Exploring the Intensive and Extensive Margins of World Trade

Gabriel J. Felbermayr and **Wilhelm Kohler**

University of Tübingen

Abstract: World trade evolves at two margins. Where a bilateral trading relationship already exists it may increase through time (intensive margin). But trade may also increase if a trading bilateral relationship is newly established between countries that have not traded with each other in the past (extensive margin). We provide an empirical dissection of post–World War II growth in manufacturing world trade along these two margins. We propose a "corner-solutions version" of the gravity model to explain movements on both margins. A Tobit estimation of this model resolves the so-called "distance puzzle". It also finds more convincing evidence than recent literature that WTO-membership enhances trade. JEL no. F12, F15

Keywords: Bilateral trade; globalization; gravity model

1 Introduction

Despite the unquestionable increase in economic globalization, world trade still covers a surprisingly small part of the world. In 1950, almost 52 percent of the potential number of bilateral trade relationships did not report any manufacturing trade at all. By 1997 the share of bilateral trade relationships that were actually utilized was still no more than 58 percent.[1] Globalization thus evolves along two major margins. At the *intensive* margin, established bilateral trade relationships change their trade volume, while at the *extensive*

Remark: We wish to thank Rudi Winter-Ebmer and participants of the CESifo Venice Summer Institute Workshop on "Dissecting Globalization", July 2004 and at the joint Munich/Tübingen Internationalization Workshop at the IAW 2005, for stimulating discussions. Thanks are also due to Alan Deardorff, Michael Funke, Keith Head, Andy Rose, Farid Toubal, and John Whalley, as well as to an anonymous referee for helpful comments. Marcel Smolka has provided valuable editorial assistance. Please address correspondence to Wilhelm Kohler, Eberhard Karls University Tübingen, Nauklerstrasse 47, D-72074 Tübingen, Germany; e-mail: wilhelm.kohler@uni-tuebingen.de

[1] These numbers are from a data set to be described in detail below. Admittedly, the ratios would look much different if trading relationships were weighted by GDPs or population. The most appropriate weight to use is the amount of trade involved. We shall return to this in Section 3.

DOI: 10.1007/s10290-006-0087-3

margin new trade relationships are established, or existing ones abandoned. It is somewhat surprising that systematic treatments of this two-fold margin in the empirical literature have only just begun to emerge, Feenstra and Rose (1997) being a notable exception.[2] In gravity studies of trade, the usual approach is to restrict attention to those country pairs for which strictly positive trade flows are observed. This seems inadequate for two reasons. First, it ignores an important part of the "action" across time. And secondly, given the coexistence of the two margins, the exact interpretation of estimates obtained with this procedure is questionable, as are their statistical properties.

The purpose of this paper is two-fold. First, we introduce a vintage-accounting framework to quantify the importance of the extensive and the intensive margin in the growth of world trade since World War II. In doing so, we separate two forms in which the extensive margin may arise: through the formation of new countries, and through first-time establishment of trade relationships between preexisting countries. The second purpose is to reformulate the gravity equation to take explicit account of the dual margin of world trade. The extensive margin will appear in what we call a corner solution of our generalized gravity model. We discuss econometric implications of this model and then present a consistent estimation of this model, based on a comprehensive panel data set for post–World War II world trade in manufactures.

Among other things, our approach allows us to readdress the so-called "puzzling persistence of distance", i.e., recent econometric evidence from the gravity equation suggesting that the elasticity of bilateral trade with respect to distance has increased (in absolute value) over time. This evidence, surveyed by Disdier and Head (2004), seems at odds with received wisdom and extraneous evidence that highlights improvements in transport and communication as a key force of globalization. Buch et al. (2004) argue against the notion of a "distance puzzle" in which globalization may work as much through affecting the intercept of the gravity equation, as through a change in the estimated distance elasticity. However, many things other than transport technology may affect the intercept. At least, an increasing (absolute value of the) distance elasticity constitutes a puzzle in which it implies that technical progress has been biased toward short distances, controlling for potentially confounding factors such as membership in regional

[2] See also Wang and Winters (1992), Evenett and Venables (2002), and Haveman and Hummels (2004).

trade agreements. This fact is difficult to reconcile with the largely undisputed fact that the past five decades have witnessed enormous progress in air and sea transport and long-distance communication.

We argue that the distance puzzle may simply reflect a mis-specification of the gravity equation that arises from inadequate treatment of the aforementioned dual margin of world trade. More specifically, we show that in the corner-solutions version of the gravity model the usual linear estimator as well as the nonlinear least squares estimator imply a mis-specification bias that shows up in the form of the distance puzzle. More importantly, the conventional approach of estimating the gravity equation does not disentangle the extensive and the intensive margins of world trade. As a result, the coefficient estimates are devoid of a clear theoretical interpretation. Based on our corner-solutions model, we achieve such a disentangling by means of a Tobit estimation approach. In addition to resolving the distance puzzle, our results also indicate that WTO membership has been more conducive to trade than would appear from previous evidence presented by Rose (2004).

Our paper is closely related to Helpman, Melitz and Rubinstein, henceforth called HMR (2004). However, there are several important differences. First, our novel vintage-accounting framework allows for a complete decomposition of the growth of world trade into movements on the intensive and the extensive margins, including an important dissection of the extensive margin into what we call the extensive margin proper, and a "pseudo-extensive" margin which arises from the emergence of new countries. We thus provide a richer dissection of world trade than do HMR. Secondly, while HMR treat zeros in their trade data as missing values, applying a Heckman sample selection procedure, we take the zeros at face value. In particular, we argue that observed zeros contain valuable information which should be exploited for efficient estimation, and we do so using a Tobit approach to estimate our corner-solutions model.[3] We do acknowledge that the zeros may also reflect mis-reporting and mis-measurement, particularly that of small and poor countries. But when confronted with zeros in the trade data, simply treating them as missing values is hardly less arbitrary than our choice of treating them as a corner solution. While our approach poses econometric problems in that the log of zero is not defined, it does allow us to extract

[3] While formally closely related to the Tobit model, the Heckman procedure addresses the problem of nonrandom selection of some country pairs with positive *or* zero trade volumes into the sample. The corner-solutions model, in contrast, starts from the assumption that all relevant data is observed, and that the problem is not one of sample selection but rather of how to deal with the censoring of the dependent variable at zero.

more information from the data, particularly relating to the role of distance and other variables affecting the extensive margin of world trade. Finally, in contrast to our paper, HMR do not address the time-varying nature of the distance coefficient.

The importance of extensive versus intensive margins in understanding the evolution of international trade volumes has been recognized in several recent studies. Fundamentally, trade growth can be decomposed in several ways, depending on the researcher's interest and the nature of data available. In the present study, we use aggregate data with countries as units of observation, distinguishing between changes in the number of active bilateral trade relationships (extensive margin), and the growth of trade volumes in existing relationships (intensive margin). Bernard et al. (2006) focus on firm-level data and decompose the growth in U.S. exports into entry and exit, respectively, of firms into and out of exporting (extensive margin), and changes in foreign sales that exporting firms achieve (intensive margin). In a similar vein, Hummels and Klenow (2005) draw on highly disaggregated product-level data to distinguish between the variety dimension of U.S. trade growth (extensive margin), the quality (price) and the quantity dimension (intensive margins). Another popular way of looking at the evolution of world trade is to distinguish between trade in final and intermediate goods, as emphasized by Yi (2003). In this context, the intensive margin is related to changes in trade volumes, based on a given pattern of vertical specialization, while the extensive margin addresses the question of whether a certain input is traded internationally or not. In each of these cases, trade costs play an important role in the story. Bernard et al. (2006) emphasize that reductions in trade costs may boost industry productivity, even without intra-firm productivity gains, through entry and exit at the extensive margin alone. Hummels and Klenow (2005) point out that adverse terms-of-trade effects are much less likely if a country's trade grows at the extensive product margin, increasing the range of products traded rather than the volume of trade within a given range. And finally, Yi (2003) demonstrates that an increase in vertical specialization at the extensive margin of intermediate inputs trade may explain the large and nonlinear responsiveness of trade volumes with respect to reductions in trade barriers. Interestingly, the studies cited above refrain from using corner-solutions models in their econometric analyses, looking at each margin separately. Hence, this paper provides a methodological contribution that extends beyond the role of distance as a trade barrier on the country-level extensive margin that we focus on.

The paper is structured as follows. Section 2 discusses the methodological problem underlying the distance puzzle and gives intuitive reasons why a consistent incorporation of the dual margin of world trade should be an important part of the solution. Section 3 provides an in-depth dissection of world trade growth from 1950 to 1997, documenting the relative importance of the extensive and the intensive margins of globalization. Section 4 introduces the corner-solutions gravity model and discusses the relevant econometric issues leading to the Tobit approach. Section 5 presents empirical results from a Tobit estimation, and Section 6 summarizes the results and their implications for future research.

2 A Methodological View on the Distance Puzzle

Fundamentally, the distance puzzle may be seen as a reflection of missing trade, meaning that observed trade through time increasingly falls short of what early estimates of the gravity model would predict, based on the evolution of time-varying determinants of trade, like income. Provided these estimates are accepted on face value, missing trade gets picked up by an increasing barrier effect attributed to time-invariant distance. If there is a common trend in missing trade and income, then the income coefficient might pick up part of it, if allowed to be time-variant and estimated jointly with the distance coefficient, thus taking some of the explanatory burden off the distance variable. However, if we still estimate an increasing role of distance through time, and if we have sufficient extraneous evidence on a falling time trend of transport costs, then the simple verdict is that gravity theory performs badly, and increasingly so over time. The challenge then is to find the “missing factors” reflected in seemingly missing trade. Any reformulation of the theory that takes the explanatory burden from missing trade off the distance variable would be seen as an improvement.

Why should we expect the dual margin of trade to play an important role in this attempt? The intuition runs as follows. Gravity theory maintains that trade is the result of mass attraction and resistance from geographical distance. If attraction in some cases is not strong enough to generate trade at all, then ignoring such cases altogether implies that we systematically overestimate the force of attraction, or—equivalently—underestimate the trade-inhibiting force of distance. Now suppose that the prevalence of such “zero trade” cases falls through time, say because of technological improvements in transport and communication. Then, the bias toward overestimation of attraction and underestimation of distance falls through time. The distance puzzle may thus be a reflection of an ever smaller underestimation of the distance barrier, as world trade expands on the extensive margin. However, it is not clear a priori to what extent the bias (and its change through time) arises with the income coefficient, or the distance coefficient. The answer requires empirical analysis of the dual margin.[4]

There are, of course, other factors that might play a role. An argument often made relates to trade liberalization. For large distances, tariffs are generally a relatively small part of overall trade costs. If tariffs are equal for all trade relationships to start with, and if they are reduced by an equal (absolute) amount, then—other things equal—the percentage reduction in the destination price is larger for low-distance trade. Tariff liberalization will then have a disproportionately large impact on short-distance trade. In the gravity model, this might be picked up by an increasing role of distance as a trade-inhibiting factor through time. However, this route of explanation hinges on an equal level of tariffs to start with, and on an equal amount of tariff liberalization across all trade relationships, which seems questionable. In any case the effect is easily controlled for in the empirical analysis pursued in this paper.

