic inflation* (Unpublished paper) Bank for International Settlements, Basel, Switzerland (March).
Braude, K., & Flug, K. (2012). The interaction between monetary and fiscal policy: Insights from two business cycles in Israel.
Bruno, V., & Shin, H. S. (2015). Capital flows and the risk-taking channel of monetary policy. *Journal of Monetary Economics, 71*, 119–132.
Bruno, M., & Fischer, S. (1984). *The inflationary process in Israel: Shocks and accommodation*. MA: National Bureau of Economic Research Cambridge.
Calvo, G. (1978). On the time consistency of optimal policy in a monetary economy. *Econometrica, 46*, 1411–1428.
Calvo, G. (1992). Are high interest rates effective for stopping high inflation? Some skeptical notes. *The World Bank Economic Review, 6,* 55–69.
Calvo, G. A., & Carlos A. V. (2001). Inflation stabilization and BOP crises in developing countries. In J. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* Elsevier.
Calvo, G. (2016). From chronic inflation to chronic deflation: Focusing on expectations and liquidity disarray since WWWII. NBER working paper 22535.
Cerutti, E., Claessens, S., Laeven, L., (2015a). The use and effectiveness of macro prudential policies: New evidence. IMF Working Papers 15/61, International Monetary Fund. 1, 2.1.
Chen, N., Jean, I., & Scott, A. (2004). Competition, globalization and the decline of inflation. *Journal of International Economics, 63*(1), 93–118, 312.
Clarida, R. J. (2008). Reflections on monetary policy in the open economy. In J. Frankel & R. Clarida (Eds.), *NBER International Seminar on Macroeconomics* 2008.
Cukierman, A. (2017). Money growth and inflation: Policy lessons from a comparison of the US since 2008 with hyperinflation Germany in the 1920s. *Economic Letters, 154,* 109–112.
Cukierman, A., Lustenberger, T., & Meltzer, A. H. (2018). The permanent-transitory confusion: Implications for tests of market efficiency and for expected inflation during turbulent and tranquil times. CEPR Discussion Paper 13187.
De Grauwe, P. (2011). The European Central Bank as a lender of last resort. VoxEU, 18th August.
Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy, 91,* 401–419.
Dornbusch, R., & Sebastian, E. (1989). *The Macroeconomics Populism in Latin America*. WPS 316, World Bank.
Elkayam, D., & Ilek, A. (2016). Estimating the NAIRU for Israel, 1992–2013. *Israel Economic Review, 14*(1), October.
Fleming, M. (1962). Domestic financial policies under fixed and under floating exchange rates. IMF Staff Papers 9, Nov, 369–380.
Flood, R., & Garber, P. (1984). Collapsing exchange-rate regimes, some linear examples. *Journal of International Economics, 17,* 1–13.
Fratzscher, M., Lo Duca, M., & Straub, R. (2013). On the international spillovers of US quantitative easing.
Friedman, M. (1971). Government revenue from inflation. *Journal of Political Economy, 79*(4), 846–856.
Gali, J. (2008). *Monetary policy, inflation, and the business cycle: An introduction to the New Keynesian Framework*. Princeton, NJ: Princeton University Press.
Gopinath, G. (2004). Lending booms, sharp reversals and the real, exchange dynamics. *Journal of International Economics, 62*(1), 1.
Karni, E. (1983). Revenue from inflation. Yediot Aharonot.
Krampf, A. (2018). *The Israeli path to neoliberalism: The state, continuity and change*. Routledge.
Krugman, P. (1979). A model of balance-of-payments crises. *Journal of Money, Credit, and Banking, 11,* 311–325.
Krugman, P., Obstfeld, M., & Melitz, M. (2015). *International economics: Theory and policy* (10th ed.). London: Pearson.

Kydland, F., & Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy, 85,* 473–493.
Leiderman, L. (Ed.). (1999). *Inflation and disinflation in Israel*. Bank of Israel.
Leviathan, N. (1984). Dollarization as monetary reform. *The Economics Quarterly, 119*, 848–858.
Liviatan, N. (1984). Dollarization as a monetary reform. *The Economic Quarterly*, 848–858, (Hebrew).
Liviatan, N., & Piterman, S. (1986). Accelerating inflation and balance-of-payments crisis, 1973-1984. In Yoram Ben-Porath (Ed.), *The Israeli economy: Maturing through crises* (pp. 320–346). Cambridge, MA: Harvard University Press.
Loungani, P., & Razin, A. (2001). How beneficial is foreign direct investment for developing countries. *Finance and Development, 38*(2).
Loungani, P., Razin, A., & Yuen, C.-W. (2001). Capital mobility and the output-inflation trade off. *Journal of Development Economics, 64,* 255–274.
Mankiw, N. G. (2010). *The trilemma of international finance*. Available at: http://www.nytimes.com/2010/07/11/business/economy/11view.html.
Mishkin, F. S. (2007a). Is financial globalization beneficial? *Journal of Money, Credit, and Banking, 39*(2–3), 259–294.
Mishkin, F. S. (2007b). *Remarks given at the Annual Macro Conference*. Federal Reserve Bank of San Francisco, San Francisco, California.
Mundell, R. A. (1963). Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economics, 29,* 475–485.
Obstfeld, M. (1996). Models of currency crises with self-fulfilling features. *European Economic Review, 40,* 1037–1047.
Razin, A. (2014). *Understanding global crises: An emerging paradigm.* MIT Press.
Razin, A. (2018a). *Israel and the world economy: The power of Globalization.* MIT Press.
Razin, A. (2018b). Israel's immigration story: Winners and losers. *Israel Economic Review, 15*(1), 73–106.
Razin, A. (2019). Israel's struggle towards macroeconomic stability: Analytical essay. *Israel Economic Review, 7*(1).
Razin, A., & Efraim, S. (1993). *The economy of modern Israel: Malaise and promise*. University of Chicago Press.
Rey, H. (2015). Dilemma not trilemma: Global financial cycles and monetary policy independence. NBER working paper 21162.
Ribon, S. (2017). *Why the Bank of Israel intervenes in the foreign exchange market, and what happens to the exchange rate*. Bank of Israel DP No. 2017.04 February.
Rogoff, K. (2003). Disinflation: An unsung benefit of globalization? *Finance and Development, 40*(4), 55–56.
Rogoff, K. (2004). Globalization and global disinflation, in *Proceedings of the 2003 Jackson Hole Symposium Sponsored by the Federal Reserve Bank of Kansas City*. Federal Reserve Bank of Kansas City, Monetary Policy and Uncertainty: Adapting to a Changing Economy.
Romer, D. (1993). Openness and inflation: Theory and evidence. *Quarterly Journal of Economics, CVII*(4), 869–904.
Romer, D. (1998). A new assessment of openness and inflation: Reply. *Quarterly*.
Sargent, T. J. (1999). *The conquest of American inflation*. Princeton University Press.
Schneider, M., & Aaron, T. (2004). Balance sheet effects. Bailout Guarantees.
Sorezcky, A. (2015). Was the Bank of Israel intervention in the foreign exchange market effective? In A. Offenbacher (Ed.), *Maintaining price stability: The Bank of Israel's sixth decade*. The Bank of Israel Press.
Tetlow, R., & Ironside, B. (2007). Real-time model uncertainty in the United States: The fed, 1996–2003. *Journal of Money, Credit and Banking, 39*(7), 1533–1561.
Wynne, M. A., & Erasmus, K. K. (2007). Openness and inflation. Federal Reserve Bank of Dallas, Staff Paper No. 2. Zhang, Yu. 2011. Essays on banking. Ph.D. Reserve Bank of Dallas, Staff Paper 2.

