EXPLORING THE INTENSIVE AND EXTENSIVE MARGINS OF WORLD TRADE

GABRIEL J. FELBERMAYR
WILHELM KOHLER

EXPLORING THE INTENSIVE AND EXTENSIVE MARGINS OF WORLD TRADE

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 classification: F12, F15 .

Keywords: bilateral trade, globalization, gravity model.

Gabriel J. Felbermayr
European University Institute
Villa San Paolo
Via Piazzuola 43
50133 Firenze
Italy
gfelberm@iue.it

Wilhelm Kohler
Eberhard Karls University Tübingen
Nauklerstrasse 47
72074 Tübingen
Germany
wilhelm.kohler@jku.at

.

1 Motivation

Despite much talk about economic globalization and the oft-quoted spectacular post-World-War-II growth in trade flows, world trade at any point in time covers a surprisingly small part of the world. In 1950, almost 52 percent of the potential number of bilateral trading relationships did not report any manufacturing trade at all. By 1997 the potential number of bilateral trading relationship had increased by more than 500 percent, but the share of bilateral trading 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 trading relationships change their trade volume, while at the *extensive* margin new trading relationships are established, or existing ones abandoned, partly based on the formation of new countries.

It is somewhat surprising that this two-fold margin has so far not been systematically explored in empirical studies of international trade. In particular, it has received little or no attention in applications of the gravity-based approach that for various reasons has recently experienced a remarkable revival. One finds occasional reference to the large and time-varying number of zero entries in bilateral trade matrices,[2] but a systematic treatment of the issue is still missing. 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, the aforementioned numbers strongly suggest that in doing so one 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 quesitonable, as are their statistical properties.

In this paper, we provide descriptive evidence on the quantitative importance of the extensive and the intensive margin in the growth of world trade since World-War-II. In doing

---

[1] These numbers are from a data set to be described in detail below.

[2] See, for instance, Evenett & Venables (2002) who employ an ad-hoc model to explain the expansion of a country's exports of disaggregated commodities into new markets. A further case in point is Haveman & Hummels (2004) whose imperfect-specialization-based foundation of the gravity equation is motivated by its better consistency with the large number of zero entries in bilateral trade matrices. Note that, while these studies focus on disaggregate trade where zero entries seem more likely, the above-mentioned numbers refer to aggregate manufacturing trade.

so, we carefully separate two forms in which the extensive margin arises: it may be due to the formation of new countries, or may arise in the form of trade being taken up for the first time between pre-existing countries where the trading relationship was previously dormant. We also look at the amount of trade generated by newly established trading relationships. Our analysis reveals the extent to which the two margins have shaped the enormous growth of world trade in various episodes between 1950 and 1997.

We then move on to a simple theoretical formulation of the gravity equation which takes explicit account of the aforementioned "dual margin" of world trade. In doing so, we emphasize the role of distance-related transportation cost. This is motivated by what has sometimes been called the *"distance-puzzle"*, i.e., econometric evidence suggesting that in the context of the gravity model the elasticity of bilateral trade with respect to distance has increased over time. Such evidence has recently been emphasized by Brun et al. (2004) and Coe et al. (2002), corroborating earlier evidence presented by Frankel (1997).[3] Leamer & Levinsohn (1995) were the first to draw attention to this trend, concluding that, contrary to popular notions of globalization, the world is not "getting smaller". They argue, in line with the gravity approach, that the driving force behind globalization is not lower distance-barriers, but increased dispersion of economic mass around the globe. More recently, Buch et al. (2004) argue against the notion of a distance-puzzle in that globalization may work as much through affecting the intercept of the gravity equation, as through a change in the estimated distance-elasticity. However, the finding that today distance apparently restricts bilateral trade more than in earlier periods still constitutes a puzzle, since it suggests that, *other things equal,* distance-related trade barriers have become more important. This is difficult to reconcile with the undisputed fact that the past 5 decades have witnessed enormous progress in transport and communication which should reduce, rather than increase, the trade-barrier effect of distance.[4]

There is, thus, still a puzzle to be resolved. Indeed, we interpret this puzzle as a clear indication of an empirical failure of the *conventional* gravity approach. There may be several

---

[3] It is interesting and important to note that distance appears to have lost its trade-restrictive power, not only over the recent decades, but also over a longer centenary horizon; see Irwin & Terviö (2002).

[4] See, however, Hummels (1999) who qualifies this view by direct evidence on transport costs.

routes of remedy that seem promising and worth pursuing. We argue that an adequate treatment of both the extensive and the intensive margins of world trade is one of them. Following Wooldridge (2002), we present a general model where zero-bilateral trade arises as a "corner-solution" in an environment of gravity-type forces on trade. The basic assumption is that both margins are governed by the same underlying determinants. Our estimation equation involves a latent variable and a rudimentary modeling of the transport sector. We provide an in-depth discussion of the econometric issues raised by such a model. We show that both, the usual linear estimator as well as the non-linear least squares estimator imply a mis-specification that shows up in the form the distance-puzzle. More importantly, however, the conventional approach does not disentangle the extensive and the intensive margins of world trade. Based on our corner-solutions gravity model, we achieve such a disentangling by means of a Tobit estimation approach. A key result of our analysis is that the distance-puzzle does, indeed, disappear. Incorporating the extensive margin of world trade thus constitutes a clear improvement of the gravity approach.

The paper is structured as follows. Section 2 first discusses the methodological problem underlying the so-called distance-puzzle and gives intuitive reasons why an extension of theory towards a consistent treatment of the "dual margin" of world trade should , a-priori, be expected to form an important part of the solution. Section 3 describes our trade data and then 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 the data set used, followed by a discussion of the regression results. Section 6 concludes the paper by a summary of our results and their implications for future research.

2 A methodological view on the distance-puzzle

Fundamentally, the so-called distance-puzzle is simply a reflection of "missing trade", "or missing globalization". In other words, observed trade through time increasingly falls short of what estimates of the gravity model would predict, based on the evolution of *time-varying* determinants of trade. If the model is deemed correct regardless, then this shows up as an

increasing estimated importance of the *time-invariant* inhibiting force of distance. What makes this a "puzzle" is that it is grossly at odds with the widely held view, and lots of direct evidence, that advancements in transport and communication should have made distance a *less* important barrier through time. If we accept this as a valid a-priori conviction, then the simple verdict is that the theory performs badly, and increasingly so over time. The appropriate response, therefore, is trying to improve the theory, finding 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 in the theoretical specification of the model.[5]

Why, then, should we expect the "dual margin" of trade to play an important role in this attempt? The intuition is quite simple and runs as follows. If trade is the result of mass attraction and resistance from geographical distance, as suggested by gravity theory, and if attraction in some cases is not strong enough to generate trade at all, then ignoring all such cases 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 overestimation bias in the force of attraction also falls through time. However, if we take the estimated force of attraction on face value, the result is that an ever increasing volume of seemingly "missing trade" gets shouldered onto the time-invariant distance-variable.