A further point relates to the degree of product differentiation and substitutability. If goods from different locations are distant substitutes, then increasing distance might not reduce trade by much. If they become closer substitutes, then distance plays an ever increasing role, until—in the limit with perfect substitution—trade in any good occurs only with the closest location. The distance puzzle could thus be due to a long-run decline in the importance of product differentiation in traded goods. However, this explanation seems questionable, since it would appear at least as plausible that the role of product differentiation in trade has increased, rather than declined over recent decades. Moreover, if valid, the argument implies that the empirical performance of the gravity equation would fall through time. If anything, the empirical literature points to the opposite.

A final argument relates to foreign direct investment (FDI). If trade and FDI are substitutes, and if during the past decades FDI has systematically replaced long distance trade more than short distance trade, then

[4] There is an additional motivation for looking at the extensive margin which has to do with welfare. If trade is based on product differentiation along the lines of the Dixit–Stiglitz approach, as suggested by gravity theory (see below), then expansion of world trade on the extensive margin seems particularly important from a welfare perspective, since it increases the degree of product differentiation.

this effect might get picked up by a rising estimated distance coefficient in a gravity equation. However, the relationship between trade and FDI is far from clear cut, both theoretically and empirically. The same holds true for the role of distance in FDI. While there is no doubt that distance and trade costs play a role for FDI, particularly through the well-known proximity-concentration trade-off, it is not at all clear why FDI should be more attractive for long-distance markets. Market-size and monitoring costs play a role as well. Moreover, FDI may be used as a platform to serve other markets, in which case it is complementary to trade (Neary 2005). Empirically, distance appears to play an ambiguous role for FDI, which would cast additional doubt on whether it is a promising candidate for resolving the distance puzzle.[5]

Overall, our priors are that the distance puzzle establishes a convincing case for exploring the dual margin in world trade, and to extend the gravity model in such a way that it consistently captures simultaneous movements on both margins through time. We therefore move on to a detailed empirical account of such movements, followed by a reformulation of the gravity equation, including a discussion of the econometric issues arising from the dual margin, and empirical panel estimation, using a Tobit estimation approach.

3 Decomposing the World Trade Expansion

Available data sources treat any country pair for which there is no reported trade as a missing observation. This must be distinguished from explicit recording of zero trade. Unfortunately, data based on explicit reporting of zero trade are not available. However, an extensive scrutiny of the IMF Direction of Trade Statistics (DoTS) by Gleditsch (2002), based on a detailed comparison with other data (e.g., UN COMTRADE and WTO), shows that 80 percent of all observations coded as missing do in fact represent zeros. In this paper, we therefore rely on the DoTS and, following Coe et al. (2002) and Brun et al. (2004), we treat missing observations as zero trade. However, while the evidence provided by Gleditch does provide some reassurance, this strategy admittedly involves a measurement problem. We shall return to it when discussing econometric issues below.

[5] The role of distance for FDI has been examined, for instance, by Markusen (2002: ch.10) and by Egger and Pfaffermayr (2004), with ambiguous results.

Import data are usually more reliable than export data since imports constitute a tax base, while no comparable incentive for correct reporting exists on the export side. This is also reflected by a larger number of country pairs with positive trade, if bilateral trade data are constructed from import data alone. Better reliability has prompted some researchers to construct trade data from imports alone; see Coe et al. (2002) and Brun et al. (2004). But imports are evaluated c.i.f., including costs of transport and insurance. Hence, using such data in a gravity regression yields an inconsistent estimator for the distance coefficient, as distance will be correlated with the error term. Using export data may involve a cost in terms of larger errors, but avoids this correlation and thus yields consistent estimates. In this paper, we follow Rose (2004) in using an average of c.i.f. import and f.o.b. export values, in order to obtain a maximum number of observations.[6] In the sequel, T_{ij} is defined as the sum of bilateral exports and imports recorded by countries i and j, divided by 4.[7] The data cover 1950–1997 trade in manufactures and are in constant U.S. dollars, based on the U.S. CPI $(1983 = 100)$.

We define the vintage of a trade relationship as the earliest time at which trade may occur between a specific pair of two countries, based on a) their independent existence and b) their principal openness. A country is judged open if it reports trade with at least one other country. Of course, an open country need not trade with all other open economies. We use $N_{t,h}$ to denote the number of trade relationships of vintage h that are active at time t. This will typically be lower than the number of potential vintage-h trade relationships at time t, which is denoted by $V_{t,h}$. Total world trade at time t may thus be written as $T_t \equiv \sum_{h=t_0}^{t} \overline{T}_{t,h} N_{t,h} \equiv \overline{T}_t \sum_{h=t_0}^{t} n_{t,h} \zeta_{t,h} V_t$, where $n_{t,h} \equiv \frac{N_{t,h}}{V_{t,h}}$ is defined as the share of active trading relationships within vintage h, and $\zeta_{t,h} \equiv V_{t,h} \big/ \sum_{h=t_0}^{t} V_{t,h}$ is defined as the share of vintage h in the potential number of trading relationships. $\overline{T}_{t,h}$ is the average bilateral trade volume based on trading relationships of vintage h, and $\overline{T}_t$ is the average volume of bilateral trade at time t across all vintages: $\overline{T}_t \equiv \sum_{h=t_0}^{t} \overline{T}_{t,h} \big/ \sum_{h=t_0}^{t} N_{t,h}$. Finally, $V_t \equiv \sum_{h=t_0}^{t} V_{t,h}$ denotes the overall number of potential trading relationships, and t_0 denotes the "beginning of time". Obviously, $h \leq t$.

[6] We use the data kindly made available by Andy Rose on his website. See http://faculty.haas.berkeley.edu/arose/RecRes.htm. Choosing a sample period from 1950 to 1997 avoids incomplete recordings at the beginning and toward the end.

[7] Where only 3 (or 2) observations on bilateral trade are available, their sum is divided by 3 (or 2).

We now call

$$\Delta T_{\mathrm{int},t} \equiv \textstyle\sum_{h=t_0}^{t-1} \left(\overline{T}_{t,h} N_{t-1,h} - \overline{T}_{t-1,h} N_{t-1,h}\right) \tag{1}$$

a movement of world trade on the intensive margin, where preexisting relationships vary in trade volumes. Accordingly, variations in the number of active trading relationships,

$$\Delta N_t \equiv \textstyle\sum_{h=t_0}^{t-1} (N_{t,h} - N_{t-1,h}) + n_{t,t} \Delta V_t = \Delta N_{\mathrm{x},t} + n_{t,t} V_{t,t} \,, \tag{2}$$

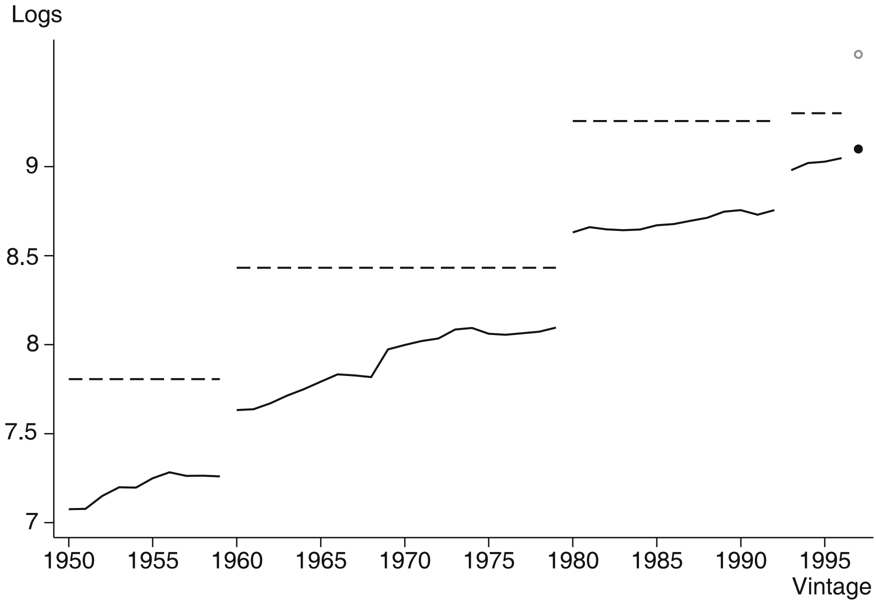
are called movements on the extensive margin of world trade, where $\Delta V_t \equiv V_t - V_{t-1}$ is the change in the number of potential trade relationships, due to the formation of new countries or disappearance of old ones. If no old countries disappear, then $\Delta V_t \equiv V_{t,t}$, as assumed in the above equation. Equation (2) separates two types of extensive margins. At the extensive margin proper, denoted by $\Delta N_{\mathrm{x},t}$, the utilization of potential trade relationships between preexisting countries and vintages $h \leq t-1$ changes from $t-1$ to t. By way of contrast, ΔV_t captures the first-time emergence of new trading partners, and disappearance of existing ones. This is certainly exogenous to gravity theory. We therefore call ΔV_t the pseudo-extensive margin of world trade. But the extent to which new potential trade relationships become active at once, i.e., the term $n_{t,t}$, is treated as endogenous below.

Defining $N_t \equiv \sum_{h=t_0}^{t} n_{th} \zeta_{th} V_t$, world trade growth may now be decomposed as follows:

$$\begin{aligned} \Delta T_t \equiv \Delta \overline{T}_t N_{t-1} + \Delta N_t \overline{T}_t \equiv \Delta \overline{T}_t N_{t-1} \\ + (\Delta N_{\mathrm{x},t} + n_{t,t} \Delta V_t) \overline{T}_t \,, \end{aligned} \tag{3}$$

where ΔN_t is taken from (2). With this decomposition, changes on the extensive margin are "weighted" by end-of-period average trade volumes, while changes on the intensive margin are "weighted" by beginning-of-period numbers of trade relationships. The decomposition could equally well be defined with opposite weighting. Indeed, one could construct more complex weighting schemes that maintain the identity between observed trade growth and the component terms. However, the component terms are used only for descriptive purposes and do not enter the estimation procedure below. Hence, the weighting chosen in (3), while in some sense arbitrary, does not give rise to a measurement bias. But in this descriptive section it is still important to be aware of the arbitrariness of the decomposition chosen, and to duly recognize its implications when interpreting the numbers; see below.