The Permanent-Transitory Confusion: Implications for Tests of Market Efficiency and for Expected Inflation During Turbulent and Tranquil Times

Alex Cukierman, Thomas Lustenberger and Allan Meltzer

6 August 2018

Personal note on the history of the paper by Alex Cukierman

The origin of this paper goes back to an old unpublished manuscript by Cukierman and Meltzer (1982). A couple of years before that time Karl Brunner and Allan Meltzer became aware of the importance of the permanent-transitory confusion. I first discovered the universality of this confusion for the formation of expectations and for economic behavior when, as a visiting scholar at Carnegie-Mellon during the end of the seventies and beginning of the eighties, I started to interact with Allan and Karl on this topic. This interaction culminated in a number of joint published papers. The Cukierman and Meltzer (1982) paper was a later spinoff of this research effort and was never completed mainly because the research attention of both Allan and myself had turned to other topics and I had returned to Tel-Aviv.

But I always felt that the ideas in our unpublished manuscript are sufficiently important to justify bringing it up to date and amplifying its message with empirical work. This is particularly important for younger generations of economists who, due to the early criticism by the rational expectation school that adaptive expectations are not rational, might not be aware of the fact that Muth (1960) provided a statistical foundation for the permanent-transitory confusion in which adaptive expectations are rational.

When the organizers of the conference on "Expectations: Theory and applications in historical perspectives" suggested I write a paper for the conference I felt the time had come to do that. Unfortunately, on May 8, 2017, my long-time friend and collaborator Allan Meltzer passed away. My young collaborator, Thomas Lustenberger, joined me in this effort. The current updated and expanded version of the paper integrates the analysis in the original manuscript with some newer empirical work. Beyond the personal loss I feel along with Allan's family, the history of the paper is one illustration of the process through which the research torch is being passed across generations of economists. I view it as a personal token of appreciation to Allan Meltzer who had a major influence on my professional life.

Allan Meltzer passed away on May 8 2017.

This research project was conducted while Thomas Lustenberger was affiliated with the Swiss National Bank and the University of Basel.

A. Cukierman (✉)
Interdisciplinary Center and CEPR, Tel-Aviv University, Tel Aviv, Israel
e-mail: alexcuk@tauex.tau.ac.il

© Springer Nature Switzerland AG 2020

A. Arnon et al. (eds.), *Expectations*, Springer Studies in the History of Economic Thought,
https://doi.org/10.1007/978-3-030-41357-6_12

1 Introduction

In making decisions, economic agents and policymakers have to form expectations about the future. The importance of expectations is pervasive and of paramount importance for current decision making. Following are some well-known illustrations: When deciding how to allocate an increase in income between consumption and savings, individuals need to evaluate the permanence of this increase. A worker's decision about whether to accept or reject a poor job offer depends on his perception about the permanence of this condition. A firm's investment decision following strong demand for its product depends on its perception of the persistence of this state. When confronted with a strong economy, monetary policymakers may consider an increase in the policy rate. But if they believe the strength is temporary they are likely to postpone the increase. Similar considerations apply to contractionary fiscal policies.

Generally, even when they possess full information about current and past realizations of relevant variables, individuals remain uncertain about their permanence. In many cases, individuals detect the permanence of changes by observing the persistence of those changes over time. As a consequence, when permanent changes occur they are recognized only gradually. Adaptive expectations capture this sluggishness by making the difference between the current and the previous period's forecasts a positive function of the forecast error committed in the previous period. Muth (1960) has shown that when a stochastic variable is composed of a random walk and a white noise process, none of which is ever observed separately, adaptive expectations are rational in the sense that they utilize all available information in an efficient manner. For brevity, the paper refers to this residual uncertainty as the "permanent-transitory confusion" (PTC).

The first part of the paper reviews the history and past applications of Muth type adaptive expectations and considers their implications for standard tests of market efficiency. Using data on inflationary expectations from the Israeli capital market, the second part examines the performance of Muth's model in tracking those expectations during the turbulent 1985 Israeli stabilization as well as during the stable 2003–2018 period.

Tests of efficiency in the treasury bill market as predictors of inflation proceed by regressing the current realization of inflation on a lagged capital market variable that embody the preceding period's expectation of inflation. Relying on Fisher's theory of interest, this signaling variable is taken to be the lagged value of the nominal interest rate. In tests of efficiency of foreign exchange markets, the signaling variable is taken to be the forward exchange rate leading to formulations in which the current rate of

T. Lustenberger
University of Basel, Basel, Switzerland
e-mail: thomas.lustenberger@unibas.ch

Swiss Financial Market Supervisory Authority FINMA, Bern, Switzerland

A. Meltzer
Carnegie-Mellon University, Pittsburgh, USA

change in the exchange rate is regressed on the rate of change implied by the past forward rate. In either case, the appearance of serial correlation in the residuals of those regressions is considered as evidence against market efficiency. The intuition supporting this view is that, if markets were efficient, rational individuals should have used it in their predictions leading to the disappearance of serial correlation. A central result of the first part of this paper is that, in the presence of the permanent-transitory confusion, the appearance of serial correlation in finite samples does not necessarily imply that markets are inefficient.

The paper is organized as follows. Section 2 reviews the history and past applications of adaptive expectations and presents Muth's (1960) statistical foundations for it. Section 3 reviews standard tests of market efficiency in the treasury bill and in the foreign exchange markets. Section 4 contains a main result of the paper. It proposes a generalization of the tests in Sect. 3 and uses it to show that, following the realization of large permanent changes, the appearance of serial correlation in the residuals of the regressions used to implement those tests does not necessarily indicate that markets are inefficient.

An attractive feature of Muth (1960) foundation for adaptive expectations is that it relates the speed of learning about permanent changes to the relative size of the variability of the permanent component of a shock to the variability of the transitory component of the shock. Sections 5 and 6 use data on Israeli inflation expectations from the capital market along with this relation in order to estimate the implicit speed of learning about changes in inflation and to examine the performance of Muth's adaptive expectations model in tracking the evolution of capital market expectations. Section 5 focuses on the period before and after the 1985 stabilization that led, after a while, to a substantial decrease in expected inflation. The numerical exercise suggests that adaptive expectations provide a good approximation for the evolution of capital market expectations during this period. Section 6 applies a similar methodology to the stable inflation targeting period between 2003 and 2018. The numerical exercise supports the conclusion that, during this period, capital market participants considered all deviations from the inflation target as transitory. This is followed by concluding remarks.

2 Adaptive Expectations Through the Ages and Muth Model of the Permanent-Transitory Confusion

2.1 Adaptive Expectations

Adaptive expectations have been around for over a century. Although their roots go back to Fisher (1911), they gained prominence and became operational in macroeconomics with the empirical work of Cagan (1956) on hyperinflations during the twentieth century and Friedman (1957) research on the permanent income hypothesis. Cagan used adaptive expectations to characterize the links between actual inflation in the past and inflationary expectations during the hyperinflation. Friedman applied

them to model and estimate the links between perceived future permanent income and past realizations of actual income.

The basic idea of adaptive expectations is quite intuitive. It states that, when new information about a variable that is being forecasted becomes available over time, individuals adjust their expectations about the future realization of this variable in proportion to the forecast error committed in the previous period. For this reason, the process is also frequently characterized as an "error correction process." Formally, adaptive expectations are given by

$$y_t^e - y_{t-1}^e = \theta\left(y_t - y_{t-1}^e\right) \tag{1}$$

where y_t is the actual realization of a variable y in period t and y_t^e is the forecast of that variable given the information available in period t. The adaptive expectations coefficient, θ, characterizes the speed with which the public incorporates recent developments into its forward-looking expectations. In empirical applications, θ is usually assumed to be bounded between 0 and 1. Moving y_{t-1}^e to the right-hand side, lagging by one period in order to express y_{t-1}^e in terms of y_{t-1} and y_{t-2}^e, inserting the resulting expression into Eq. (1), and repeating this procedure ad infinitum y_t^e can be rewritten in the integral form

$$y_t^e = \sum_{i=0}^{\infty} \theta(1-\theta)^i y_{t-i}. \tag{2}$$

With the onset of the rational expectations revolution, Lucas (1972) and others criticized adaptive expectations on the ground that they were backward rather than forward looking. Rational expectations imply that

$$y_t^e \equiv E_t y_{t+1}$$

where $E_t y_{t+1}$ is the expected value of y_{t+1} given the information available up to and including period t. As shown in the next subsection, and as recognized later, the criticism above is not justified in the presence of the permanent-transitory confusion.