There are, of course, other deficiencies of the model that might play a role. A quick explanation that is often brought up in connection with the distance puzzle relates to trade liberalization. If distance is large, then tariffs are a relatively small part of overall trade costs. If tariffs are equal for all trading relationships to start with, and if they are reduced by an equal (absolute) amount for all trading relationships, then – other things equal – the percentage reduction in the destination price will be larger for low-distance trading relationships. Tariff liberalization will then have a disproportionally large impact on short-distance trade. In the gravity model, this effect might be picked up by an increasing role of distance as a trade-

---

[5] See Trefler (1995) who similarly notes "missing trade" as an important apparent deficieny of the Heckscher-Ohlin trade model.

inhibiting factor through time. However, this route of explanation importantly hinges on an equal level of tariffs to start with, and on an equal amount of tariff liberalization across all trading relationships, which seems questionable.

One might also look for foreign direct investment (FDI) as a potential explanation for the distance-puzzle. The enormous increase in FDI as such, however, is not enough. First, the relationship between trade and FDI is far from clear-cut, even qualitatively. If they are substitutes, then an explanation of the puzzle would require that FDI has systematically replaced long distance trade more than short distance-trade. While distance and trade costs no doubt play a role for FDI, particularly through the well-known proximity-concentration trade off, it is not at all clear why FDI and subsequent host-country sales, instead of exports, should be more attractive for long-distance markets. Theory tells that 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. Empirically, distance appears to play an ambiguous role for FDI, which would cast additional doubt on whether it is a promising route to follow for solving the distance-puzzle for trade.[6]

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 thus 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 estimation based on panel data, and using a Tobit estimation approach.

3 Dissecting the dual margin of world trade

Expansion of world trade evolves at two margins. At the *extensive margin* new trading relationships between countries are opened up, or existing relationships are closed, while at the *intensive margin* the level of trade based on existing relationships is increased or reduced. In empirical studies of bilateral world trade this is reflected by a large and time-

---

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

varying number of zero-entries in trade matrices, and by unbalanced panels. To the best of our knowledge, there is no systematic account of the role that these two margins have played in the growth of overall world trade after World-War-II. Moreover, as we argue below, the coexistence of the two margins has not been adequately dealt with by the econometric techniques employed in estimating gravity equations on world trade. In this section we shed light on the empirical relevance of the issue by dissecting the growth of world trade along the lines of the aforementioned two margins.

3.1 Data issues

We start off with a few data issues. Empirical studies based on the gravity approach use *dyadic* observations on trade, whereby the unit of observation is a country pair. Any empirical investigation of the dual margin must rely on *observations* of dormant trading relationships. However, the data sources for dyadic trade often do not include such observations, but simply treat any country pair for which there is no reported trade during a given year as a *missing* observation. This may be a prime motivation for restricting empirical applications of the gravity approach to observations where bilateral trade flows are strictly positive. We have argued above that this is problematic as it ignores a lot of action over time. However, replacing missing values by zeros may be no less problematic, unless the data source provides information on whether the respective observations are genuinely missing or have been explicitly reported was zero.

In this paper, we rely on IMF Direction of Trade Statistics (DoTS) which does not provide such information. However, upon extensive scrutiny of the DoTS, Gleditsch (2002) concludes that 80% of all observations coded as missing do in fact represent zeros. He reaches this conclusion after a careful comparison of the DoTS with other data bases, such as the COMTRADE basis of the UN, data from the WTO or national accounts. Hence, we feel comfortable enough in assuming zero trade for all missing observations in the IMF DoTS. Coe et al. (2002) and Santos and Tenreyro (2004) adopt the same strategy.

Dyadic trade data typically pose a further data problem in that information provided by exporting countries often does not fit the data released by importing countries; see for instance Rozanski and Yeats (1994). However, international organizations have set up a

number of task forces to harmonize national standards and definitions[7], so that one should expect data quality to have improved over time. Moreover, while published data certainly remains prone to measurement errors, whether or not a country pair was trading at a certain point in time does not depend on standards or definitions and should – in principle at least – be easy to verify, since any transaction is always recorded twice, once in the exporting country, and then again in the importing country.

There is a presumption that import data are more reliable than export data, since most governments closely track imports which constitute a tax base, while no comparable incentive exists on the export side. This is reflected by the fact that the number of country pairs reporting positive trade is larger if based on import data than if based on exports. Better reliability has prompted some researchers to construct dyadic trade data from imports alone; see Coe et a. (2002) and Brun et al. (2004). This strategy, however, has an important shortcoming itself. Import flows are typically evaluated c.i.f., including in particular costs of transport and insurance. Using such data in a regression of bilateral trade on some proxy of transportation costs (distance, say) yields an inconsistent estimator for the distance-coefficient, as distance will, almost by definition, be correlated with the error term. Using more error-prone export data most likely results in a lower efficiency of the estimation procedure, but yields consistent estimates, since f.o.b. evaluation as such preserves standard regularity assumptions on the error term. Faced with the choice between efficiency and consistency, researchers often opt for consistency. 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.[8] Extensive checks have revealed that using f.o.b. data alone, while reducing the number of observations, does not affect the core results. In the sequel, the variable T_{ij} is defined as bilateral trade between country i and j, defined as the sum of recorded exports plus imports by the two countries (whence each trade flow is recorded four times) divided by 4.[9] The data cover 1950-1997 and are in constant US dollars, based on the US CPI (1983=100).

---

[7]See http://unstats.un.org/unsd/ for details.

[8]Indeed, we use those same data which have generously been 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 towards the end.

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

3.2 The vintage of trading relationships

We define the "vintage" of a trading relationship as the earliest time at which trade may occur between any 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 will all other open economies. We denote the number of active trading relationships of vintage h by $N_{t,h}$. This will typically be lower than the number of potential vintage-h trading relationships at time t, which is denoted by $V_{t,h}$. Total world trade at time t may thus be written as

$$T_t \;=\; \sum\nolimits_{h=t_0}^{t} \bar{T}_{t,h} N_{t,h} = \bar{T}_t \sum\nolimits_{h=t_0}^{t} n_{t,h} \zeta_{t,h} V_t, \tag{1}$$

$$\text{with} \quad n_{t,h} \;\equiv\; \frac{N_{t,h}}{V_{t,h}} \quad \text{and} \quad \zeta_{t,h} \equiv V_{t,h} \Big/ \sum\nolimits_{h=t_0}^{t} V_{t,h}\,, \tag{2}$$

where $\bar{T}_{t,h}$ is the average trade volume based on trading relationships of vintage h, and $\bar{T}_t$ is the average volume of trade at time t across all vintages: $\bar{T}_t = \sum_{h=t_0}^{t} \bar{T}_{t,h} \big/ \sum_{h=t_0}^{t} N_{t,h}$. Moreover, V_t denotes the overall number of potential trading relationships, $V_t \equiv \sum_{h=t_0}^{t} V_{t,h}$, and t_0 denotes the "beginning of time". In the second line, $n_{t,h}$ is defined as the share of active trading relationships within vintage h, and $\zeta_{t,h}$ is defined as the share of vintage h in the potential number of trading relationships. Obviously, $h \leq t$.