Figure 1: *The Extensive Margin of Merchandize World Trade (Looking at Vintages up to 1950 / 1960 / 1980 / 1993 / 1997)*



Figures 1 through 3 provide descriptive evidence on the role that these margins have played in post–World War II evolution of world trade in manufactures. Figure 1 highlights the extensive margin, looking at the increasing utilization over time t of potential trading relationships, singling out five different groups of vintages. The left-most horizontal line gives $\sum_{h=t_0}^{1950} V_{t,h}$, while the line below gives $\sum_{h=t_0}^{1950} N_{t,h}$ for t up to 1959. Note the difference between t_0 ("beginning of time") and the beginning of the sample period which is 1950. The corresponding lines for $1960 \leq t \leq 1979$ depict $\sum_{h=t_0}^{60} V_{t,h}$ and $\sum_{h=t_0}^{60} N_{t,h}$, respectively, and analogously for vintages up to 1980 and 1993 further to the right.[8] These lines do not fully trace out the evolution of world trade at the extensive margin. Indeed, they do not even look at trade as such, but simply count trading relationships for four cumulative groups of vintages, in order to illustrate the type of movement that occurs at the extensive margin. The jumps in the horizontal lines indicate

[8] These points in time have been chosen primarily to illustrate the definitions of the extensive and the pseudo-extensive margin. A more continuous way of tracing the pseudo-extensive margin in the evolution of world trade is presented in Figures 3a and 3b below.

movements on the pseudo-extensive margin when moving from one group to the next, while the gap between the two lines at any point in time reveals the extent to which potential trading relationships of the respective groups of vintages have not yet become active.

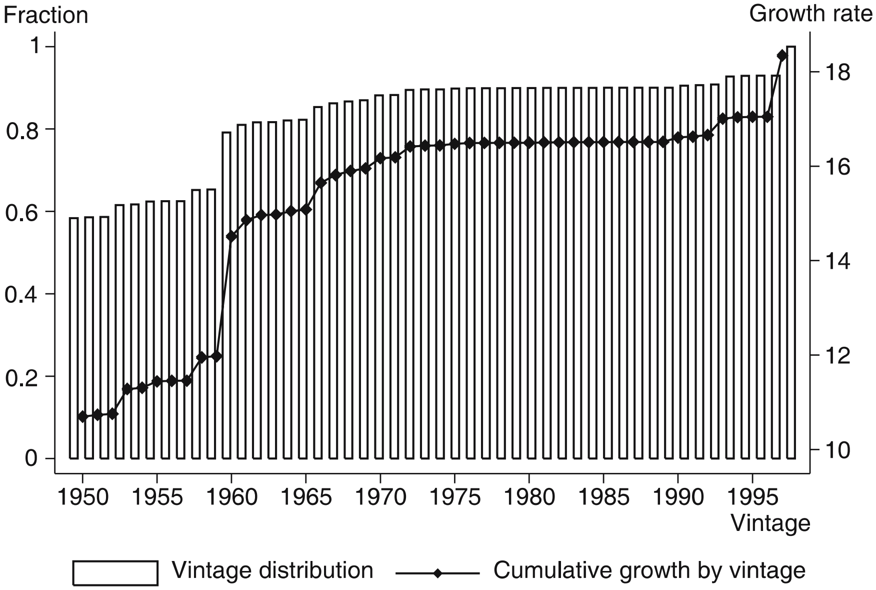
Thus, of the potential trading relationships based on the group of vintages up to 1950, about 40 percent had not yet become active by 1950 (the numbers are in natural logs). By 1959, the active relationships within this group have grown by about 20 percent, narrowing the gap of unutilized relationships to 30 percent. Including new vintages emerging between 1950 and 1959, mainly from colonies becoming independent, the maximum number of trading relationships increases by about 65 percent, as indicated by the second horizontal line starting at 1960. However, the number of active relationships increases only by about 40 percent (comparing the solid lines between 1959 and 1960). Hence, the gap of unutilized relationships within this larger group of vintages (comparing the dashed and solid line) is still 43 percent in 1960, but is reduced substantially in the next two decades. Analogous interpretations hold for the subsequent groups of vintages, including $1960 \leq h \leq 1979$ and $1980 \leq h \leq 1992$. Notice the relatively large jump at the sample end of 1997, reflecting the emergence of new independent states from disintegration of the former Soviet Union and Yugoslavia.[9] Obviously, these are institutional shocks driven by forces other than those of the intensive margin. Our statistical model therefore does not purport to explain ΔV_t. On the other hand, the number of new vintages that are immediately active, $n_{t,t}$, is treated as an endogenous variable.[10] In terms of Figure 1, what we want to explore is the time-varying role of distance in explaining the gap between the horizontal and the solid line. Disintegration of countries is also likely to increase the share of low-distance trade relationships. But, as will be seen below, such a change in the sample composition does not, in and of itself, generate any estimation problem regarding the distance coefficient.

Figure 1 clearly shows that during our sample period there was significant change at both types of extensive margins, and there is still much room for further movements, even at the end of the sample period in 1997. It does not, however, tell us how much trade was involved in movements at the extensive margin. One would probably expect the amount of trade generated

[9] It should be emphasized that COMECON countries did not belong to the sample before opening up and disintegration.

[10] Referring to the world trade matrix, $\Delta V_t > 0$ amounts to adding new rows and columns, while $n_{t,t}$ gives the share of new cells emerging in “filled form”.

Figure 2: *Growth of Merchandize World Trade*
(Vintage Distribution of Cumulative Growth 1950–1997)



at this margin to be rather small.[11] Figure 2 therefore moves to a somewhat more comprehensive perspective by asking a very simple question: What is the contribution of vintages up to H toward the overall cumulative growth in world trade between 1950 and 1997? The bars indicate the frequency distribution, plotting the ratio of $\sum_{h=t_0}^{H} \overline{T}_{1997,h}N_{1997,h} - \sum_{h=t_0}^{1950} \overline{T}_{1950,h}N_{1950,h}$ to $\sum_{h=t_0}^{1997} \overline{T}_{1997,h}N_{1997,h} - \sum_{h=t_0}^{1950} \overline{T}_{1950,h}N_{1950,h}$ on the left-hand vertical axis, with H going from 1950 to 1997 on the horizontal axis. Using the right-hand axis, the line depicts the growth rate for trade of vintages up to H, i.e., $\sum_{h=t_0}^{H} \overline{T}_{1997,h}N_{1997,h} \Big/ \sum_{h=t_0}^{1950} \overline{T}_{1950,h}N_{1950,h} - 1$ between 1950 and 1997 of vintages up to H, i.e., $\sum_{h=t_0}^{H} \overline{T}_{1997,h}N_{1997,h} \Big/ \sum_{h=t_0}^{1950} \overline{T}_{1950,h}N_{1950,h} - 1$.

Figure 2 is perhaps best understood by looking at extreme cases. If all growth had happened entirely at the intensive margin, then the distribution function would be degenerate, with all mass concentrated at vintages up

[11] The working paper version of this paper Felbermayr and Kohler (2004) presents a view on the amount of trade generated through movements at the extensive margin, relative to the intensive margin, for the five vintage groupings depicted in Figure 1.

to $H = 1950$. If growth had evolved in a completely symmetric way at the extensive margin only, then we would have a linear increase of the bars, and a straight line for the growth rates. Vertical jumps at interior points indicate movements at the extensive margin (proper plus pseudo), while flat segments indicate prevalence of the intensive margin. For example, based on vintages up to 1950 the growth rate of world trade was about 1.100 percent, contributing about 60 percent to cumulative growth during the entire sample period. Including vintages 1950–1960, the growth rate increases to about 1.420 percent, contributing a further 20 percent to overall cumulative growth. Cumulative growth of trade from 1950 to 1997 involves both, episodes where the contribution was more important on the extensive margin (late 1950s and early 1960s, as well as the 1990s) and an interim period dominated by the intensive margin.

Finally, Figures 3a and 3b depict a decomposition of world trade growth into its constituent parts according to equation (3) above. Plotting $t = 1959, ..., 1997$ on the horizontal axis, Figure 3a reveals how the cumulative difference of actual trade from the initial volume for 1950, indicated by bars, may be decomposed into changes at the respective mar-

Figure 3a: *Decomposing the Evolution of Merchandize World Trade (Cumulative Difference to 1950, in Trillion Real Dollars)*

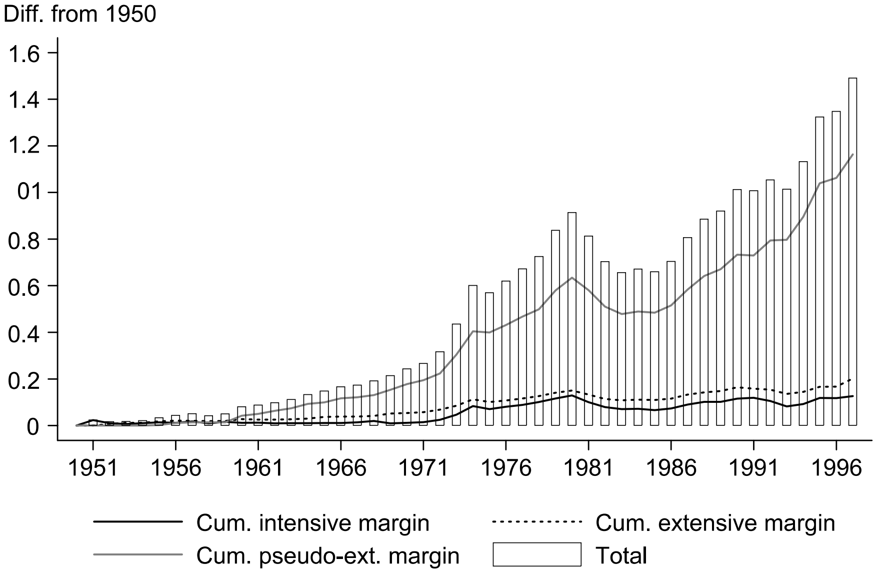
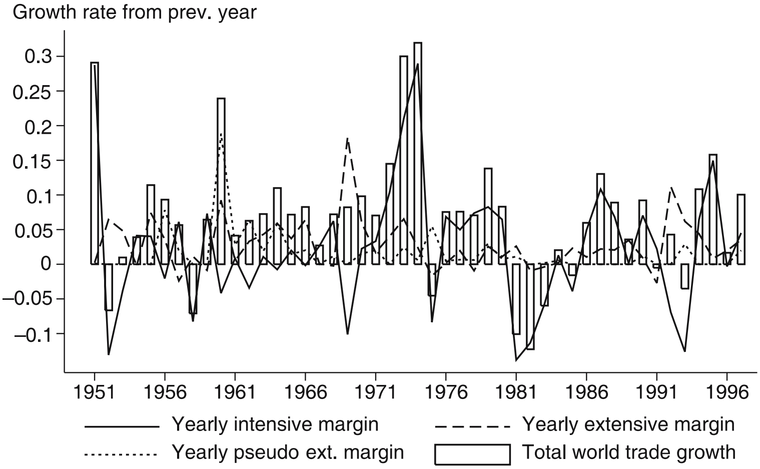