2.2 *The Permanent-Transitory Confusion and Muth (1960) Statistical Foundations for It*

The permanent-transitory confusion (PTC) refers to the widespread fact that knowledge of current and past changes in a stochastic variable normally leaves a margin of uncertainty about how much of those changes will persist into the future and how much are just temporary changes that will fade away as the future unfolds. The PTC is a pervasive fact of life that confronts investors, consumers, producers, and policymakers when they make current decisions. In a path breaking article, Muth

(1960) developed the following stylized statistical model for the PTC.[1] The model postulates that the stochastic variable, y_t, is the sum of two stochastic components none of which is ever observed separately. One is a random walk that persists into the future, and the other is a transitory white noise that appears in period t and does not persist at all into the future. More formally

$$\begin{aligned} y_t &= y_t^p + y_t^q, \\ \Delta y_t^p &\sim N\left(0, \sigma_p^2\right), \\ y_t^q &\sim N\left(0, \sigma_q^2\right), \\ \Delta y_t^p &\text{ and } y_t^q \text{ are mutually independent.} \end{aligned} \tag{3}$$

Here Δy_t^p is the first difference of the random walk (permanent) component and y_t^q is the white noise (transitory) component. Muth (1960) has shown that the **forward-looking** optimal predictor of $y_{t+j}, j \geq 1$ given the information set, $I_t \equiv \{y_t, y_{t-1}, y_{t-2}, \ldots\}$, available in period t is identical to the adaptive expectation process in Eqs. (1) and (2).[2] Furthermore, the coefficient θ is an increasing function of the ratio, a, between the variance, σ_p^2, of the innovation to the random walk component and the transitory variance, σ_q^2, and is given by

$$\theta = \sqrt{a + \frac{a^2}{4}} - \frac{a}{2}, a \equiv \frac{\sigma_p^2}{\sigma_q^2}. \tag{4}$$

Muth's optimal predictor has some notable and convenient features that are briefly summarized in what follows. First, it implies that it is optimal to utilize **all past** observations on y_t in order to forecast the **future**. Second, Eq. (2) implies that it is a Koyck lag with geometric weights that decrease the more distant in the past is the observation on y. Third, the weights sum up to one. Fourth, the larger is the adaptive expectations coefficient, θ, the larger is the sum of the weights on the most recent past in comparison to the more distant past. Consequently, the larger is θ, the faster is the speed at which individuals detect a permanent change when such a change has occurred implying that θ characterizes the speed of learning. Finally, it is not surprising that θ and the ratio, a, between the permanent and the transitory variances are positively related. The higher is a, the higher is the signal-to-noise ratio implying that optimal learning should be faster.

The more general message of the preceding discussion is that, although predictors of the future are forward looking, they normally rely on past information since the

[1] Although this article is relatively less known (and quoted) than Muth (1961), Econometrica article that inspired the rational expectations revolution in macroeconomics its contribution is, nonetheless, not less important.

[2] Statistically minded readers may note that this optimal predictor is the expected value of $y_{t+j}, j \geq 1$ conditional on the information set, $I_t \equiv \{y_t, y_{t-1}, y_{t-2}, \ldots\}$. Due to the normality assumption, this conditional expected value is linear in the elements of the conditioning set and the weights are those that minimize the variance of forecasts around this expected value.

past contains useful, albeit imperfect, information about the future. During the early days of the rational expectations revolution, some economists criticized adaptive expectations on the ground that they are backward rather than forward looking. This criticism is probably based on perfect foresight models like that of Barro and Gordon (1983) that do not feature stochastic terms. In such models, rational expectations reduce to the, known with certainty, values of relevant variables as predicted by such models. But once the more realistic existence of stochastic terms and the PTC are incorporated into models, the role of **past information in predicting the future** becomes essential. Muth's predictor provided an early convenient way to capture the main features of the PTC and to relate it to natural intuition. But it is by no means, the only way to do that. A multi-variables generalization is provided by the Kalman filter (Kalman 1960).[3]

2.3 Past Applications of Muth's Predictor

Lucas and Rapping (1969) develop a model of employment/unemployment in which individuals decide how much of their employment efforts to allocate to the present versus the future. This decision is based on a comparison of their current wage with what they believe is their long-run normal or permanent wage rate. Brunner et al. (1980) embed this mechanism along with Friedman's permanent income hypothesis into an extended IS-LM model. They utilize Muth's predictor to characterize the behavior of individual expectations about permanent income and permanent wages. Cukierman (1982) uses it to investigate the behavior of relative prices and of the allocative efficiency of the price system in the presence of the PTC about individual prices in a Lucas (1973) type multi-markets model.

3 Tests of Market Efficiency in the Treasury Bills and Foreign Exchange Markets

To test for the efficiency of short-term treasury bill rates as predictors of future inflation Fama (1975) relied on Fisher's (1930) theory of interest according to which those rates reflect the sum of the equilibrium real interest rate and the rate of change in the real value of money expected to realize over the life of the bill. The efficient markets or rational expectations hypothesis implies that in a linear regression of the rate of change in the real value of money on a previous market forecast of this change, there should be no correlation in the residuals. Sample evidence of serial correlation in the residuals is taken to imply that individuals do not utilize all currently available information in an efficient manner since errors of forecast can be reduced by using the

[3] A compact useful presentation of the Kalman filter appears in Chapter 21 of Ljungqvist and Sargent (2000).

information contained in the persistent deviations of actual values from the forecast values implied by observable market values.

The simplicity of the test and the intuitive appeal of Fama's interpretation led to its application in other asset markets. Hamburger and Platt (1975) used current values of forward rates on treasury bills to forecast future spot rates. They found evidence of positive serial correlation in the residuals from some of their regressions and corrected for this "inefficiency" using the first-order Cochrane–Orcutt procedure. Frenkel (1977, 1979) and many others subsequently used very similar procedures to test for the efficiency of forward rates as predictors of future spot exchange rates. Figlewski and Wachtel (1981) tested the rationality of individual price expectations by checking whether forecast errors are serially correlated and found those errors to be serially correlated. They concluded that survey respondents did not use all available information and that, consequently, the rational expectations (RE) hypothesis is violated.

3.1 *Fama (1975) Early Efficiency Test of Current Interest Rates as Predictors of Future Inflation*

Fama (1975) tested the efficiency of one month treasury bills (TB) as predictors of the decrease in the real value of money over the remaining life of a bill as follows. The starting point of the test is the theory by Fisher (1930) according to which

$$\Delta_{t+1} = -r_t + R_t \tag{5}$$

where Δ_{t+1}, R_t, and r_t are the decrease in the real value of money between month t and month $t+1$, the nominal and real rates at time t, respectively, and second-order terms have been dropped. R_t is observed on the market at time t but Δ_{t+1} and r_t are stochastic variables at that time. The test consists in running the regression

$$\Delta_{t+1} = \alpha_0 + \alpha_1 R_t + \varepsilon_t \tag{6}$$

Under rational expectations cum risk neutrality and the additional assumption that the real rate is constant the hypothesis that short-term nominal rates are efficient predictors of the upcoming monthly inflation reduces to a test of the joint hypothesis that $\alpha_1 = 1$, $\varepsilon_t = \Delta_{t+1} - E_t \Delta_{t+1}$ is a serially uncorrelated forecast error and α_0 is an estimate of minus the (assumed) constant real rate.[4] Note, in particular, that detection of serial correlation is taken as evidence against market efficiency. In Fama's words (1975, p. 273):

[4]Fama and others subsequently extended the test to fluctuating real rates. The central point of the next section applies to those extensions as well.