We now call

$$\Delta T_{\mathrm{int},t} \equiv \sum\nolimits_{h=t_0}^{t-1} \left(\bar{T}_{t,h} N_{t-1,h} - \bar{T}_{t-1,h} N_{t-1,h} \right) \tag{3}$$

a movement of world trade on the *intensive margin*, whence pre-existing relationships vary in trade volumes. Accordingly, variations in the number of active trading relationships,

$$\Delta N_t \equiv \sum\nolimits_{h=t_0}^{t-1} \left(N_{t,h} - N_{t-1,h} \right) + n_{t,t} \Delta V_t = \Delta N_{\mathrm{x},t} + n_{t,t} V_{t,t} \tag{4}$$

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 trading 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 (4) separates two types of extensive margins. At the extensive margin *proper*,

denoted by $\Delta N_{\mathrm{x},t}$, the utilization of potential trading relationships between *pre-existing* countries and vintages $h \leq t-1$ changes from $t-1$ to t. In terms of theory, this should obviously be treated as endogenous. By way of contrast, ΔV_t which captures the first-time emergence of new trading partners, and disappearance of existing ones, is certainly an exogenous event, at least for the gravity theory of trade. We therefore call ΔV_t the *pseudo*-extensive margin of world trade. However, the extent to which new potential trading relationships, ΔV_t, become active at once, i.e., the term $n_{t,t}$, must be seen as endogenously determined according to the gravity approach. This is exactly what the approach suggested below does.

Defining $N_t \equiv \sum_{h=t_0}^{t} n_{th} \zeta_{th} V_t$, we may now decompose the growth in world trade according to (1) as follows:

$$\Delta T_t = \Delta \bar{T}_t N_{t-1} + \Delta N_t \bar{T}_t = \Delta \bar{T}_t N_{t-1} + (\Delta N_{\mathrm{x},t} + n_{t,t} \Delta V_t) \bar{T}_t, \tag{5}$$

where all changes are from $t-1$ to t, and ΔN_t is taken from (4). Notice that 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 trading relationships.

3.3 Empirics of the dual margin

Figures 1 through 3 provide a succinct but comprehensive overview of the role that these margins have played in the post-World-War-II evolution of world trade in manufactures, based on the data set described above. Figure 1.a highlights the extensive margin, looking at the increasing utilization over time t of potential trading relationships for different groups of vintages. The leftmost 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 numbers are in natural logs. 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. The jump in the horizontal lines are exogenous movements on the pseudo-extensive margin, 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. For instance, of the maximum

number of trading relationships based on vintages up to 1950, about 40 percent had not yet become active by 1950. By 1959, the number of active relationships has increased by about 20 percent, narrowing the gap of unutilized relationships to 30 percent. Including vintages 1950-1960, we observe a 1960-jump in the maximum number of trading relationships by about 65 percent (relative to vintages up to 1950). Including these additional vintages, the number of active relationships has increased by about 40 percent between 1959 and 1960, which implies a larger gap of unutilized relationships equal to about 43 percent in 1960. Figure 1.a conveys a simple and important message: There was significant change at both types of extensive margins, with the 1950s and 1960s, as well as the 1990s, showing stronger movements on the extensive margin proper than the 1970s and 1980s. Moreover, there is still much ground to be covered at the extensive margin, even at the end of the sample in 1997.

Fig. 1.a: The extensive margin of world trade expansion 1950-1997
Actual and potential number of trading pairs by 'country vintage'

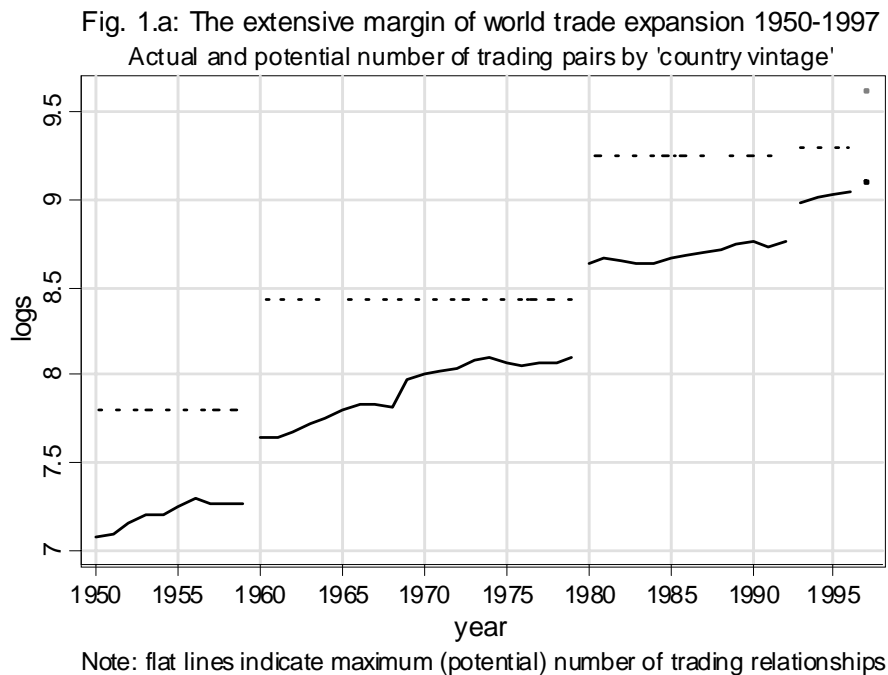Note: flat lines indicate maximum (potential) number of trading relationships

Figure 1.a deliberately ignores the volume of trade involved in movements at the extensive margin. Figure 1.b brings $\bar{T}_t$ into the picture, revealing that the evolution of world trade volume is heavily dominated by movements at the intensive margin. The bars indicate the overall volume of world trade (index of exports plus imports with 1950 set equal to 1), while the lines indicate $\sum_{h=t_0}^{H} \bar{T}_{t,h} N_{t,h}$ for $H = 1950, 1960, 1980, 1993$, in line with figure 1.a. The extent to which actual trade at any point t (on the horizontal axis) exceeds trade based

10

on vintages $h \leq H < t$ reveals the quantitative importance (judged by trading volumes) of movements that have taken place between H and t on the extensive margin. In other words, as opposed to figure 1.a, movements at the extensive margin are now weighted by the relative amount of trade involved. Moreover, in figure 1.b we can no longer identify the extensive margin proper, separately from the pseudo-extensive margin. For instance, by 1980 the volume of world trade had increased to about 12 times its 1950-level. The volume of trade based on vintages only up to 1960, the increase vis à vis 1950 was only about 11-fold, the remainder being due to vintages 1960-1980. While figure 1.b indicates that, judged by trade volumes, the intensive margin is more important, it must be emphasized that the extensive margin in and of itself, i.e., independent on the trade volume affected, is a key explanatory target of the gravity approach which has so far almost entirely been neglected. And figure 1.a clearly indicates that there is a lot of movement to be explained.