Figure 3b: *Decomposing Yearly Growth of Merchandize World Trade (Yearly Growth Rates 1950–1997)*



gins, indicated by lines. The intensive margin line depicts $(\overline{T}_t - \overline{T}_{1950}) \sum_{h=t_0}^{1950} N_{1950,h}$, while the line for the extensive margin proper plots $\overline{T}_t \left(\sum_{h=t_0}^{1950} N_{t,h} - \sum_{h=t_0}^{1950} N_{1950,h}\right)$, and the pseudo-extensive margin is plotted as $\overline{T}_t \left(\sum_{h=t_0}^{t} N_{t,h} - \sum_{h=t_0}^{1950} N_{1950,h}\right)$. By definition, the three lines add up to the bars. Figure 3b plots the yearly growth rates. Extreme values of yearly growth tend to be dominated by the intensive margin, the exception being the period from the late 1950s to the late 1960s, as well as the early 1990s. But even for the 1970s and 1980s, the extensive margin did play a role, as evidenced by the difference between the bar-values and the solid line.

Figure 3a indicates a surprisingly large contribution of the extensive margin, relative to Figure 2. One should, however, be aware of the specific decomposition chosen in (3). For the extensive margin, the contribution to the growth of trade is weighted by the current average trade volume $\overline{T}_t$, while the intensive margin receives base-period weights from the 1950 number of active trading relationships, i.e., $\sum_{h=t_0}^{1950} N_{1950,h}$. For instance, from 1950 to 1980 world trade has increased by about 0.93 trillion (real 1983) dollars. If all new trading relationships emerging and becoming active between 1950 and 1980 had been trading on the average 1980 level, $\overline{T}_{1980}$, then this margin alone (pseudo-extensive margin) would imply an increase by as much as about 0.65 trillion real dollars. Taking only vintages up to 1950 and the increase in their utilization (extensive margin proper) adds a further increase by 0.14 trillion real dollars. About that same increase is observed on the intensive margin, where we look at the trading relationships that had been active already in 1950, assuming that they have increased their trading volume according to $\overline{T}_{1980} - \overline{T}_{1950}$.

Our analysis clearly suggests that the extensive margin is important quantitatively, both along the cross-sectional and the time-series dimension of the data. First, even in the last year of the time span under investigation, only about 58 percent of potential bilateral trading relationships are actually utilized. Second, about 40 percent of world trade growth from 1950 to 1997 comes from movements at the extensive margin. Admittedly, this result is "contaminated" by the pseudo-extensive margin where new countries emerge, due to decolonialization and other fundamental changes in the political environment. But even from 1970 to 1997, where the pseudo extensive margin arguably plays a minor role in our data, about 15 percent of total world trade growth is attributable to the extensive margin. The empirical importance of the extensive margin calls for a proper modeling of the dual margin in gravity-type investigations of world trade. This relates to both, the conceptual framework and the econometric implications. In the following sections of the paper, we undertake such a modeling effort.

4 Modelling the Dual Margin of Globalization

Established trade theory does not go very far in explaining movements on the dual margin of world trade. In a large class of models, the focus lies squarely on countries' overall trade. As noted recently by Deardorff (2004), the comparisons are mostly done globally, and not by pairs of countries, thus leaving bilateral trade undetermined. This seems justifiable on the grounds that bilateral trade of a country is largely irrelevant for its level of welfare, but in many respects bilateral trade is important.[12]

Common sense suggests that bilateral trade is importantly determined by trade costs related to geographic distance and transport. Traditional trade

[12] A case in point is the trade potential that arises if a country opens up to world trade in the process of systemic change. Prominent examples are Eastern European countries in the 1990s.

theory almost entirely neglects all such costs, the important exception, of course, being the gravity approach where distance as a trade-inhibiting force combines with economic mass of two countries to explain bilateral trade. The gravity force arises, whenever trade is based on perfect specialization, meaning that each good (with a sufficiently narrow definition) is produced in only one country, but consumed everywhere (say due to identical preferences). Such a case arises, almost by definition, in a love-for-variety product differentiation framework with increasing returns to scale, which is the usual theoretical justification of the gravity equation going back to Anderson (1979). It also arises with comparative-advantage-based trade, provided domestic trade is treated on an equal footing with foreign trade, or if the number of goods and countries is very large, relative to the number of factors; see Deardorff (1998).

The gravity equation derived using models of monopolistic competition assumes that trade costs are of the iceberg type and that there is no fixed cost of entering a certain market. With CES preferences, this model cannot account for zero trade. By way of contrast, our corner-solution approach to gravity importantly hinges on zero trade as an equilibrium outcome. Theoretical foundations of the gravity equation that allow for zero trade have been provided by Feenstra et al. (2001), Eaton and Kortum (2002), Haveman and Hummels (2004), and more recently Helpman et al. (2004). In this paper, we abstain from providing an explicit structural model of trade leading to an estimable gravity equation that allows for zero bilateral trade volumes. For our purposes it suffices to work with an reinterpretation of the standard gravity equation as derived, e.g., by Baier and Bergstrand (2001) and reviewed by Feenstra (2004).

Suppose that the bilateral trade potential is determined according to the conventional gravity equation. It is, thus, positive regardless of the magnitude of real trade costs. Our key assumption in this paper is that, for reasons external to each exporting firm, positive trade between two countries arises only if their bilateral trade potential exceeds some dyad-specific threshold value $\overline{T}_{ijt}$. One may think of various explanations for the existence of such a threshold value, the principal line of reasoning is as follows. Maintaining trade relationships requires certain infrastructure and institutions that facilitate an efficient flow of information and exchange of ideas, in addition to the best-practice transport of goods. Such institutions need to be present both, in the export and import country. While the private costs of information flows and transport are captured in the conventional gravity equation determining the trade potential, the underlying infrastructure and institutions often exhibit characteristics of public goods. In other words, they involve fixed costs (largely independent of the magnitude of potential trade) that are borne by the public sector. Important examples are international payment systems, legal agreements, consulates, or the activities of the respective chambers of commerce. The assumption here is that individual firms will not be able to trade in the absence of services provided by such institutions. A further key assumption is that the benefits accruing to both countries from such institutions depends on how much trade they are likely to support once established. And this is where the gravity-determined trade potential comes into play. Rational governments will invest into these institutions only if the expected benefit from a materializing trade potential in present value terms exceeds the investment cost. These will, in turn, determine the threshold value $\overline{T}_{ijt}$ that the trade potential needs to surpass for governments to invest into the required institutions.

For the present purpose, we need not provide an explicit model determining the threshold value $\overline{T}_{ijt}$, but it would seem natural to assume that it depends on geographical barriers such as distance, and on the two countries' GDP levels. It is thus specific to each dyad. Armed with these ideas, we may now modify the standard gravity equation by postulating the following system of equations

$$T^*_{ijt} = \frac{Y_{it} Y_{jt}}{p_i^\sigma \bar{Y}_t} \left(\frac{\theta_{ijt}}{P_{jt}} \right)^{1-\sigma}, \tag{4}$$

$$T_{ijt} = \frac{T^*_{ijt}}{T^*_{ijt} - \overline{T}_{ijt}} \max\left(0,\, T^*_{ijt} - \overline{T}_{ijt}\right), \tag{5}$$

where the first expression explains the trade potential T^*_{ij}. It is taken from Baier and Bergstrand (2001: 9–10), where it is derived as an approximation to a gravity equation that includes multilateral trade resistance, proposed by Anderson and van Wincoop (2003).[13] The equation is based on product differentiation on the firm level, with $\sigma > 1$ as the elasticity of substitution

[13] Our gravity equation is log-linear in distance. This is a common feature in the applied literature. However, it is worth noting that the Anderson and van Wincoop (2003) version of the gravity model is highly nonlinear in distance (and other covariates). Differentiating their system of equations, one finds that the elasticity of trade with respect to distance depends on the Helpman (1987) similarity index. Potentially, changes in this index could also explain the time pattern of the elasticity of distance. However, empirically, the index exhibits very little time variance and thus does not seem a good candidate explanation for the distance puzzle. Therefore, we do not pursue this argument any further.

between varieties. While Y_i and Y_j are total GDPs of the two countries i and j, $\bar{Y}$ denotes world-GDP, $\theta_{ij} > 1$ indicates iceberg trade costs and p_i are country i's mill (ex factory) prices and P_j is country j's exact price index, based on love-of-variety preferences. Equation (5) introduces the threshold value to relate the trade potential T^*_{ijt} to actual trade T_{ijt}, in line with the reasoning proposed above. Trade will materialize in its full potential if it exceeds the threshold level $\overline{T}_{ijt}$, otherwise bilateral trade is zero. We call this the corner solutions of the gravity model.

Specifying iceberg-type trade costs determined by geographical and cultural distance, as well as membership in trade agreements, we may write

$$\theta_{ijt} = \Theta_{ijt} D^{\delta}_{ij} B^{\beta}_{ij} L^{\gamma}_{ij} A^{\alpha}_{ijt} \,, \tag{6}$$

where D_{ij} is bilateral distance, B_{ij} is a dummy for a common border, L_{ij} is a dummy for common language, and A_{ijt} denotes joint membership in trade agreements such as the WTO or the European Union. Importantly, Θ_{ij} denotes the overall level of technology pertaining to transport or communication. Substituting and taking logs gives the following equation for the bilateral trade potential

$$\begin{aligned} \ln T^*_{ijt} = \; & (1-\sigma)\, \Theta_{ijt} + \ln\left(Y_{it} Y_{jt}\right) + (1-\sigma) \\ & \times \left(\delta \ln D_{ij} + \beta \ln B_{ij} + \gamma \ln L_{ij} + \alpha \ln A_{ijt}\right) \\ & + \phi_i + \phi_j + \phi_t + u_{ijt} \,. \end{aligned} \tag{7}$$

Note that this specification includes separate sets of dummies for countries i and j as well as time dummies to control for the unobserved multilateral resistance indexes proposed by Anderson and van Wincoop (2003). Technical progress in transport or communication technologies may affect both, the constant $(1-\sigma)\,\Theta_{ij}$ and the coefficients δ and β. However, the effect of distance on transport costs, δ, is not identified in (7). Following established literature, we make the identifying assumption that σ is constant over time. In the subsequent discussion, we shall refer to the right-hand-side variables in (7) collectively as $\mathbf{X}_{ijt}$.