Nonzero autocorrelations imply that the market is inefficient; one can improve on the market's assessment of the expected value of Δ_{t+1} by making correct use of information in past values of Δ_t.

3.2 *Efficiency Tests of Forward Premia as Predictors of Future Spot Exchange Rates*

The forward premium is the difference between the current forward and spot exchange rates. Similarly, to the case of nominal rates as predictors of future inflation, efficiency tests in the foreign exchange market are based on the notion that current forward market quotations embody expectations about future spot rates. Provided expectations are rational and market participants are risk neutral the forward premium should provide an unbiased estimate of the current market assessment of the change in the spot rate between the future maturity period of the forward rate and the current spot. More precisely, consider the regression of the change in the log of the spot exchange rate on the forward discount (expressed in log form)[5]

$$s_{t+1} - s_t = \alpha + \beta(f_t - s_t) + u_{t+1} \tag{7}$$

Here, s_t is the log of the spot price of foreign currency at time t, f_t is the log of the one-period forward exchange rate at time t, and u_{t+1} is the regression disturbance. The general idea is that under risk neutrality and rational (or efficient) expectations, the log of the forward rate provides an unbiased forecast of the log of the future spot exchange rate implying that u_{t+1} is a serially uncorrelated forecast error with zero mean. Translated into statistical hypothesis testing this implies the commonly tested null hypothesis that $\alpha = 0$, $\beta = 1$ and u_{t+1} has mean zero and is serially uncorrelated.[6] The intuition underlying the null hypothesis is that under risk neutrality and rational expectations, the forward premium should equal the expected value as of period t of the spot rate in period $t + 1$ implying that α should equal zero and that β should be equal to one.[7] Rational or efficient expectations also imply that u_{t+1} should have a zero mean and be serially uncorrelated since the contrary would imply that

[5]For simplicity of exposition, we focus in the text on the one period ahead forward premium as a predictor of the change in the exchange rate between the current and the next period. However, all the discussion that follows in the text also applies to the k periods ahead forward premium. In this case, Eq. (7) is simply replaced by

$$s_{t+k} - s_t = \alpha + \beta(f_{t+k} - s_t) + u_{t+k}.$$

[6]This equation is the canonical regression used in the voluminous literature on the forward premium puzzle. See Chinn (2009), Eq. (2) and the adjoining discussion. Early formulations of the test were done in levels rather than in actual and expected rates of change (Frenkel (1977) and Frenkel (1979)).

[7]Subsequent literature such as Fama (1984) recognized the potential existence of risk aversion by introducing a risk premium into regression (7). A survey of this literature appears in Engel (1996).

individuals do not efficiently utilize all the information available in period t violating the assumption of rational expectations.

3.3 A More General Reformulation of Market Efficiency Tests

The common feature in efficiency tests of short rates as predictors of inflation and of forward rates as predictors of the future rates is that, in both cases, currently observed market variables contain information about current expectations of future variables. This idea is captured more precisely by the following formalization

$$x_t = c_0 + cy_t^e \tag{8}$$

where x_t is a market variable observed at time t and c_0, c are constant coefficients that depend on the particular model under consideration. Solving for y_t^e in terms of x_t

$$y_t^e = -\frac{c_0}{c} + \frac{1}{c}x_t. \tag{9}$$

Consider the identity,

$$y_{t+1} = y_t^e + \left(y_{t+1} - y_t^e\right). \tag{10}$$

Replacing the first y_t^e on the right-hand side of this identity by Eq. (9)

$$y_{t+1} = -\frac{c_0}{c} + \frac{1}{c}x_t + \left(y_{t+1} - y_t^e\right) \tag{11}$$

we obtain a general formulation that subsumes the efficiency tests of the two preceding subsections as particular cases. It states that the realization of y_{t+1} is a linear function of period's t observed market variable, x_t, plus a forecast error, $y_{t+1} - y_t^e$. Efficiency of this more general model can be tested by running the regression

$$y_{t+1} = \beta_0 + \beta x_t + u_{t+1} \tag{12}$$

and by testing the restrictions on β_0, β and u_{t+1} implied by market efficiency for each of the models subsumed under the general formulation in Eq. (11).

When $y_{t+1} = \Delta_{t+1}$, $x_t = R_t$ and $c = 1$ Eq. (11) reduces to the regression used by Fama to test the efficiency of current rates in predicting future inflation (compared to Eq. (6)).[8]

When $y_{t+1} = s_{t+1}$, $x_t = f_t$, $c_0 = 0$ and $c = 1$, Eq. (11) reduces to the canonical regression used to test the efficiency of the forward premium in predicting future spot exchange rates (compare to Eq. (7)).

4 The Impact of Occasionally Large Permanent Shocks on the Serial Correlation in Forecast Errors: The Case of Finite Samples

Using Muth (1960) type optimal adaptive expectations, this section shows that in finite samples that are occasionally subject to the realization of relatively large permanent shocks, estimated forecast errors will be serially correlated even when expectations are rational and markets are efficient. The wider implication of this result is that detection of such serial correlation does not necessarily indicate that markets are inefficient. To demonstrate this statement, we focus on the general formulation of tests of market efficiency (Eq. 11) in the presence of Muth's specification of the PTC (Eq. 3). Equation (2) along with the optimality of those expectations implies

$$y_t^e = E_t y_{t+1} = \sum_{i=0}^{\infty} \theta(1-\theta)^i y_{t-i}. \tag{13}$$

Period's $t+1$ forecast error is given by

$$u_{t+1} = y_{t+1} - \sum_{i=0}^{\infty} \theta(1-\theta)^i y_{t-i}. \tag{14}$$

Following simple but tedious algebraic manipulations, the forecast error can be rewritten[9]

$$\begin{aligned} u_{t+1} &= \overbrace{y_{t+1}^q - \sum_{i=0}^{\infty} \theta(1-\theta)^i y_{t-i}^q}^{\equiv Q_{t+1}} + \overbrace{\sum_{i=0}^{\infty} (1-\theta)^i \Delta y_{t+1-i}^p}^{\equiv P_{t+1}} \\ &= Q_{t+1} + P_{t+1} \end{aligned} \tag{15}$$

[8] As was the case before the estimate of $\beta_0 = c_0$, provides an estimate of minus the (assumed) constant real rate of interest.

[9] Details appear in Sect. 1 of the Appendix.

The first two terms on the right-hand side of this expression summarize the impact of period's $t + 1$ transitory component and of all past transitory components on period's $t + 1$ forecast error. The last term summarizes the impact of all past innovations to the permanent component up to and including period $t + 1$ on this forecast error. An immediate consequence of Eq. (15) is that the current forecast error depends on **all** the past history of shocks to both the permanent and the transitory components of y. Since individuals never observe (not even ex post) the permanent and transitory components of y separately, this should not come as a surprise.

Since the transitory shocks and the innovations to the permanent shocks have zero expected value and are serially and mutually independent,

$$\text{Cov}(u_{t+1}, u_t) = Eu_{t+1}, u_t = EQ_{t+1}Q_t + EP_{t+1}P_t \tag{16}$$

and

$$\text{Var}(u_{t+1}) = E\{Q_{t+1} + P_{t+1}\}^2 = E\{Q_{t+1}\}^2 + E\{P_{t+1}\}^2 \tag{17}$$

where

$$Q_{t+1} \equiv y^q_{t+1} - \sum_{i=0}^{\infty} \theta(1-\theta)^i y^q_{t-i} \text{ and } P_{t+1} \equiv \sum_{i=0}^{\infty} (1-\theta)^i \Delta y^p_{t+1-i}. \tag{18}$$

It is shown in Sect. 2 of the Appendix that in spite of the infinite series of overlapping terms between $u_{t+1.}$ and u_t, the first-order covariance between those forecast errors in the population is zero.[10] But, when a relatively large permanent innovation occurs in a finite sample, the covariance between adjacent forecast errors may be positive for a sufficiently long time to produce evidence in favor of first-order serial correlation in spite of the fact that the predictor in Eq. (13) is optimal.