Fig. 1.b: The intensive and extensive margin of world trade expansion
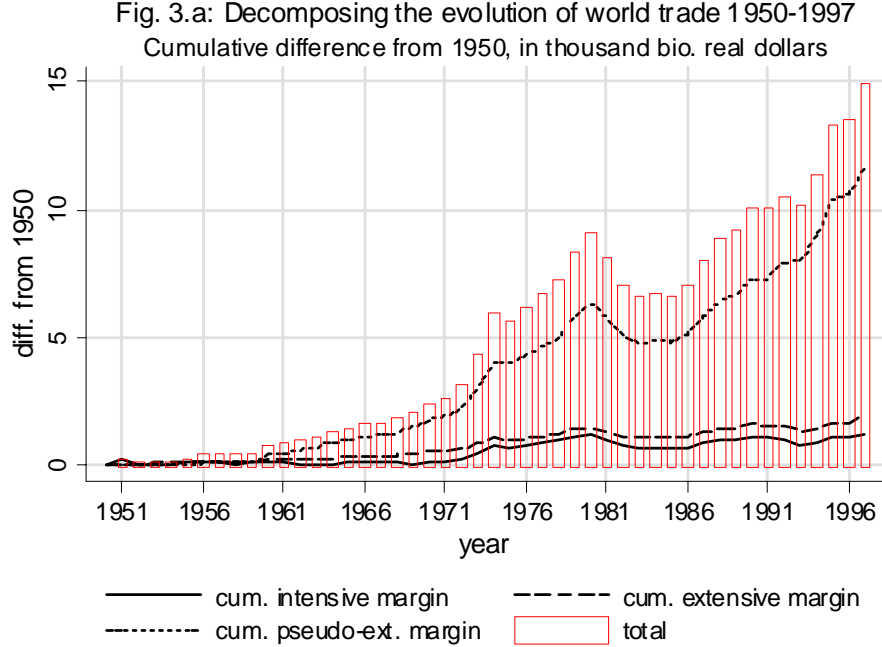Evolution of trade volumes by vintage
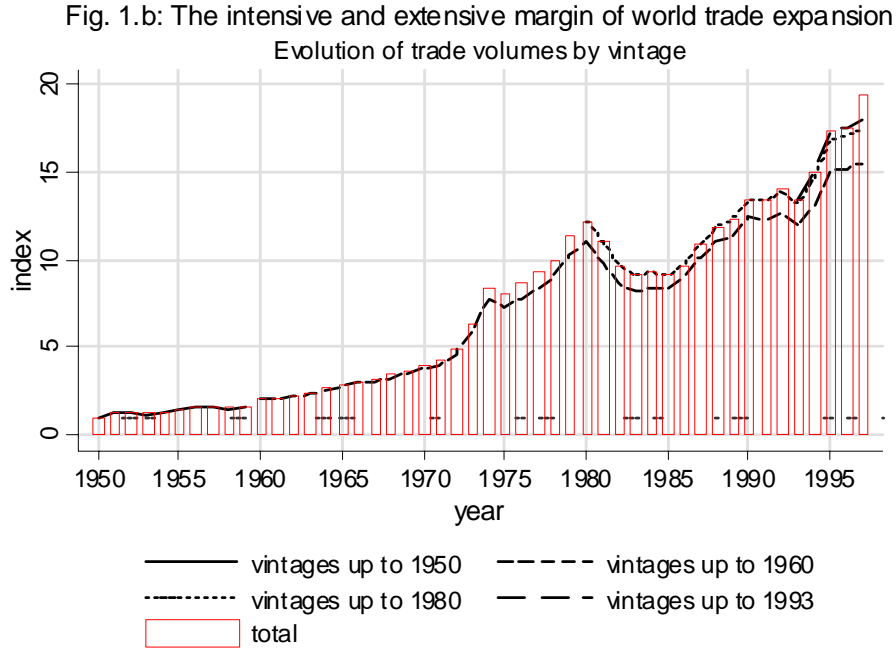
Figure 2 dissects the overall growth in world trade from 1950 up to 1997. It asks a very simple question: What is the contribution of vintages up to H towards the overall cumulative growth in world trade between 1950 and 1997? The bars indicate the frequency distribution, plotting the ratio of $\sum_{h=t_0}^{H} \bar{T}_{1997,h} N_{1997,h} - \sum_{h=t_0}^{1950} \bar{T}_{1950,h} N_{1950,h}$ to $\sum_{h=t_0}^{1997} \bar{T}_{1997,h} N_{1997,h} - \sum_{h=t_0}^{1950} \bar{T}_{1950,h} N_{1950,h}$ on the left-hand vertical axis, with H going from 1950 to 1997 on the

11

horizontal axis. Using the right-hand axis, the line depicts the growth rate for trade, again between 1950 and 1997, of vintages up to H, i.e., $\sum_{h=t_0}^{H} \bar{T}_{1997,h} N_{1997,h} \Big/ \sum_{h=t_0}^{1950} \bar{T}_{1950,h} N_{1950,h} - 1$, again with $H = 1950 \ldots 1997$ on the horizontal axis. Obviously, the bars and the line are just scalar transformations.