As we have shown in Section 3 above, our data set features a large number zero trade cases, whereby the number of cases with positive trade increases over time. Broadening our perspective, we now need to distinguish three types of events. The first is the formation of new countries. In terms of the world trade matrix, this adds new rows and columns. Although we have discussed this under the pseudo-extensive margin above (denoted by ΔV_t

in Section 3 above), it is driven by forces entirely unrelated to the gravity model, and indeed unrelated to economics. This margin is therefore treated as exogenous in our statistical model. The second type of event occurs when changes in gravity-type variables cause empty cells to be filled over time. This is the extensive margin proper, and it includes the question of whether a new cell enters in “filled form” (denoted by $n_{t,t}$ in Section 3 above), which is therefore treated as an endogenous variable in our model. And finally, there is the traditional case where a cell entry increases or falls in magnitude, which is the intensive margin.

A separate issue relates to measurement. As we have noted above, available data sets do not include explicit reporting of zero trade, but simply code all country pairs not reporting positive trade as missing observations. This precludes a multistep procedure, where the first step would explain whether a given country pair is a missing observation, followed by a separate explanation of whether a given observation involves positive or zero trade, and a final step explaining the extent of trade if positive. The first step is not only negated by lack of data, it would also feature explanatory factors for missing observations that are unrelated to the gravity equation and not of immediate interest for the present purpose. Therefore, our strategy, supported to some extent at least also by empirical evidence (see Section 3 above), is to directly treat missing observations as zero trade.

Zero trade entries in the data pose a problem when taking logarithms. One way to deal with this problem is to estimate the gravity equation in semilogarithmic form, with T_{ijt} instead of $\ln T_{ijt}$ as the dependent variable. This is the strategy followed by Eaton and Tamura (1994). However, theory suggests $\ln T_{ijt}$ as the correct left-hand variable. In line with this, the literature so far clearly shows that the semilog equation performs considerably worse than the log-log one.[14] Hence, we follow Eichengreen and Irwin (1995, 1997), in transforming the dependent variable to $\ln\left(1 + T_{ijt}\right)$, which may be justified by the following logic of interpretation. If T_{ijt} is large, then $\ln\left(1 + T_{ijt}\right)$ is approximately equal to $\ln T_{ijt}$, thus validating the common practice of interpreting estimated coefficients as elasticities. If T_{ijt} is very small, then $\ln\left(1 + T_{ijt}\right)$ is approximately equal to T_{ijt}, in which case coefficients have to be interpreted as semielasticities. Arguably, either

[14] We have experimented with the semilog specification. Qualitatively, the main patterns in the interesting regression parameters do not change. But many coefficients are estimated imprecisely and with elasticities that take implausible magnitudes.

of the two strategies is somewhat arbitrary, making the results dependent on the unit of measurement. To check the robustness, we shall express the dependent variable as $\ln\left(a + T_{ijt}\right)$, experimenting with different values for $a > 0$.[15]

There is a key statistical problem that follows from our corner-solutions version of the gravity equation. This is that the conditional mean of *actual* trade $\ln T_{ijt}$ cannot be linear in $\mathbf{X}_{ijt}$, because there is positive probability mass at $T_{ijt} = 1$, or $\ln T_{ijt} = 0$. Addressing this problem by means of nonlinear least squares (NLS), as in Santos Silva and Tenreyro (2003) or Coe et al. (2002), poses several problems. First, since $\ln T_{ijt}$ includes corner outcomes, $\ln T_{ijt} \left| \mathbf{X}_{ijt} \right.$ is likely to be heteroskedastic, which renders NLS inefficient.[16] Using weighted NLS requires an arbitrary choice of a specific model for the conditional variance, $\text{var}\left(\ln T_{ijt} \left| \mathbf{X}_{ijt} \right.\right)$, and would thus seem questionable; see Wooldridge (2002: 518 ff). More importantly, the coefficients obtained by NLS estimation of a model for $\text{E}\left(\ln T_{ijt} \left| \mathbf{X}_{ijt} \right.\right)$ are difficult to interpret. They do not tell us anything about distribution of $\ln T_{ijt} \left| \mathbf{X}_{ijt} \right.$, other than its mean. In particular, they would not allow us to empirically identify the intensive and extensive margins of world trade. In the underlying distribution, the extensive margin is given by $\Pr\left[\ln T_{ijt} \geq 0 \left| \mathbf{X}_{ijt} \right.\right]$, while the intensive margin is given by $\text{E}\left[\ln T_{ijt} \left| \mathbf{X}_{ijt}, \ln T_{ijt} > 0 \right.\right]$.

To deal with these difficulties, we treat our estimation equation as a corner-solutions model in the sense of Wooldridge (2002), which is a special case of the censored regression. In the following, we call $\mathbf{X}_{ijt}$, which stands for all exogenous explanatory variables in (7), the "the gravity force". The underlying statistical model is written as

$$\left(\ln T^{*}_{ijt} - \mathbf{X}_{ijt}\boldsymbol{\beta}\right)/\sigma \sim \text{N}[0, 1]\,, \tag{8}$$

where $\boldsymbol{\beta}$ is a vector of gravity-related parameters, and N[0, 1] denotes the standard-normal distribution. In other words, the mean of trade potential $\ln T^{*}_{ijt}$, conditional on gravity forces $\mathbf{X}_{ijt}$, is equal to $\mathbf{X}_{ijt}\boldsymbol{\beta}$, with variance equal to σ^2.

[15] Helpman et al. (2004) circumvent these difficulties by interpreting all observations with $T_{ijt} = 0$ as missing observation. However, this comes at the cost of sacrificing all information present in zero trade data, which clearly runs counter to our corner-solutions interpretation of the gravity equation.

[16] Heteroskedasticity arises because at the "corner", i.e., where $\text{E}\left(\ln T_{ijt} \left| \mathbf{X}_{ijt} \right.\right) = 0$, we observe only one-sided deviations. More generally, for values of explanatory variables leading to a lower $\text{E}\left(\ln T_{ijt} \left| \mathbf{X}_{ijt} \right.\right) = 0$, the variance of the error term is smaller.

The relationships between the parameters $\boldsymbol{\beta}$ and the two margins mentioned above are as follows. The overall effect is given by

$$\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt}) = \Phi(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)[\mathbf{X}_{ijt}\boldsymbol{\beta} + \sigma\lambda\,(z)]\,, \quad \text{with} \tag{9}$$

$$\partial\,\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt})/\partial X^{r}_{ijt} = \beta^{r}\Phi(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)\,, \tag{10}$$

where $z \equiv \mathbf{X}_{ijt}\boldsymbol{\beta}\big/\,\sigma$, and a superscript r denotes a specific explanatory variable r. The term $\lambda(z)$ is the inverse Mills ratio, $\lambda(z) =: \phi(z)/\Phi(z)$, with ϕ and Φ denoting the standard normal density and the standard normal distribution function, respectively.[17] Given (8), the conditional probability of positive trade, $\Phi\left(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma\right)$, is less than 1, which gives rise to attenuation.[18] The reason for this is straightforward. Suppose $\beta^{r} > 0$, and consider a reduction in X^{r}_{ijt}. The marginal effect on the conditional mean of the uncensored variable T^{*}_{ijt} must clearly be larger (in absolute value) than the effect on the conditional mean of the "cornered" variable T_{ijt}. However, the economic interpretation of (10) in the gravity context is not straightforward, as it is an amalgam of the intensive and the extensive margin.

The intensive margin relates to the expected value of trade, conditional on the gravity force $\mathbf{X}_{ijt}$ (see Greene 2003: 670):

$$\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt}, \ln T_{ijt} > 0) = \mathbf{X}_{ijt}\boldsymbol{\beta} + \sigma\lambda\,(z)\,, \quad \text{with} \tag{11}$$

$$\partial\,\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt}, \ln T_{ijt} > 0)/\partial X^{r}_{ijt} = \beta^{r} - \beta^{r}\lambda(z)[z + \lambda\,(z)]\,, \tag{12}$$

where $\lambda(z)$ again is the inverse Mills ratio, with $z \equiv \mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma$. The extensive margin relates to the probability of a given country pair already having taken up a trade relationship, again given the gravity force and time-specific effects:

$$\Pr(\ln T_{ijt} \geq 0|\mathbf{X}_{ijt}) = \Phi(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)\,, \tag{13}$$

where the marginal coefficient, $\partial\Pr[\ln T_{ijt} \geq 0|\mathbf{X}_{ijt}]/\partial X^{r}_{ijt}$, may be derived by differentiating the equality $\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt}) = \Pr(\ln T_{ijt} \geq 0|\mathbf{X}_{ijt})$

[17] The first term in (9) is the discrete part of the censored distribution, measuring the probability mass at zero, while the second term is the conditional mean of the corresponding truncated normal distribution. $\lambda(y)$ gives the hazard function of the standard-normal distribution; see Greene (2003: 762–763).

[18] The interpretation of $\Phi(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$ follows from $\Pr\left[\left(\ln T^{*}_{ijt} - \mathbf{X}_{ijt}\boldsymbol{\beta}\right)/\sigma < -\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma\right] = \Pr\left(\ln T^{*}_{ijt} < 0\right) = \Phi(-\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$. Due to symmetry, we have $\Phi(-\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma) = 1 - \Phi$ $(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$. Hence, $\Phi(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$ is the complementary probability for $\Phi(-\mathbf{X}_{ijt}-/\sigma)$, and thus the probability of positive trade.

$\times \mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt}, \ln T_{ijt} > 0)$, and invoking the associated marginal effects; see Wooldridge (2002: 523). Note that the marginal effects are not constant on either of the two margins. Moreover, it is clear from (13) that the extensive margin defined in this way does include the term $n_{t,t}$ in equation (3), but not the pseudo-extensive margin ΔV_t, as repeatedly mentioned above.