The reason is that the public is unable to fully identify permanent changes even after the fact. They learn gradually, but optimally, according to Eq. (13), by observing that y maintains a value that is greater (or lower) than expected for some time. If the learning parameter, θ, is sufficiently low, econometricians that implement market efficiency tests may find evidence of serially correlated forecast errors in finite samples that are dominated by the realization of a large permanent shock.

To show that expost forecast errors appear to be serially correlated under the circumstances just described, we focus on the coefficient of correlation between adjacent forecast errors following the realization in period t of a relatively large permanent innovation, Δy^p_t. In order to focus on the impact of a large permanent shock in comparison to the normal variabilities of both shocks, we assume that all the other realizations of the transitory and permanent innovations are equal to their respective standard deviations. The formula for this conditional (on a large Δy^p_t) coefficient of correlation is

[10] It is likely that this is the case also for higher-order covariances between forecast errors.

$$\rho_j\left(\Delta y_t^p\right) \equiv \frac{E\left\{u_{t+j+1}u_{t+j}|\Delta y_t^p\right\}}{\sqrt{E\left(u_{t+j+1}\right)^2 E\left(u_{t+j}\right)^2}} j \geq 0 \tag{19}$$

where the symbol E stands for the expected value over the distributions of both the permanent and transitory shocks. It is shown in Sect. 3 of the Appendix that this coefficient is given by

$$\rho_j\left(\Delta y_t^p\right) = (1-\theta)^{2(j+1)}\left[\frac{\left(\Delta y_t^p\right)^2}{\sigma_q^2} - \frac{\sigma_p^2}{\sigma_q^2}\right] \tag{20}$$

Note that, when the squared ratio of period's t permanent shock to the transitory variance is identical to the signal-to-noise ratio, $\frac{\sigma_p^2}{\sigma_q^2}$, $\rho_j\left(\Delta y_t^p\right)$ is zero. This provides a "normal" benchmark value for $\rho_j(\Delta y_t^p)$. But, following a large realization of this squared ratio in comparison to the signal-to-noise ratio $\rho_j(\Delta y_t^p)$ is positive.[11] Due to gradual learning, it is largest in the period immediately following the realization of the large permanent shock. It then gradually declines to zero as the impact of the shock on current expectations fades into the past. When the learning parameter, θ, is relatively low (or equivalently $\frac{\sigma_p^2}{\sigma_q^2}$ is low), this positive sample correlation may persist for quite a while before it finally converges to its normal zero value. On the other hand, the likelihood that a relatively large value of the permanent shock occurs lower when $\frac{\sigma_p^2}{\sigma_q^2}$ is low.

The upshot is that although the probability of a large realization of Δy_t^p is low when $\frac{\sigma_p^2}{\sigma_q^2}$ is low, if such a low probability event does occur, it induces in finite samples persistent measured serial correlation in forecast errors. Figure 1 illustrates the behavior of $\rho_j\left(\Delta y_t^p\right)$ for $\frac{\left(\Delta y_t^p\right)^2}{\sigma_q^2} = 3$ and $\theta = 0.01$. The figure shows that following the realization of this large permanent shock the covariance between forecast errors is larger than the variance of those errors $\left(\rho_j\left(\Delta y_t^p\right) > 1\right)$ for over 50 periods after the realization of this shock in spite of the rationality of expectations. On the other hand, once the speed of learning rises above 0.2, most of this persistence vanishes given the same value of $\frac{\left(\Delta y_t^p\right)^2}{\sigma_q^2}$.

The more general lesson from this exercise is that in tests of efficiency of the treasury bill market, the failure to reject serial correlation can be missleading if applied to samples taken shortly after violent changes in the purchasing power of money. Similarly, the serial correlation test may yield wrong conclusions about the efficiency of the foreign exchange market if applied during or shortly after large permanent changes in the exchange rate. Interestingly, Frankel and Poonawala (2006), Table 2, reject the null hypothesis of no serial correlation in forecast errors at the 5% significance level for India, Indonesia, and Turkey. Our analysis implies that this

[11] Note that, since it depends on a particular realization of the innovation to the permanent component, $\rho_j(\Delta y_t^p)$ is not necessarily smaller than one.

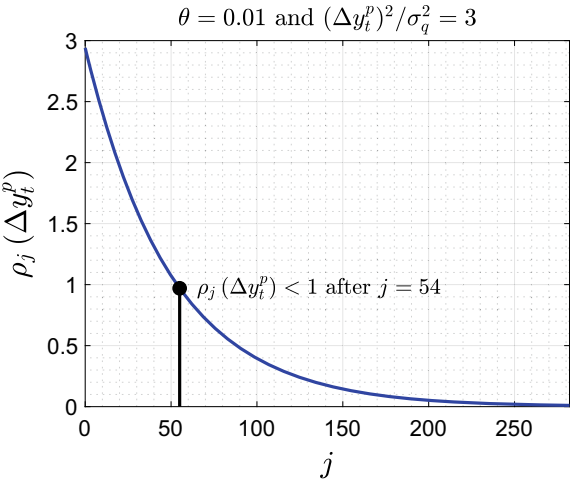

Fig. 1 Conditional coefficient of correlation

finding does not necessarily imply that foreign exchange markets in those countries are inefficient.

5 Turbulent Times: The Behavior of Expected Inflation During Disinflation

Since about the mid-nineties, the Bank of Israel has been deriving estimates of expected inflation from the difference between the yields to maturity on indexed and non-indexed government bonds. Due to the absence of long-term nominal bonds at the start of the period, those estimates, also known as breakeven inflationary expectations, were initially limited to forecast horizons of one year. But, as inflation subsided at the beginning of the twenty-first century, the Israeli treasury issued nominal bonds with longer maturities making it possible to derive longer-term inflationary expectations from the bond market up to a horizon of ten years.

5.1 An Empirical Application to the Israeli 1985 Cold Turkey Stabilization

A "cold turkey" or "shock" stabilization refers to a situation in which high inflation is stabilized very aggressively within a short period of time. Following seven years with yearly rates of inflation of 100% or more and several failed attempts to stabilize inflation Israel finally managed to stabilize it in July 1985 bringing the rate of inflation down from about 400% to almost zero within a couple of months. This dramatic drop

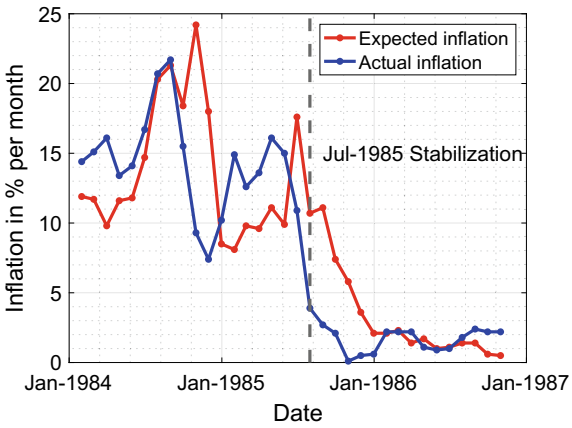

Fig. 2 Expected three months ahead and actual inflation from January 1984 to October 1986

was achieved through the simultaneous deployment of conventional measures like restrictive fiscal and monetary policies as well as less conventional measures such as temporary controls on prices, wages, and the exchange rate.[12]

It can be concluded with the benefit of hindsight that the 1985 cold turkey stabilization produced a large permanent drop in the rate of inflation. However, at the time of the stabilization, there was substantial uncertainty about the extent to which this dramatic drop will persist. This uncertainty was induced by wide gyrations in inflation and several failed attempts to stabilize prior to the 1985 successful stabilization. It is therefore instructive to examine the behavior of inflationary expectations before and after the 1985 stabilization.