Fig. 2: Growth of world trade 1950-1997

vintage distribution of cumulative growth rates
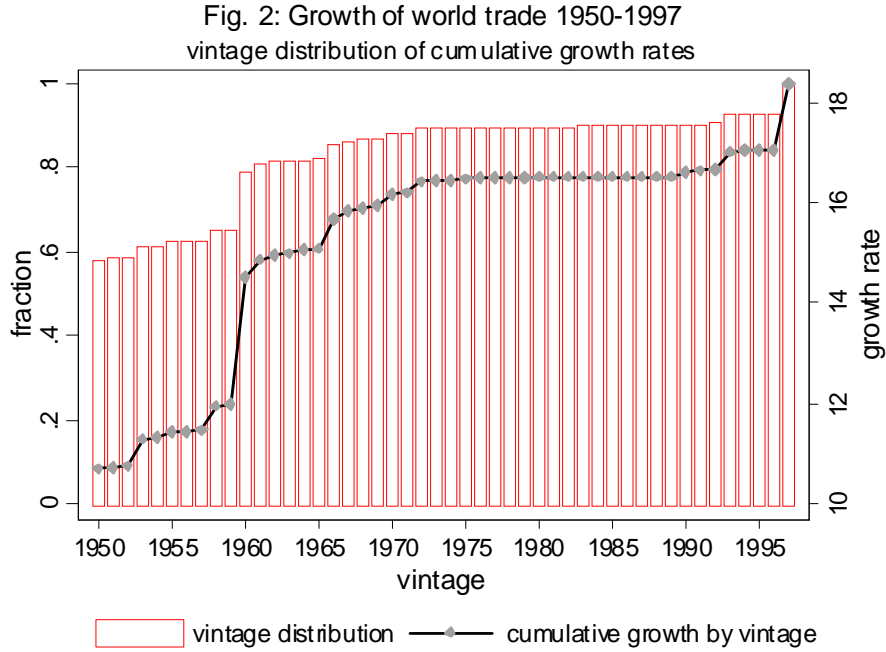
To understand figure 2, it is perhaps helpful to look at extreme reference cases. For instance, if all growth in world trade had happened entirely at the intensive margin, then the distribution function would be degenerate, with the all "mass" concentrated at vintages up to $H = 1950$. On the other hand, if all growth had been due in a completely symmetric way to movements at the extensive margin, then we would have a linear increase of the bars and a straight line for the growth rates. Moreover, vertical jumps at interior points indicate movements at the extensive margin (proper plus pseudo), while flat segments indicate prevalence of the intensive margin. More specifically, based on vintages up to 1950 the growth rate of world trade between 1950 and 1997 was about 1100 percent, contributing about 60 percent to cumulative growth during the entire sample period. Including vintages 1950-1960, the growth rate increases to about 1420 percent, contributing a further 20 percent to overall cumulative growth. The message conveyed is that the 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.

Fig. 3.a: Decomposing the evolution of world trade 1950-1997
Cumulative difference from 1950, in thousand bio. real dollars

Finally, figures 3 depicts a decomposition of world trade growth into its constituent parts according to equation (5) above. Plotting $t = 1959 \ldots 1997$ on the horizontal axis, figure 3.a reveals how the cumulative difference of actual trade from the initial volume for 1950, indicated by bars, is decomposed into changes at the respective margins, indicated by lines. More specifically the intensive margin line in figure 3.a depicts $(\bar{T}_t - \bar{T}_{1950}) \sum_{h=t_0}^{1950} N_{1950,h}$, while the line for the extensive margin proper plots $\bar{T}_t \left(\sum_{h=t_0}^{1950} N_{t,h} - \sum_{h=t_0}^{1950} N_{1950,h} \right)$, and the pseudo-extensive margin is plotted as $\bar{T}_t \left(\sum_{h=t_0}^{t} N_{t,h} - \sum_{h=t_0}^{1950} N_{1950,h} \right)$. By definition, the three lines add up to the bars. Figure 3.b plots the yearly growth rates at these same margins according to equation (5), dividing each component change by the actual volume of world trade in the previous period. Again, the lines must add up to the bars, indicating yearly changes in total world trade volumes. The first and important impression we obtain from figure 3.b is that growth rates do vary a lot across years, as does the relative importance of the three constituent margins. Extreme values of overall yearly growth tend to be dominated by movements at 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

13

margin did play a role, as evidenced by the difference between the bar-values and the solid line. Quite clearly, however, the big dent in world trade occurring in the early 1980s was due to a movement, not at the extensive, but at the intensive margin.

Figure 3.b: Decomposing growth rates of world trade
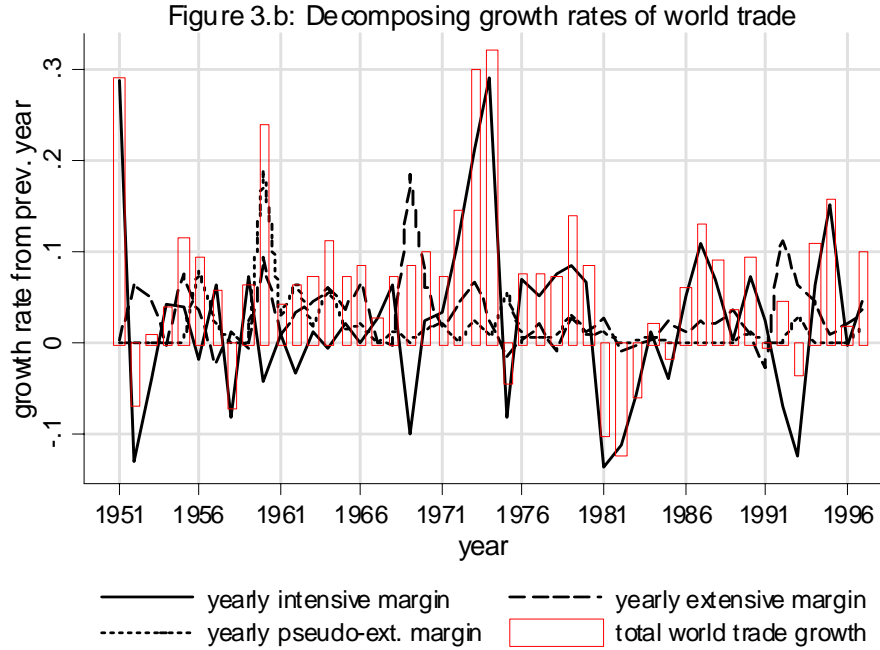
When looking at figure 3.a, one needs to be aware of the fact that the contribution of the extensive margins to the growth of trade from 1950 is weighted by the current average trade volume $\bar{T}_t$, while the contribution of the intensive margin receives base-period weights from the 1950 number of active trading relationships, i.e., $\sum_{h=t_0}^{1950} N_{1950,h}$. Given the large movements in average trading volumes, over time, the two extensive margins thus receive large weights which may, at first sight, cause a misleading impression of a dormant, rather than dominant behavior of the intensive margin, as suggested by figure 3.b. However, once the implication of weighting is realized, figure 3.b has a revealing interpretation. For instance, from 1950 to 1980, world trade has increased by about 9.3 thousand bio (real 1983) dollars. If all newly emerging trading relationships since 1950 that have become active by 1980 had been trading on the average 1980 level, $\bar{T}_{1980}$, then this margin alone (pseudo-extensive margin) would imply an increase by as much as about 6.5 thousand bio dollars. Taking the 1950-1980 increase in the number of active relationships only of vintages up to 1950 (extensive margin proper), adds a further increase by 1.4 thousand bio dollars. About that same increase is ob-

14

served on the intensive margin, assuming that the 1950 number of active trading relationships had increased their trading volume according to $\bar{T}_{1980} - \bar{T}_{1950}$.