From equation (11), we may write

$$\begin{aligned} &\ln T_{ijt} = \mathbf{X}_{ijt}\,\boldsymbol{\beta} + \sigma\lambda(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma) + \varepsilon_{ijt} \\ &\text{with} \quad \mathrm{E}(\varepsilon_{ijt}|\mathbf{X}_{ijt}, \ln T_{ijt} > 0) = 0\,. \end{aligned} \tag{14}$$

It is clear that running OLS of $\ln T_{ijt}$ on $\mathbf{X}_{ijt}$ for $\ln T_{ijt} > 0$ amounts to omitting the variable $\lambda(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$. If the covariance between $\mathbf{X}_{ijt}$ and $\lambda(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$ is nonzero, then the coefficients $\boldsymbol{\beta}$ are inconsistently estimated. Moreover, regressing $\ln T_{ijt}$ on $\mathbf{X}_{ijt}$ using all of the data, including where $T_{ijt} = 1$, will not consistently estimate $\boldsymbol{\beta}$ either, since $\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt})$ is nonlinear in $\mathbf{X}_{ijt}$, $\boldsymbol{\beta}$ and σ.

The bias involved becomes clear from looking at the simple univariate case, using y and x to denote the dependent and independent variable. The model then becomes $\mathrm{E}(y|x, y > 0) = \beta x + \sigma\lambda(\beta x/\sigma)$. Ignoring the second term and running a regression of y on x for all observations $y > 0$ introduces the aforementioned omitted-variable bias. The estimated coefficient $\hat{\beta}$ is $\hat{\beta} = \beta + \sigma \operatorname{cov}[x, \lambda\,(\beta x/\sigma)]\,/\operatorname{var}(x)$. Since $\lambda\,(c) = \phi\,(c)\,/\Phi\,(c) > 0$ and $\lambda'\,(c) = -\lambda\,(c)\,[c + \lambda\,(c)] < 0$, we know that $\operatorname{cov}[x, \lambda\,(\beta x/\sigma)] > 0$. Moreover, we know that in a sample where no corner solutions appear $\mathrm{E}(y|x, y > 0) = \mathrm{E}(y|x) = \beta x$. Hence, since $\operatorname{cov}[x, \lambda(\beta x/\sigma)] > 0$, we conclude that $\hat{\beta} > \beta$. The omitted-variable bias causes an overestimation of β. In the long run, the extensive margin should disappear, as trade eventually does cover the whole world, whence $\hat{\beta} \to \beta$ with time. Accordingly, $\hat{\beta}$ falls toward β as time unfolds.

5 Estimation Results

We now proceed toward estimating our corner-solutions model on a subset of the data set used by Rose (2004). In particular, we focus on the time span between 1970 and 1990, because in that period very few new countries came into existence. Hence, we may expect that our results are note overly contaminated by the pseudo-extensive margin (see above). The summary statistics of the dependent variables and the covariates used in all of the models are presented in Table 1. As argued in the introduction, we expect that, among other things, our estimation sheds new light on the distance puzzle in that we may separately identify the time pattern of the distance coefficient at the extensive and the intensive margin of trade. A similar expectation also pertains to the role of WTO membership for bilateral trade.

Table 1: *Summary Statistics*

	Model			
	Restricted sample, $T > 0$ only		Full sample including $T = 0$	
Ln real value of (exports+imports)	10.0369	3.1836	5.5360	5.5233
Share of active trade relationships in sample	1.0000	0.0000	0.5620	0.4961
Ln distance	8.1778	0.8010	8.4887	0.8126
Ln distance × time	95.7212	49.9438	100.3139	51.3477
Dummy: both countries in WTO	0.4787	0.4995	0.3581	0.4794
Dummy: no country in WTO	0.4423	0.4967	0.4781	0.4995
#0,1,2: sum island	0.3409	0.5364	0.3963	0.5633
#0,1,2: sum landlocked	0.2482	0.4668	0.3585	0.5415
#0,1,2: sum Sachs-Warner open countries	0.7001	0.6463	0.5091	0.6210
Ln product of real GDPs	48.0002	2.5641	47.4561	2.6695
Dummy: both countries in same RTA	0.4777	0.4995	0.4154	0.4928
Ln price of oil	3.0924	0.4476	3.1016	0.4476
Ln real world GDP	16.8174	0.2186	16.8210	0.2165
Dummy: common border	0.0299	0.1704	0.0195	0.1383
Dummy: colonial ties	0.0240	0.1530	0.0136	0.1157
Dummy: common language	0.2015	0.4011	0.1734	0.3786
Number of observations	102,823		186,419	

Source: We use the data kindly made available by Rose on his website. See http://faculty.haas.berkeley.edu/arose/RecRes.htm

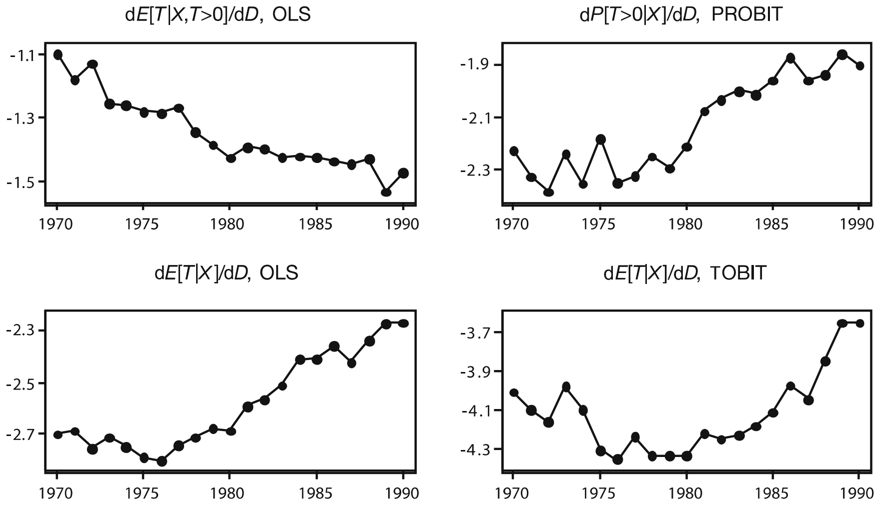
Before turning to the estimation results of our corner-solutions model, we provide some "naive" exploration of the time dependency of the distance coefficient, using OLS and Probit techniques. Thus, Figure 4a shows estimates of the distance coefficient when (7) is applied to the data repeatedly, year-by-year. The upper-left panel plots the estimates obtained from OLS, restricting the sample to country pairs that do in fact report positive trade. This ignores the extensive margin of trade, which we have emphasized in our accounting framework for trade in Section 3 and which we have argued to be a promising candidate for resolving the distance puzzle in Section 2. The regressions include a list of standard covariates (see Table 2) and a comprehensive set of country-fixed effects. The absolute value of the distance coefficient increases over time from about 1.10 to approximately 1.50. While the coefficient is at the higher range of the studies discussed in the meta analysis by Disdier and Head (2004), the time pattern neatly highlights what the literature has dubbed the distance puzzle. We shall turn to the question of statistical significance in this time trend below. For the present moment, the reader should contrast the upper-left panel in Figure 4a with that in Figure 4b. The latter shows time-varying estimates of the distance coefficients $\tilde{\delta}_t$ obtained in a panel framework. Time dependence of distance is captured by substituting

$$(1-\sigma)\,\delta \ln D_{ij} = \left(\tilde{\delta}_{1970}\phi_{1970} + \tilde{\delta}_{1971}\phi_{1971} + \ldots + \tilde{\delta}_{1990}\phi_{1990}\right)\ln D_{ij} \quad (15)$$

into (7).

In contrast to the year-by-year estimates, the panel estimation allows for time dependency only of the distance coefficient, in addition to year-fixed effects that account for the influence of the business cycle. All other coefficients, including the large array of country-fixed (importer plus exporter) effects, are constrained to be constant over time. This procedure leads to precise estimation of country-fixed effects, thus controlling for country-specific unobserved variables that are time-invariant. Interestingly, while the absolute value of the distance coefficient still rises over time, the pattern is less clear-cut than in Figure 4a. Moreover, the series seems to exhibit a structural break in the late seventies, rather than a linear time trend. Column (1) in Table A1 in the Appendix shows that the distance coefficients are estimated with satisfactory precision. The important thing to note here is that the distance puzzle survives the inclusion of fixed effects, both quantitatively and qualitatively. It also survives other, less fundamental methodological

Figure 4a: *Year-by-Year Estimates of the Distance Coefficient*



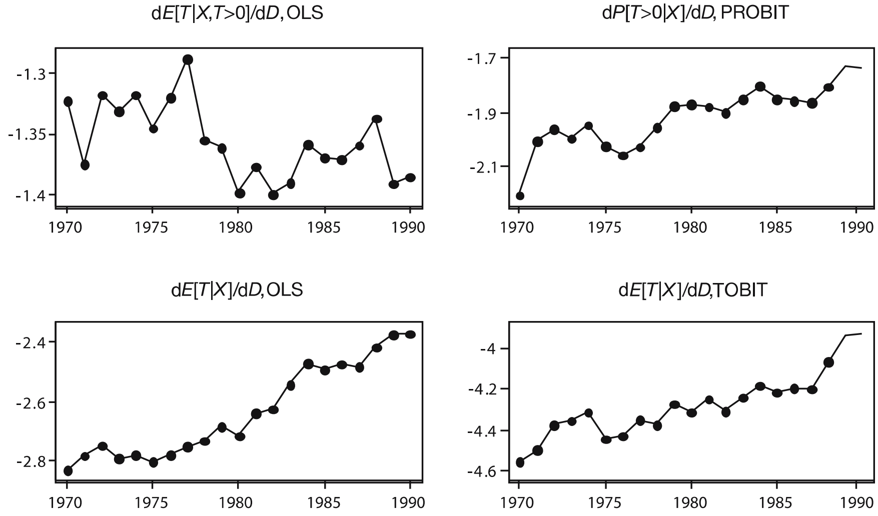
$\mathrm{d}E[T|X,T>0]/\mathrm{d}D$, OLS

$\mathrm{d}P[T>0|X]/\mathrm{d}D$, PROBIT

$\mathrm{d}E[T|X]/\mathrm{d}D$, OLS

$\mathrm{d}E[T|X]/\mathrm{d}D$, TOBIT

Figure 4b: *Panel Estimates of the Distance Coefficient*



$\mathrm{d}E[T|X,T>0]/\mathrm{d}D$, OLS

$\mathrm{d}P[T>0|X]/\mathrm{d}D$, PROBIT

$\mathrm{d}E[T|X]/\mathrm{d}D$, OLS

$\mathrm{d}E[T|X]/\mathrm{d}D$, TOBIT

Table 2: *Regression Results: Naive Regressions of the Intensive vs. the Extensive Margin, Panel Estimations (Dependent Variable: ln of Real Bilateral Trade)*