Although capital market inflationary expectations were not calculated on a systematic basis prior to the mid-nineties, they were occasionally estimated also prior to that time. In particular, Table 2.2 in Cukierman (1988) provides average monthly breakeven expected inflation over a three month horizon along with average monthly inflation over the same horizon between January 1984 and October 1986. Figure 2 plots actual and previously expected average inflation at monthly rates for this period.

Perusal of the figure suggests that breakeven inflationary expectations lagged behind changes in the actual rate of inflation. It is likely, therefore, that this gradual adjustment of expectations indicates that expectations are adaptive and that Muth's model of the PTC may provide a reasonable approximation to the behavior of actual inflation and of breakeven expectations during the time period displayed in the figure. To examine this possibility, the next subsection utilizes the data on actual and expected inflation underlying Fig. 2 to estimate the learning parameter, θ, and the variances, σ_p^2 and σ_q^2, of the permanent and transitory shocks to the components of inflation over this period.

[12] A detailed description of the 1985 stabilization appears in Bruno and Piterman (1988).

5.2 *Estimation of the Learning Parameter During the 1985 Stabilization*

Since observations on breakeven expectations are available, the parameter θ that fits the data best can be estimated from Eq. (1) where y_t and y_t^e stand now for actual and expected inflation. Estimation of σ_p^2 and of σ_q^2 requires the prior estimation of the variance of the first difference of actual inflation. Taking the first difference of y_t from Eq. (3),

$$\Delta y_t = \Delta y_t^p + y_t^q - y_{t-1}^q. \tag{21}$$

Equations (3) and (21) imply that the variance, $\sigma_{\Delta y}^2$, of Δy_t is

$$\sigma_{\Delta y}^2 = \sigma_p^2 + 2\sigma_q^2. \tag{22}$$

$\sigma_{\Delta y}^2$ is estimated by taking first differences of y_t and by calculating the variance of those differences over the sample period. It is shown in Sect. 4 of the Appendix that Eq. (4) is equivalent to

$$a \equiv \frac{\sigma_p^2}{\sigma_q^2} = \frac{\theta^2}{1-\theta}. \tag{23}$$

Finally, given the estimates of θ and of $\sigma_{\Delta y}^2$ Eqs. (22) and (23) are used to obtain estimates of σ_p^2 and of σ_q^2. The estimated values are $\theta = 0.32$, $\sigma_p^2 = 0.49$, $\sigma_q^2 = 3,34$ implying that the signal-to-noise ratio, $a = \frac{\sigma_p^2}{\sigma_q^2}$, is 0.14. Given the estimate of the learning parameter, θ, simulated values of the breakeven expectations are calculated by using expected inflation and actual figures in Eq. (1).

Figure 3 shows simulated values of the three months ahead capital market expectations along with the actual values of those expectations. It is apparent from the figure that Muth's stochastic structure with $\theta = 0.32$ performs quite well in tracking actual values of those expectations particularly following the July 1985 stabilization. This conclusion is also backed by the finding that the ratio between the sum of squared deviations of simulated from actual values of expectations and the variance of actual expectations is only **0.15**. This evidence supports the conclusion that the stochastic structure postulated in Muth (1960) fits the data around the 1985 stabilization of inflation reasonably well. More precisely, it implies that inflation during the 1984–1986 period can be characterized as the sum of a random walk and of a white noise (Eq. 3).

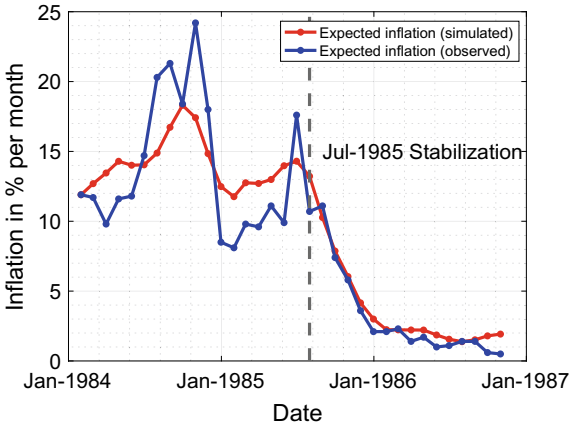

Fig. 3 Actual and simulated three months ahead inflation expectations from January 1984 to October 1986

6 The Performance of Adaptive Expectations During Tranquil Times: Israel 2003–2018

Although the July 1985 stabilization permanently reduced inflation below 20% per year, inflation converged to the vicinity of the 2% international standard only at the beginning of the twenty-first century. An, initially informal, inflation target regime was inaugurated in the mid-nineties. The target was first used as an instrument for reduction of expectations without excessive decreases in economic activity and was gradually decreased from year to year when the previous year's target was attained. It finally converged to a long-run fixed inflation target of 2% central target with an allowable band between 1 and 3% at the beginning of 2003. From that point and on, actual inflation remained most of the time within this target range.[13]

The main objective of this section is to examine empirically the ability of Muth's adaptive expectations model to provide a characterization of capital market expectations and to estimate the speed of learning, θ, during the 2003–2018 tranquil period. During this period, 10 as well as one year ahead capital market inflationary expectations are available. Figure 4 shows actual and long-term expected capital market expectations along with the fixed inflation target for this period.

6.1 Estimation of the Learning Parameter During the Tranquil 2003–2018 Period

The discussion in the section focusses on 10 years ahead expectations but results for a one year horizon are briefly reported as well. The methodology for estimation of

[13] A detailed description of the convergence process and other details appear in Cukierman and Melnick (2015).

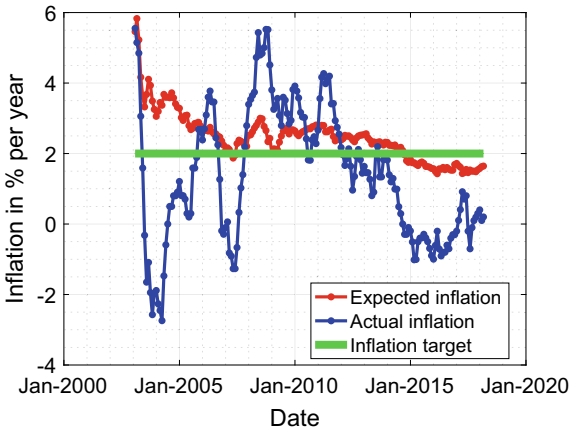

Fig. 4 Expected ten years ahead and actual inflation from January 2003 to February 2018

the parameters θ, σ_p^2, σ_q^2 and a is similar to the estimation of those parameters in the turbulent period discussed in the previous section except for the fact that during this period there was a fixed inflation target. To take into consideration the existence of a pre-announced fixed long-term target, δ, during the tranquil period, the actual rate of inflation (denoted now π_t) is respecified as,

$$\begin{aligned} \pi_t &= \delta + y_t = \delta + y_t^p + y_t^q \\ \pi_t^e &= \delta + y_t^e \end{aligned} \tag{24}$$

where the stochastic properties of y_t and of its constituent components are given in Eq. (3). That is, actual inflation is equal to a full certainty known in advance permanent target plus a stochastic deviation, y_t, that possesses the stochastic properties postulated by Muth to describe the PTC. Consequently, the optimal forecast of π_t^e is given by the second line in Eq. (24). Rearranging Eq. (24)

$$\begin{aligned} y_t &= \pi_t - \delta \\ y_t^e &= \pi_t^e - \delta. \end{aligned} \tag{25}$$

Since y_t has the same stochastic properties as in the previous section, the parameters θ, σ_p^2, σ_q^2 and a can be estimated by applying the procedure from that section to $\pi_t^e - \delta$. It is shown in Sect. 5 of the Appendix that the estimates of θ and of $\sigma_{\Delta y}^2$ obtained by using y_t and y_t^e are identical to the estimates using the original actual and expected inflation figures π_t and π_t^e. Hence, the procedure used in the previous section for estimation purposes can be applied directly to the expected inflation figures before the transformations in Eqs. (25).