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. Obviously, such models are of little help in explaining movements on the dual margin of world trade that we have identified above.

Common sense suggests that bilateral trade is importantly determined by natural barriers to trade, particularly trade costs related to geographic distance and transport. Indeed, the prime reason why traditional trade theory leaves bilateral trade undetermined is that it has so far almost entirely neglected all such costs. The important exception, of course, is the gravity approach where distance is usually added as a trade-inhibiting force, alongside the "economic mass" of two countries as the key force generating bilateral trade. It has been noted several times that something like this gravity force becomes apparent, whenever – and for whatever reason – 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). With comparative-advantage-based trade, it arises if 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).[10] The problem with the traditional approach based on product differentiation is that it fails to explain zero bilateral trade, which is an

---

[10] A full-fledged incorporation of trade costs into a generalized theory of comparative advantage has recently been provided by Deardorff (2004).

important characteristic of real world trade data, as emphasized above. Alternative theoretical foundations of the gravity equation that do allow for zero trade have been provided by Eaton & Kortum (2002), Feenstra et al. (2001) and Haveman & Hummels (2004).

In this section, we abstain from providing an explicit structural model of trade leading to the gravity equation, but simply work with a general reduced form relationship which, given the above mentioned literature, is consistent with a relatively broad range of models permitting zero bilateral trade as an equilibrium outcome. Our primary aim is to add an explicit, if rudimentary, treatment of the transport sector which helps us explore the role that distance plays in a gravity equation, with an eye on the dual margin of world trade

4.1 A simple gravity framework

In order to capture the possibility of zero bilateral trade, we specify the gravity model as a "corner-solutions-model". Denoting the volume of bilateral trade between country i and j at time t by T_{ijt}, we define a latent variable T^*_{ijt} which may take any value on the real line. Our model implies that only if this latent variable exceeds some threshold variable $\bar{T}_{ijt}$, will the respective country pair report positive trade. More specifically, we stipulate

$$T^*_{ijt} = \frac{Y_{it}Y_{jt}}{\sum_k^N Y_{kt}} \cdot \Pi(\mathbf{S}_{ijt}) \cdot F\left[C\left(\mathbf{D}_{ij}, \mathbf{P}_t, t\right)\right], \tag{6}$$

$$T_{ijt} = \max\left(0, T^*_{ijt} - \bar{T}_{ijt}\right). \tag{7}$$

In the remaining analysis, we allow for a constant term in (6) and set $\bar{T}_{ijt} = 0$. It should be noted at the outset that in this framework globalization may show up in a change in the the elasticities underlying (6), as well as through a change in the threshold level $\bar{T}_{ijt}$ in equation (7).

Our maintained hypothesis thus states that latent trade between countries i and j is governed by the traditional *gravity force*, i.e., the "economic mass" measured by product of GDP of the trade partners, $Y_{it}Y_{jt}$, relative to world GDP, $\sum_k^N Y_{kt}$. For the sake of simplicity, we resort to the base-line specification where the trade elasticity with respect to gravity is

equal to unity.[11] Note that the number of countries, N, does not necessarily coincide with the number of *trading* countries.

In addition to gravity, bilateral trade is also influenced by *structural characteristics* of the two economies. We denote these by a vector of determinants $\mathbf{S}_{ijt}$, entering the multiplicative gravity equation through some functional form $\Pi(\cdot)$, which we need not further specify for the present purpose. In addition to the two countries' trade policy stance, such as their membership in regional trade agreements or the WTO and more direct measures of their trade policies, $\mathbf{S}_{ijt}$ also includes structural differences between the two countries, such their relative factor endowments. A theoretical justification for this is found in Helpman & Krugman (1985, ch.8) who show that in a model with a homogeneous good and a differentiated-goods-sector, and with Heckscher-Ohlin-driven inter-industry trade, any given volume of trade is consistent with alternative combinations of mass-attraction (dispersion of endowment levels across countries) and structural differences (endowment ratios) between countries. More specifically, a lower mass-attraction from a less even country-dispersion may be compensated by an increase in structural difference. In the above formulation, this would be captured by the corresponding elasticity of the function $\Pi(\cdot)$. In the following general discussion, we write these elasticities in vector form as $\boldsymbol{\pi}$.

Finally, we allow *trade costs* $C(\mathbf{D}_{ij}, \mathbf{P}_t, t)$ to enter the gravity equation through some functional form $F(\cdot)$. In turn, trade costs are determined by geographical and cultural *distance*, grouped together in the time-invariant vector $\mathbf{D}_{ij}$, and by prices for crucial inputs in transportation, such as the price of oil and the rental price of transport and communications equipment. We write these prices in vector form as $\mathbf{P}_t$. The expression $C(\mathbf{D}_{ij}, \mathbf{P}_t, t)$ is interpreted as a minimum cost function for overcoming the distance $\mathbf{D}_{ij}$ for the sake of trade, given relevant prices and technology prevailing at time t. We narrow our focus on geographical distance, so that $\mathbf{D}_{ij}$ becomes a scalar. Moreover, we assume that the production function for transport services is constant returns to scale and technological progress in the transport sector is Hicks neutral and independent of distance. Writing A_t for the productivity level in

---

[11]See Evenett & Keller (2002) for a discussion of alternative cases where incomplete specialization leads to a coefficient of less than unity.

transportation, we then have

$$C_{ijt} = C\left(D_{ij}, \mathbf{P}_t, t\right) = D_{ij} A_t^{-1} c\left(\mathbf{P}_t\right), \tag{8}$$

where c has the usual interpretation of a minimum unit-cost function. Substituting this relationship into the gravity equation and taking logs, we obtain

$$\ln T_{ijt}^* = -\delta \ln A_t + \ln\left(\frac{Y_{it} Y_{jt}}{\sum_k^N Y_{kt}}\right) + \boldsymbol{\pi} \ln \mathbf{S}_{ijt} + \delta \ln D_{ij} + \delta \boldsymbol{\alpha} \ln \mathbf{P}_t, \tag{9}$$

where δ is the elasticity of trade with respect to trade costs C_{ij}, and $\boldsymbol{\alpha}$ is a vector of cost shares for oil and equipment, as implied by the production function for transport services.