	(1) Intensive margin only OLS $E(T\|X,T > 0)$	(2) Extensive margin only PROBIT $P(T > 0\|X)$	(3) Both margins OLS $E(T\|X)$
Ln distance	−1.3114	−0.7305	−3.2627
	(0.0389)	(0.0340)	(0.0902)
Ln distance × time	−0.0043	0.0055	0.0231
	(0.0018)	(0.0011)	(0.0023)
Dummy: both countries in WTO	0.0934	0.1506	0.5020
	(0.0846)	(0.0243)	(0.0945)
Dummy: no country in WTO	0.0004	0.0507	0.0679
	(0.0705)	(0.0180)	(0.0682)
#0,1,2: sum island	−0.9691	−0.5821	−1.9431
	(0.4754)	(0.0462)	(0.1682)
#0,1,2: sum landlocked	−0.3076	5.6736	5.6883
	(0.0945)	(0.2352)	(0.1947)
#0,1,2: sum Sachs-Warner open countries	−0.0622	−0.0072	−0.0767
	(0.0329)	(0.0093)	(0.0332)
Ln product of real GDPs	0.5697	−0.0982	−0.2325
	(0.0422)	(0.0029)	(0.0094)
Dummy: both countries in same RTA	−0.2760	0.1421	0.4912
	(0.0544)	(0.0349)	(0.0803)
Ln price of oil	0.2970	0.0358	0.3341
	(0.0153)	(0.0053)	(0.0158)
Ln real world GDP	5.0138	1.2527	6.5843
	(0.2669)	(0.0995)	(0.2667)
Dummy: common border	0.3309	−0.6421	−0.9815
	(0.1337)	(0.0301)	(0.2468)
Dummy: colonial ties	1.2220	0.2732	1.2828
	(0.1144)	(0.0376)	(0.1881)
Dummy: common language	0.3605	0.0261	0.3413
	(0.0557)	(0.0184)	(0.0740)
Number of observations	102,823	170,432[a]	186,419
Number of dyads	7,539	8,154	10,416
Pseudo R-squared	0.7265	0.7116	0.7799
RMSE	1.6673	–	2.5932
Pseudo log likelihood	–	−33,693	–

Note: Robust standard errors (clustering by country pairs) in parentheses. Exporter and importer fixed effects, as well as a linear time trend are included in each regression but not reported. Marginal effects are evaluated at sample means. For dummy variables marginal effects are for discrete change of dummy variable from 0 to 1.

[a] Probit estimation is over fewer observation than OLS in column (3) since the presence of exporter and importer dummies completely predicts some outcomes, so that they drop out of the sample.

modifications, as discussed in Brun et al. (2004) and the meta analysis of Disdier and Head (2004).[19]

Column (1) in Table 2 replaces (15) with a parametric specification of the time dependency of the distance coefficient. In particular, we model a linear trend in the distance coefficient.[20] While the estimates shown in the upper-left panel of Figure 4b suggest caution with the use of a linear trend, this procedure allows for a simple way of testing whether there is a statistically significant trend in the coefficients. Indeed, the time-distance interaction is statistically significant at the 1 percent level and enters with a negative sign. But the time effect is small in magnitude: It would take about 300 years for the distance coefficient to double. However, there is no evidence whatsoever pointing to a decline in the absolute value of the distance coefficient over time. The remaining covariates included in the regression have the usual signs and are mostly estimated with high precision.

Next, we turn to the time dependency of the distance coefficient in Probit estimations of the extensive margin. We replace the left hand side of (7) by an indicator variable which is unity if a given trade relationship was active at time *t*, and zero otherwise. The resulting specification is estimated using a Probit model, employing the same covariates as in the OLS regressions discussed above. The upper-right panels in Figures 4a and 4b show that the distance coefficient obtained in this model falls in absolute value over time. Column (2) in Table 2 establishes that the trend in the estimates is statistically and economically significant. If the trend were to continue, the panel estimates would imply that by the early years of the next century distance no longer has any explanatory power for predicting the existence of bilateral trade relationships between countries. These results are in stark contrast to the ones discussed above, and they provide some tentative evidence for a declining role of distance along the extensive margin.

The natural next step is to include observations with zero trade and run an OLS regression, using the same covariates as above. The results are shown in the lower-left panels of Figures 4a and 4b. It turns out that the extensive

[19] We have run a large number of robustness checks on the persistence of the distance puzzle in our data, using specifications that vary in parsimony (basic gravity-type regressors versus additional regressors in line with Table 1), in treatment of global year-specific events (including the price of oil and real world GDP versus unspecified year-specific effects), and error specification (pooled OLS versus random effects estimator). The distance puzzle shows up as a robust phenomenon across all estimation procedures. More detailed results may be found in Felbermayr and Kohler (2004).

[20] $(1-\sigma)\delta \ln D_{ij}$ is substituted by $\bar{\delta}_1(\ln D_{ij} \times \mathit{time}) + \bar{\delta}_2 \ln D_{ij} + \bar{\delta}\mathit{time}$ in equation (7).

margin dominates the overall assessment of the time dependency of the distance effect. The results indicate a downward trend in the absolute value of the distance coefficient which is statistically and economically significant. However, it is important to remember from our discussion above that OLS does not yield consistent estimates if we include zero trade observations. The lower-right panels in Figures 4a and 4b show that there is evidence for a decrease in the importance of distance in line with the OLS and Probit estimates presented above. The time trend is statistically significant and economically meaningful. If this trend were to continue, distance would lose its bite on bilateral trade volumes around the year 2200.

We now turn to the time pattern of the distance effect as revealed by a consistent estimation of our corner-solutions model. Table 3 shows regression results for the corner-solutions model. Column (1) reports the marginal effects of the covariates on the trade potential. As argued above, these are difficult to interpret economically. Column (2) presents coefficients obtained by differentiating $E(T|X, T > 0)$. We interpret this column as a consistent estimate of the *intensive margin*. Column (3) presents the coefficients relating to $E(T|X)$, which is the unconditional expectation of bilateral trade. Finally, column (4) contains estimates of the probability of an interior solution, i.e., $P(T > 0|X)$, which we interpret as the *extensive margin*. All estimates reported are marginal coefficients. Since the model is nonlinear, they are not constant. The values shown are those for the respective sample means.[21]

Importantly, we find a positive time trend in the distance coefficient which is statistically and economically significant. This implies that the importance of distance, if estimated in a consistent way, is indeed falling over time, not only on the extensive margin, but also on the intensive margin. While the exact time pattern of the coefficients may be nonlinear in the year-by-year model (Figure 4a, lower-right panel), there is no remaining evidence whatsoever indicating the presence of a distance puzzle. Moreover, when looking at the panel model (Figure 4b, lower-right panel), there seems to be a linear time trend; this also bears out in the results presented in Table 3.

As similar conclusion also holds for the effect of WTO membership on bilateral trade. Evidence presented by Rose (2004), based on a standard OLS estimation of the equation restricted to cases where reported trade is positive, seems to indicate that there is no statistically significant effect of

[21] In computing the partial derivatives, sums of binary variables which take integer values over the interval [0, 1] are treated as continuous variables.

Table 3: *Regression Results: The Corner Solutions Model. (Dependent Variable: ln of Real Bilateral Trade)*

	(1)	(2)	(3)	(4)
	Latent variable	Expectation conditional on interior solution	Expectation unconditional on interior solution	Probability of interior solution
	$E(T^*\|X)$	$E(T\|X,T > 0)$	$E(T\|X)$	$P(T > 0\|X)$
Ln distance	−5.0747	−2.7061	−3.7978	−0.3423
	(0.1026)	(0.0547)	(0.0768)	(0.0069)
Ln distance × time	0.0215	0.0115	0.0161	0.0015
	(0.0080)	(0.0043)	(0.006)	(0.0005)
Dummy: both countries in WTO	2.1130	1.1721	1.6256	0.1348
	(0.1815)	(0.1007)	(0.1396)	(0.0116)
Dummy: no country in WTO	1.2386	0.6611	0.9266	0.0834
	(0.1557)	(0.0831)	(0.1165)	(0.0105)
#0,1,2: sum island	−1.6625	−0.8865	−1.2441	−0.1121
	(0.1459)	(0.0778)	(0.1092)	(0.0098)
#0,1,2: sum landlocked	−2.7534	−1.4682	−2.0606	−0.1857
	(0.1115)	(0.0595)	(0.0835)	(0.0075)
#0,1,2: sum Sachs-Warner open countries	1.9643	1.0474	1.4700	0.1325
	(0.1346)	(0.0718)	(0.1007)	(0.0091)
Ln product of real GDPs	0.5014	0.2674	0.3753	0.0338
	(0.0201)	(0.0107)	(0.015)	(0.0014)
Dummy: both countries in same RTA	0.8828	0.4758	0.6660	0.0587
	(0.1254)	(0.0676)	(0.0946)	(0.0083)
Ln price of oil	0.5595	0.2983	0.4187	0.0377
	(0.0371)	(0.0198)	(0.0278)	(0.0025)
Ln real world GDP	7.8197	4.1698	5.8520	0.5274
	(0.5047)	(0.2691)	(0.3777)	(0.0340)
Dummy: common border	−2.2660	−1.0664	−1.5137	−0.1723
	(0.4177)	(0.1966)	(0.2791)	(0.0318)
Dummy: colonial ties	0.8053	0.4486	0.6231	0.0512
	(0.2660)	(0.1482)	(0.2058)	(0.0169)
Dummy: common language	0.4637	0.2515	0.3516	0.0306
	(0.0813)	(0.0441)	(0.0617)	(0.0054)
Number of observations	186,419			
Number of dyads	10,416			
Pseudo R-squared	0.7265			
RMSE	1.6673			
Pseudo log likelihood	−345,840			

Note: Robust standard errors (clustering by country pairs) in parentheses. Exporter and importer fixed effects, as well as a linear time trend are not reported. Marginal effects are evaluated at sample means. For dummy variables marginal effects are for a discrete change of dummy variable from 0 to 1.