The estimated values are $\theta = 0.01$, $\sigma_p^2 = 0.00002$, $\sigma_q^2 = 0.13$ implying that the signal-to-noise ratio, $a = \frac{\sigma_p^2}{\sigma_q^2}$, is 0.0001. In sharp contrast to the turbulent 1985 high inflation period, the variance of the stochastic permanent component is almost zero

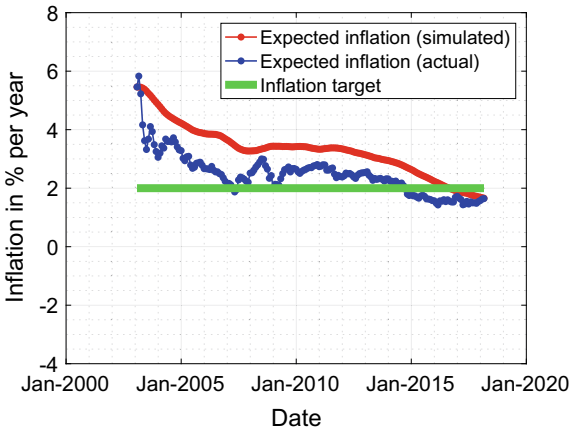

Fig. 5 Actual and simulated ten years ahead inflation expectations from January 2003 to February 2018

implying that the speed of learning about this component is extremely slow. In other words, during the tranquil period, capital market participants practically considered all deviations from the long-term 2% inflation target as transitory supporting the view that long-term inflationary expectations were well anchored to the 2% target. Figure 5 shows actual and simulated values of long-term expectations.

The ultimate emergence of a well-maintained inflation targeting regime neutralized the impact of the stochastic PTC on long-term expectations replacing it by a non-stochastic permanent inflation target of 2% implying that Muth (1960) process does a poor job of characterizing the behavior of ten years expectations data. This conclusion is backed by the finding that the ratio between the sum of squared deviations of simulated from actual values of expectations and the variance of actual expectations is a huge **1.62**.

Estimation results for one year ahead inflation expectation are broadly similar except that the speed of learning and the signal-to-noise ratio are somewhat higher. They are $\theta = 0.02, \sigma_p^2 = 0.00005, \sigma_q^2 = 0.13$ implying that the signal-to-noise ratio, $a = \frac{\sigma_p^2}{\sigma_q^2}$, is 0.0004. But, due to an almost doubling of the speed of learning, the fit of simulated expectations is better than in the case of ten years ahead expectations. This is reflected in the finding that the ratio between the sum of squared deviations of simulated from actual values of expectations and the variance of actual expectations for the one years ahead expectations drops to **0.84**.

The upshot from this experiment is that, in contrast to the turbulent period, Muth (1960) stochastic assumptions and optimal predictor does not capture the behavior of both long- and short-term inflationary expectations well during the tranquil period. Instead, it supports the view that capital markets participants considered all deviations for the inflation target as transitory. The broader consequences of this finding are discussed in the concluding section.

7 Concluding Remarks

This paper reviewed the history of adaptive expectations as a vehicle for modeling the permanent-transitory confusion. A central result of the first part of the paper is that, in the presence of this confusion, the appearance of serial correlation in tests of market efficiency based on finite samples does not necessarily imply that markets are inefficient. This implies that the detection of serial correlation in tests of efficiency in the treasury bill and in the foreign exchange markets does not necessarily imply that the expectations embodied in interest rates and in forward exchange rates are not rational in the sense that they disregard relevant information.

Although the early rational expectations literature criticized adaptive expectations on the ground that they are backward rather than forward looking, the work of Muth (1960) demonstrated that in the presence of the permanent-transitory confusion the optimal forecast of the future relies on information from the past. Muth considered only the case of a single stochastic variable in which the permanent component is a random walk and the transitory component is a white noise process. But the work of Kalman (1960) on the Kalman filter and subsequent literature suggest that, generally, optimal forecasts of the future rely on available past and current information and that this statement is true for a large class of more general processes that include both stationary and non stationary stochastic processes.[14] The crucial feature underlying this regularity is that the stochastic variables considered are composed of shocks with different degrees of persistence none of which is observed separately.[15]

Using Israeli data on inflationary expectations from the capital market, the second part of the paper examines the performance of adaptive expectations in tracking those expectations during the 1985 Israeli stabilization as well as during the tranquil stable inflation targeting period. Adaptive expectations perform quite well prior to and shortly after the cold turkey 1985 stabilization but not during the tranquil inflation targeting period (2003–2018).

As a matter of fact in the latter period, the empirical results are consistent with the view that individuals in the capital market believed that the long-run inflation rate is given by the two percent pre-announced stable inflation target and interpreted any deviation of inflation from this target as temporary. The wider economic implication is that, during the stable inflation targeting period, capital market expectations were well anchored.[16] At the technical level, this suggests that an expectation process in which the only permanent component is the pre-announced fixed inflation target and the temporary component is stationary is likely to produce a better fit for modeling the behavior of capital market expectations.

---

[14] One example is Chapter 21 of Ljungqvist and Sargent (2000).

[15] Furthermore, as demonstrated by Friedman (1979), serial correlation may also arise when a slope coefficient of an economic model changes permanently. The reason is that an econometrician using least square becomes aware of the change only gradually as post-change observations cumulate over time.

[16] This conclusion is consistent with results obtained in Cukierman and Melnick (2015).

Acknowledgements We benefited from useful discussions on tests of market efficiency in the foreign exchange market with Menzie Chinn. The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the SNB.

Appendix

Derivation of Equation (15)

Inserting Eq. (3) into Eq. (14) for all t

$$u_{t+1} = y^p_{t+1} + y^q_{t+1} - \sum_{i=0}^{\infty} \theta(1-\theta)^i \left(y^p_{t-i} + y^q_{t-i}\right).$$

Grouping all the transitory terms into one expression and all the first differences of the random walk component into another expression and rearranging

$$u_{t+1} = Q_{t+1} + P_{t+1}$$

where Q_{t+1} and P_{t+1} are given by Eq. (18) in the text. QED.

Proof that $Eu_{t+1}u_t = 0$

It is convenient to first prove the following Lemma

Lemma 1 $(1-\theta)\sigma_p^2 - \theta^2\sigma_q^2 = 0$

Proof Rearranging Eq. (4) in the text

$$\theta + \frac{a}{2} = \sqrt{a + \frac{a^2}{4}}.$$

Raising both sides of this equation to second power, cancelling terms and noting that $a \equiv \frac{\sigma_p^2}{\sigma_q^2}$

$$\theta^2 = \frac{\sigma_p^2}{\sigma_q^2}(1-\theta).$$

The proof is completed by moving σ_q^2 to the left hand side of this equation.

Since all the terms in P_{t+1} are statistically independent from the terms in Q_{t+1}

$$Eu_{t+1}u_t = EQ_{t+1}Q_t + EP_{t+1}P_t. \tag{26}$$

Using the definitions of Q_{t+1} and of P_{t+1} from Eq. (18) in the text, it can be shown after some tedious algebra that

$$EQ_{t+1}Q_t = -\frac{\theta\sigma_q^2}{2-\theta}, \tag{27}$$

$$EP_{t+1}P_t = \frac{\sigma_p^2(1-\theta)}{\theta(2-\theta)}. \tag{28}$$

Substituting those expressions into Eq. (26)

$$Eu_{t+1}u_t = \frac{(1-\theta)\sigma_p^2 - \theta^2\sigma_q^2}{\theta(2-\theta)}.$$

By Lemma 1, the numerator of this expression is zero. Since the denominator is positive $Eu_{t+1}u_t = 0$. QED

Derivation of $\rho_j(\Delta Y_t^p)$ (Eq. 20)