Equation (9) shows how the trade-effect of technological improvements in transport and communication might be reflected in estimation results. Typically, the productivity level A_t will not be observable directly. Hence, one might be tempted to recover this effect from the residuals in an estimation of a gravity equation like (9), with $-\delta \ln A_t$ excluded and focusing on observations where $T_{ijt} > 0$, as usual in the literature. More specifically, provided that the estimation controls for all observable determinants of bilateral trade that vary *only* along the time dimension, such as $\mathbf{P}_t$, estimates of time-varying intercepts might be equated with $-\delta \ln A_t$.[12] In addition, one may wish to allow for the elasticity δ to parametrically vary across time, as in Brun et al. (2004), and to incorporate this information when recovering A_t from the time-varying intercepts. Provided $\delta < 0$, any increase in A_t over time would boost bilateral trade, as expected.

However, since equation (9) relates to latent trade T_{ijt}^*, any procedure based on actual observations where $T_{ijt} > 0$ is questionable. We shall return to this in more detail in the following subsection. Intuitively, against the background of the extensive margin of world trade mentioned above, it is easy to recognize that there is an additional channel which would escape any such procedure. In equation (7), if the term $\bar{T}_{ijt}$ is allowed to vary along the time dimension, it essentially captures the same information as $-\delta \ln A_t$. Hence, an increase in A_t is equivalent to a reduction in $\bar{T}_{ijt}$, the threshold level of latent trade, beyond

---

[12] This is the view taken, for instance, by Buch et al. (2004).

which positive trade occurs. In other words, technological improvements in transport and communications are also directly related to the extensive margin of international trade, which should, therefore, be explicitly taken into account by an appropriate econometric strategy.

4.2   Econometric issues

From an econometric point of view, one might be tempted to treat movements on the extensive margin of world trade as a sample selection problem. This would imply that we have a separate theory explaining the establishment of a trading relationship between any two countries, independent on the gravity equation which explains the extent of trade, once such a relationship exists. It is not obvious however, why the establishment of trading relationships should be governed by "laws" fundamentally different from those of the gravity approach. A preferable approach is to rely on a unified gravity model to explain both, the emergence of a trading relationship where none exists to start with (extensive margin of world trade), and the extent of trade where it exists (intensive margin). This is, indeed, what we have done above.

The corresponding statistical model is what Wooldridge (2002) calls the *corner-solution model*. This is a special case of the *censored regression*, and it allows for a full-fledged analysis of the extensive and the intensive margin of world trade expansion within a unified gravity approach.[13] For simplicity we now use $\mathbf{X}_{ijt}$ to denote the vector of all exogenous explanatory variables (all in natural logs) in an equation like (9), although some of these (like distance) are time-invariant. We refer to these variables, collectively, as "the gravity force". The underlying statistical model is written as

$$\left(\ln T^*_{ijt} - \mathbf{X}_{ijt}\boldsymbol{\beta}\right)\big/\,\sigma \sim \mathrm{N}[0,1], \tag{10}$$

where $\boldsymbol{\beta}$ is a vector of gravity-related parameter, and $\mathrm{N}[0,1]$ denotes the standard-normal

---

[13]It should perhaps be noted that one should not look at the problem as one of a mere *sample defect*, as in the usual censored regression context. Specifically, the problem is not one of erroneously coding "negative trade" with $T_{ij} = 0$, as would be the case in a censored distribution with bottom-coding at $T^*_{ij} = 0$. Instead, the appropriate interpretation of the model is one where the endogenous outcome includes the possibility of a *corner solution* with $T_{ij} = 0$, depending on the explanatory variables.

distribution.[14] In statistical terms, the mean of *latent* trade $\ln T_{ijt}^*$, conditional on gravity forces $\mathbf{X}_{ijt}$, is equal to $\mathbf{X}_{ijt}\boldsymbol{\beta}$, with variance equal to σ^2. Writing the model in this way allows a direct application of results on censored standard-normal distributions.

Since $\ln(0)$ is not defined, we proceed by replacing zero-trade by $T_{ijt} = 1$. An immediate consequence of this statistical model 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 $\ln T_{ijt} = 0$. One might try to address this problem by means of nonlinear least squares (NLS), as in Silva and Tenreyro (2003) or Coe et al. (2002). However, this poses several problems. First, since $\ln T_{ijt}$ includes corner outcomes, $\ln T_{ijt}|\mathbf{X}_{ijt}$ is probably heteroskedastic, which renders NLS inefficient.[15] Using weighted NLS requires an arbitrary choice of a specific model for the conditional variance, $\mathrm{var}(\ln T_{ijt}|\mathbf{X}_{ijt})$, and would thus seem questionable; see Wooldridge (2002, p. 518 ff). Secondly, and more importantly, the coefficients obtained by NLS estimation of a model for $\mathrm{E}(\ln T_{ijt}|\mathbf{X}_{ijt})$ are difficult to interpret. By definition, such a model would not allow us to estimate any feature of the distribution of $\ln T_{ijt}|\mathbf{X}_{ijt}$, other than its mean. Specifically, it does not allow us to isolate the effects corresponding to the intensive and the extensive margin of world trade. In the present context, the extensive margin is captured by $\Pr[\ln T_{ijt} \geq 0|\mathbf{X}_{ijt}]$, while the intensive margin is captured by $\mathrm{E}[\ln T_{ijt}|\mathbf{X}_{ijt}, \ln T_{ijt} > 0]$.

For a better understanding of the results below, it is worth exploring in somewhat more detail the relationships between the parameters $\boldsymbol{\beta}$ and the effects that we would associate with our two margins. First, we may note that the *overall effect* is given by

$$\mathrm{E}\left(\ln T_{ijt}|\mathbf{X}_{ijt}\right) = \Phi\left(\frac{\mathbf{X}_{ijt}\boldsymbol{\beta}}{\sigma}\right)[\mathbf{X}_{ijt}\boldsymbol{\beta} + \sigma\lambda(z)], \quad \text{with} \tag{11}$$

$$\frac{\partial\,\mathrm{E}\left(\ln T_{ijt}|\mathbf{X}_{ijt}\right)}{\partial X_{ijt}^r} = \beta^r \Phi\left(\frac{\mathbf{X}_{ijt}\boldsymbol{\beta}}{\sigma}\right), \tag{12}$$

where $z \equiv \mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma$, and a superscript r denotes a specific explanatory variable r. The term

---

[14] We shall later return to a separation between country-specific explanatory variables and year-specific 'fixed' effects (intercepts), which are meant to control in a non-parametric way for global events, such as oil-price shocks, the global business cycle, or the efficiency of the transport sector, A_t.

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

$\lambda(z)$ 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.[16] Note that, given (10), the conditional probability of positive trade, $\Phi(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$, is less than 1, which gives rise to *attenuation*.[17] The reason for this is straightforward. Suppose $\beta^r > 0$ and we consider a reduction in X_{ijt}^r. The marginal effect on the conditional mean of the uncensored latent 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 (12) in the gravity context is not straightforward, as it is an amalgam of the intensive and the extensive margin. It gives the marginal effect of gravity forces on the expected volume of trade, allowing for corner-solutions, but without explicitly separating corner cases (extensive margin) from interior cases (intensive margin).