WTO membership. In effect, our estimates are a robustness check of the Rose results, since the WTO puzzle does not disappear upon introduction of a comprehensive set of importer and exporter dummies. However, once we switch attention to the extensive margin of world trade, the effect of WTO membership is statistically significant and economically meaningful: Column (2) in Table 2 shows that compared to country pairs which do not *both* belong to the WTO, the probability to find an existing trade relationship between two member countries is 10 percentage points higher. This carries over to the corner-solutions model, where the difference between the two cases is somewhat smaller, but still highly significant. While other explanations of the WTO puzzles have been proposed, such as the importance of restricting the sample to countries that really abide by the WTO rules (see Subramanian and Wei 2004), the effect of WTO membership on the extensive margin of world trade seems a plausible, yet usually neglected possibility.

6 Conclusions

Our Tobit estimates of a corner-solutions version of the gravity model indicate that the distance puzzle is a reflection of the extensive margin of world trade expansion. This becomes most strikingly clear in our Probit estimates. There are several routes to pursue for future research. First, distance-related trade costs are likely to be nonlinear in distance. It seems questionable that the oft quoted technological improvements in transport and communication have been equally relevant for long-distance and short-distance trade. Similarly, improvements may have been differently pronounced for land and sea transport. Obviously, an improved explicit treatment of the transport sector, or—more generally—of trade costs, is required to fully address these points. Moreover, distance-related trade costs are likely to be of differing importance across goods or industries.

Meanwhile, our results should reinforce the widespread notion that the post–World War II era is, indeed, characterized by a long-run decline in the trade inhibiting force of geographic distance. This is most obvious when studying the time pattern of the distance coefficient in Probit regressions for the extensive margin of world trade. However, our view of the distance puzzle is that, even if resolved in the strict sense of the word, it serves as a “warning shot” against exaggerated views of a dramatically “shrinking world geography”. Our results do *not* warrant the conclusion that distance and trade costs have become negligibly small. The world has not yet become a “global village”, and it probably never will. Indeed, a further important conclusion of our study is that world trade, even after 5 decades of spectacular growth, is still far away from covering the whole world. The extensive margin of trade, where dormant bilateral trade relationships become utilized, still leaves much ground to be covered, and much gains from trade to be reaped.

Appendix

Table A1: *Regression results: Time Pattern of the Distance Coefficient, Panel Estimates, 1970–1990 (Dependent Variable: ln of Real Bilateral Trade)*

	(1)	(2)	(3)	(4)
	OLS	OLS	PROBIT	TOBIT
	$\mathrm{d}E(T \mid X,T > 0)/\mathrm{d}D$	$\mathrm{d}E(T \mid X)/\mathrm{d}D$	$\mathrm{d}P(T > 0 \mid X)/\mathrm{d}D$	$\mathrm{d}E(T \mid X)/\mathrm{d}D$
1970	−1.3218	−2.8284	−0.7304	−4.5582
	(0.0412)	(0.0548)	(0.0302)	(0.1113)
1971	1.3750	−2.7806	−0.6641	−4.5019
	(0.407)	(0.0558)	(0.0292)	(0.1123)
1972	1.3178	−2.7465	−0.6507	−4.3714
	(0.3957)	(0.0552)	(0.0285)	(0.1100)
1973	−1.3308	−2.7901	−0.6607	−4.3563
	0.3898	(0.0551)	(0.0293)	(0.1091)
1974	−1.3177	−2.7789	−0.6460	−4.3132)
	(0.3926)	(0.0564)	(0.0297)	(0.1084)
1975	−1.3450	−2.8006	−0.6707	−4.4468
	(0.3966)	(0.0536)	(0.0273)	(0.1072)
1976	−1.3189	−2.7764	−0.6815	−4.4320
	(0.3907)	(0.0513)	(0.0274)	(0.1046)
1977	−1.2874	−2.7006	−0.6714	−4.3519
	(0.3944)	(0.0520)	(0.0273)	(0.1038)
1978	−1.3544	−2.7324	−0.6487	−4.3709
	(0.3886)	(0.0512)	(0.0258)	(0.1030)
1979	−1.360	−2.6812	−0.6216	−4.2718
	(0.3870)	(0.0501)	(0.0248)	(0.1008)
1980	−1.3977	−2.7144	−0.6196	−4.3144
	(0.3888)	(0.0504)	(0.0245)	(0.1020)
1981	−1.3770	−2.6390	−0.6224	−4.2467
	(0.3813)	(0.0485)	(0.0240)	(0.0997)
1982	−1.3994	−2.6239	−0.6290	−4.3072
	(0.3811)	(0,0483)	(0.0242)	(0.1009)
1983	−1.3901	−2.5431	−0.6128	−4.2416
	(0.3829)	(0.0476)	(0.0232)	(0.0999)
1984	−1.3579	−2.4728	−0.5975	−4.1811
	(0.3746)	(0.0481)	(0.0229)	(0.0998)
1985	−1.3694	−2.44936	−0.6126	−4.2162
	(0.0374)	(0.0476)	(0.0234)	(0.1000)
1986	−1.3706	−2.4764	−0.6149	−4.1951
	(0.3719)	(0.0471)	(0.0235)	(0.0994)
1987	−1.3586	−2.4831	−0.6175	−4.1993
	(0.3688)	(0.0467)	(0.0237)	(0.0988)
1988	−1.3369	−2.4189	−0.5983	−4.0608
	(0.0369)	(0.0466)	(0.0233)	(0.0977)
1989	−1.3903	−2.3767	−0.5729	−3.9322
	(0.3712)	(0.0462)	(0.0226)	(0.0960)
1990	−1.3850	−2.3744	−0.5747	−3.9244
	(0.3733)	(0.0460)	(0.0210)	(0.0954)

Note: Robust standard errors (clustering by country pairs) in parentheses. The regressions contain the same list of covariates as the regressions presented in Table 2. Exporter and importer fixed effects, as well as a linear time trend are included but not reported.

References

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review* 69 (1): 106–116.

Anderson, J. E., and E. van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review* 93 (1): 170–192.

Baier, S., and J. H. Bergstrand (2001). The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity. *Journal of International Economics* 53 (1): 1–27.

Bernard, A. B., J. B. Jensen, and P. K. Schott (2006). Trade Costs, Firms and Productivity. *Journal of Monetary Economics* 53 (1): 917–937.

Buch, C., J. Kleinert, and F. Toubal (2004). The Distance Puzzle: On the Interpretation of the Distance Coefficient in Gravity Equations. *Economics Letters* 83 (3): 293–298.

Brun, J.-F., C. Carrère, P. Guillaumont, and J. de Melo (2004). Has Distance Died? Evidence from a Panel Gravity Model. *World Bank Economic Review* 19 (1): 99–120.

Coe, D. T., A. Subramanian, N. T. Tamirisa, and R. Bhavnani (2002). The Missing Globalization Puzzle. IMF Working Paper 02/171. International Monetary Fund, Washington, D.C.

Deardorff, A. V. (1998). Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? In J. A. Frankel (ed.), *The Regionalization of the World Economy.* Chicago: University of Chicago Press.

Deardorff A. V. (2004). Local Comparative Advantage: Trade Costs and the Pattern of Trade. Research Seminar in International Economics Discussion Paper No. 500. University of Michigan, Ann Arbor.

Disdier, A.-C., and K. Head (2004). *The Puzzling Persistence of the Distance Effect on Bilateral Trade.* Mimeo. University of British Columbia.

Eaton, J., and S. Kortum (2002). Technology, Geography, and Trade. *Econometrica* 70 (5): 1741–1779.

Eaton, J., and A. Tamura (1994). Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns. *Journal of the Japanese and International Economies* 8 (4): 478–510.

Egger, P., and M. Pfaffermayr (2004). Distance, Trade and FDI: A Hausman–Taylor SUR Approach. *Journal of Applied Econometrics* 19 (2): 227–246.

Eichengreen, B. (1995). Trade Blocs, Currency Blocs and the Reorientation of World Trade in the 1930s. *Journal of International Economics* 38 (1–2): 1–24.

Eichengreen, B., and D. A. Irwin (1997). The Role of History in Bilateral Trade Flows. In J. A. Frankel (ed.), *The Regionalization of the World Economy.* Chicago: University of Chicago Press.

Evenett, S. J., and A. J. Venables (2002). *Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows.* Mimeo. London School of Economics.

Feenstra, R. (2004). *Advanced International Trade – Theory and Evidence*. Princeton: Princeton University Press.

Feenstra, R., and A. K. Rose (1997). Putting Things in Order: Patterns of Trade Dynamics and Macroeconomics. NBER Working Paper 5975. National Bureau of Economic Research, Cambridge, Mass.

Feenstra, R., J. A. Markusen, and A. K. Rose (2001). Using the Gravity Equation to Differentiate among Alternative Theories of Trade. *Canadian Journal of Economics* 34 (2): 430–447.

Felbermayr, G. J., and W. Kohler (2004). Exploring the Intensive and Extensive Margins of World Trade. CESifo Working Paper No. 1276. Center for Economic Studies, Munich.

Gleditsch, K. S. (2002). Expanded Trade and GDP Data. *Journal of Conflict Resolution* 46 (5): 712–724.

Greene, W. (2003). *Econometric Analysis*. New York: Prentice Hall.

Haveman, J., and D. L. Hummels (2004). Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization. *Canadian Journal of Economics* 37 (1): 199–218.

Helpman, E. (1987). Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries. *Journal of the Japanese and International Economies* 1 (1): 62–81.

Helpman, E., M. Melitz, and Y. Rubinstein (2004). Trading Patterns and Trading Volumes. Mimeo. Harvard University.

Hummels, D., and P. J. Klenow (2005). The Variety and Quality of a Nation's Exports. *American Economic Review* 95 (3): 704–723.

Markusen, J. R. (2002). *Multinational Firms and the Theory of International Trade.* Cambridge, Mass.: MIT Press.

Neary, J. P. (2005). Trade Costs and Foreign Direct Investment. CER Working Paper 05/12. Available at ⟨http://www.ucd.ie/economic/workingpapers/WP05.12.pdf⟩.

Rose, A. (2004). Do We Really Know That the WTO Increases Trade? *American Economic Review* 94 (1): 98–114.

Santos Silva, J. M. C., and S. Tenreyro (2003). Gravity-Defying Trade. Working Paper 03, 1. Federal Reserve Bank of Boston, Boston, Mass.

Subramanian, A., and S.-J. Wei. (2003). The WTO Promotes Trade Strongly but Unevenly. NBER Working Paper 10024. National Bureau of Economic Research, Cambridge, Mass.

Wang, Z. K., and A. Winters (1992). The Trading Potential of Eastern Europe. *Journal of Economic Integration* 7 (2): 113–136.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

Yi, K.-M. (2003). Can Vertical Specialization Explain the Growth of World Trade? *Journal of Political Economy* 111 (1): 52–102.