From Eq. (15) in the text,

$$E\left[u_{t+j+1}u_{t+j}|\Delta y_t^p\right] = EQ_{t+1}Q_t + E\left[P_{t+j+1}P_{t+j}|\Delta y_t^p\right] \tag{29}$$

where

$$\begin{aligned} E\left[P_{t+j+1}P_{t+j}|\Delta y_t^p\right] = & E\left\{\Delta y_{t+j+1}^p + (1-\theta)\Delta y_{t+j}^p + ..\right\} \\ & \left\{\Delta y_{t+j}^p + (1-\theta)\Delta y_{t+j-1}^p + ..\right\} \\ & + (1-\theta)^{2j+1}\left\{\left(\Delta y_t^p\right)^2 - \sigma_p^2\right\} \end{aligned} \tag{30}$$

Taking the expected value of the product in Eq. (30), summing up the resulting infinite series and rearranging this equation reduces to

$$E\left[P_{t+j+1}P_{t+j}|\Delta y_t^p\right] = \frac{(1-\theta)\sigma_p^2}{\theta(2-\theta)} + (1-\theta)^{2j+1}\left\{\left(\Delta y_t^p\right)^2 - \sigma_p^2\right\}. \tag{31}$$

Substituting Eqs. (28) and (30) into Eq. (29), rearranging and using Lemma 1

$$E\left[u_{t+j+1}u_{t+j}|\Delta y_t^p\right] = (1-\theta)^{2j+1}\left\{\left(\Delta y_t^p\right)^2 - \sigma_p^2\right\}. \tag{32}$$

From Eq. (17) in the text,

$$Eu_t^2 = EQ_t^2 + EP_t^2 \text{ for all } t. \tag{33}$$

Using the expressions for Q_t and P_t from Eq. (18) in Eq. (33), taking expectations of the resulting expressions, rearranging and using Lemma 1 yields

$$Eu_t^2 = \frac{\sigma_q^2}{1-\theta} \text{for all } t.$$

Hence

$$\sqrt{E\left(u_{t+j+1}\right)^2 E\left(u_{t+j}\right)^2} = \frac{\sigma_q^2}{1-\theta}. \tag{34}$$

Equations (32) and Eq. (34) imply that

$$\rho_j\left(\Delta y_t^p\right) \equiv \frac{E\left\{u_{t+j+1}, u_{t+j}|\Delta y_t^p\right\}}{\sqrt{E\left(u_{t+j+1}\right)^2 E\left(u_{t+j}\right)^2}} = (1-\theta)^{2(j+1)}\left\{\frac{\left(\Delta y_t^p\right)^2}{\sigma_q^2} - \frac{\sigma_p^2}{\sigma_q^2}\right\}$$

QED

Derivation of Equation (23)

The proof is an immediate consequence of Lemma 1. QED

Proof that Using Observations on $\mathbf{y_t}$ *and* $\mathbf{y_t^e}$ *or on* $\boldsymbol{\pi_t}$ *and* $\boldsymbol{\pi_t^e}$ *Yield Identical Estimates of* $\boldsymbol{\theta}$ *and of* $\boldsymbol{\sigma^2_{\Delta y_t^p}}$

When the pair $\{\pi_t, \pi_t^e\}$ is used the estimate of θ is obtained by running the regression

$$\pi_t^e - \pi_{t-1}^e = \theta\left(\pi_t - \pi_{t-1}^e\right).$$

When the pair $\{y_t, y_t^e\}$ is used the estimate of θ is obtained by running the regression

$$y_t^e - y_{t-1}^e = \theta\left(y_t - y_{t-1}^e\right).$$

The definitions of $\{y_t, y_t^e\}$ in Eq. (25) in the text imply that the first and the second equations are identical so the estimate of θ obtained from either equation is the same.

When π_t is used to estimate $\sigma^2_{\Delta y} t$ the estimate is the sample variance of $\pi_t - \pi_{t-1}$ and when y_t is used it is the sample variance of $y_t - y_{t-1}$. Since the definitions in Eq. (25) imply

$$y_t - y_{t-1} = \pi_t - \pi_{t-1}$$

the two estimates are identical. QED

References

Barro, R. J., & Gordon, D. B. (1983). A positive theory of monetary policy in a natural rate model. *Journal of Political Economy, 91*(4), 589–610.

Brunner, K., Cukierman, A., & Meltzer, A. H. (1980). Stagflation, persistent unemployment and the permanence of economic shocks. *Journal of Monetary Economics, 6,* 467–492.

Bruno, M., & Piterman, S. (1988). Israel's stabilization: A two-years review. In M. Bruno, G. Di Tella, R. Dornbusch, & S. Fischer (Eds.), *Inflation stabilization: The experience of Israel, Argentina, Brazil, Bolivia and Mexico*. Cambridge MA and London: The MIT Press.

Cagan, P. D. (1956). The monetary dynamics of hyper-inflation. In M. Friedman (Ed.), *Studies in the quantity theory of money*. Chicago: University of Chicago Press.

Chinn, M. D. (2009). Forward premium puzzle. In K. A. Reinert & R. S. Rajan (Eds.), *Princeton encyclopedia of the world economy*. Princeton and Oxford: Princeton University Press.

Cukierman, A. (1982). Relative price variability, inflation and the allocative efficiency of the price system. *Journal of Monetary Economics, 9,* 131–162.

Cukierman, A. (1988). The end of the high Israeli inflation: An experiment in heterodox stabilization. In M. Bruno, G. Di Tella, R. Dornbusch, & S. Fischer (Eds.), *Inflation stabilization: The experience of Israel, Argentina, Brazil, Bolivia and Mexico*. Cambridge MA and London: The MIT Press.

Cukierman, A., & Melnick, R. (2015). The conquest of Israeli inflation and current policy dilemmas. In A. Offenbacher (Ed.), *Maintaining price stability: The Bank of Israel's sixth decade* (pp. 13–62).

Cukierman, A., & Meltzer, A. H. (1982). What do tests of market efficiency show? Unpublished manuscript, Carnegie-Mellon University.

Engel, C. (1996). The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance, 3,* 123–192.

Fama, E. F. (1975). Short-term interest rates as predictors of inflation. *The American Economic Review, 65*(3), 269–282.

Fama, E. F. (1984). Forward and spot exchange rates. *Journal of Monetary Economics, 14,* 319–338.

Figlewski, S., & Wachtel, P. (1981). The formation of inflationary expectations. *The Review of Economics and Statistics, 63*(1), 1–10.

Fisher, I. (1911). *The purchasing power of money.* Macmillan, New York (latest edition, A.M. Kelley, New York, 1963).

Fisher, I. (1930). *The theory of interest*. New York: Macmillan.

Frankel, J., & Poonawala, J. (2006). *The forward market in emerging currencies: Less biased than in major currencies*. NBER Working Paper Series 12496.

Frenkel, J. A. (1977). The forward exchange rate, expectations, and the demand for money: The German hyperinflation. *The American Economic Review, 67*(4), 653–670.

Frenkel, J. A. (1979). Further evidence on expectations and the demand for money during the German hyperinflation. *Journal of Monetary Economics, 5,* 81–96.

Friedman, M. (1957). *A theory of the consumption function*. Princeton: Princeton University Press.
Friedman, B. M. (1979). Optimal expectations and the extreme information assumptions of 'rational expectations' macromodels. *Journal of Monetary Economics, 5,* 23–42.
Hamburger, M. J., & Platt, E. N. (1975). The expectations hypothesis and the efficiency of the trasury bill market. *The Review of Economics and Statistics, 57*(2), 190–199.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering, 85*(Series D), 35–45.
Ljungqvist, L., & Sargent, T. J. (2000). *Recursive macroeconomic theory*. Cambridge MA and London: The MIT Press.
Lucas, R. E. (1972). Econometric testing of the natural rate hypothesis. In O. Eckstein (Ed.), *The econometrics of price determination*. Washington D.C.: Board of Governors of the Federal Reserve System.
Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review, 63*(3), 326–334.
Lucas, R. E., & Rapping, L. A. (1969). Real wages, employment and inflation. *Journal of Political Economy, 77*(5), 721–754.
Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association, 55*(290), 299–306.
Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29*(3), 315–335.