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

$$\mathrm{E}\left(\ln T_{ijt} \left| \mathbf{X}_{ijt}, \ln T_{ijt} > 0\right.\right) = \mathbf{X}_{ijt}\boldsymbol{\beta} + \sigma\lambda(z), \quad \text{with} \tag{13}$$

$$\frac{\partial\,\mathrm{E}\left(\ln T_{ijt} \left| \mathbf{X}_{ijt}, \ln T_{ijt} > 0\right.\right)}{\partial X_{ijt}^r} = \beta^r - \beta^r\lambda(z)\left[z + \lambda(z)\right], \tag{14}$$

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

$$\Pr\left(\ln T_{ijt} \geq 0 | \mathbf{X}_{ijt}\right) = \Phi\left(\frac{\mathbf{X}_{ijt}\boldsymbol{\beta}}{\sigma}\right), \tag{15}$$

where the marginal coefficient, $\partial\Pr\left[\ln T_{ijt} \geq 0 | \mathbf{X}_{ijt}\right]/\partial X_{ijt}^r$, may be derived by differentiating

---

[16] See Greene (2003, p. 762-3). The first of the two terms in (11) is the discrete part of the censored distribution, measuring the probability mass at zero, while the second, bracketed term is the conditional mean of the corresponding truncated normal distribution. Note that $\lambda(y)$ gives the hazard function of the standard-normal distribution.

[17] 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 $\Phi(-\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma)$, and thus the probability of positive trade.

tiating the equality

$$E\left(\ln T_{ijt}\,|\mathbf{X}_{ijt}\right) = \Pr\left(\ln T_{ijt} \geq 0|\mathbf{X}_{ijt}\right) \cdot E\left(\ln T_{ijt}\,|\mathbf{X}_{ijt}, \ln T_{ijt} > 0\right), \qquad (16)$$

and invoking the associated marginal effects; see Wooldridge (2002, p. 523). Note that the marginal effects are not constant on either of the two margins. We shall return to this when presenting estimation results below.

It should be noted that the usual procedure of running OLS regression on the subsample for which $\ln T_{ijt} > 0$ results in *inconsistent* estimates. From equation (13), we may write

$$\ln T_{ijt} = \mathbf{X}_{ijt}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{X}_{ijt}\boldsymbol{\beta}}{\sigma}\right) + \varepsilon_{ijt} \quad \text{with} \quad E\left(\varepsilon_{ijt}\,|\mathbf{X}_{ijt}, \ln T_{ijt} > 0\right) = 0. \qquad (17)$$

Obviously, running OLS of $\ln T_{ijt}$ on $\mathbf{X}_{ijt}$ for $\ln T_{ijt} > 0$ amounts to omitting the variable $\lambda\left(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma\right)$ in the above equation. If the covariance between $\mathbf{X}_{ijt}$ and $\lambda\left(\mathbf{X}_{ijt}\boldsymbol{\beta}/\sigma\right)$ is non-zero, then the coefficients $\boldsymbol{\beta}$ are inconsistently estimated. Moreover, regressing $\ln T_{ijt}$ on $\mathbf{X}_{ijt}$ using all of the data, i.e., including the observations where $T_{ijt} = 1$, will not consistently estimate $\boldsymbol{\beta}$ either, since $E(\ln T_{ijt}\,|\mathbf{X}_{ijt})$ is nonlinear in $\mathbf{X}_{ijt}$, $\boldsymbol{\beta}$ and σ.

It is worth considering very briefly the bias involved in the usual procedure by looking at the simple univariate case, using y and x to denote the dependent and independent variable. The model then becomes

$$E\left(y|x, y > 0\right) = \beta x + \sigma\lambda\left(\frac{\beta x}{\sigma}\right). \qquad (18)$$

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

$$\hat{\beta} = \beta + \sigma\frac{\text{cov}\left[x, \lambda\left(\beta x/\sigma\right)\right]}{\text{var}\left(x\right)}. \qquad (19)$$

Since $\lambda\left(c\right) = \phi\left(c\right)/\Phi\left(c\right) > 0$ and $\lambda'\left(c\right) = -\lambda\left(c\right)\left[c + \lambda\left(c\right)\right] < 0$, we know that $\text{cov}[x, \lambda\left(\beta x/\sigma\right)] > 0$. Moreover, we know that in a sample where no corner-solutions appear $E(y|x, y > 0) = E(y|x) = \beta x$. Hence, since $\sigma \cdot \text{cov}[x, \lambda\left(\beta x/\sigma\right)]/\text{var}(x) > 0$, we conclude that $\hat{\beta} > \beta$. The omitted-variable-bias causes an over-estimation of β. One may, however, argue that 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 towards β as time unfolds.

5 Estimation results

5.1 Data

For the left-hand side trade data, we may refer to the extensive description in section 2 above. For the right-hand side variables, we use standard data sources. GDP in purchasing power parities is taken from the Penn World Tables 6.1, and (where missing) from the World Development indicators. Data on geographic distance is taken from the CEPII homepage.[18] Data on whether a country is an island or is landlocked has been taken from the CIA fact book (2003), which also contains information on land surface, and the fraction of land surface that can be exploited for agricultural production. Countries' memberships in diverse regional trade agreements or the WTO has been taken from the websites of the respective organizations. We also use the Sachs-Warner index, as updated by Wacziarg and Welch (2003), as a proxy for the overall restrictiveness of a country's trade policy.

In our regressions, we also include a *multilateral resistance index*. An intuitive justification for this is that the gravity force between any two countries not only depends on how far apart they are from each other, but also on how far apart each of them is with respect to other countries. A more rigorous justification is found in Anderson & van Wincoop (2003).[19] In line with Brun et al. (2004), we define this index as $R_{it} = \sum_j w_{jt} D_{ij}$ for $i \neq j$, and with $w_{jt} = Y_{jt} / \left(\sum_j Y_{jt} \right)$ for all j. A similar variable is computed for country j.

Table 1 provides the mean and standard deviation of our main variables, measured at the starting year of our panel (1950), at the final year (1997), and over the entire time span of the sample. We report only on observations for which none of our variables contain missing values. The natural logarithm has been applied to continuous variables. We have constructed a proxy to capture differences between two countries labor-capital ratio by computing a structural differences index defined as the squared difference between two countries' real GDPs per capita. This measure is meant to capture Heckscher-Ohlin type determinants of trade flows. The product of the fraction of arable land in total surface is

---

[18] See http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

[19] See also Feenstra (2004, ch.5), Harrigan (2004), and Anderson & van Wincoop (2004).

meant to proxy the ease of goods transport within a country. The RTA dummy takes the value of one if both countries are in one of the following regional trade agreements: EU, EFTA, NAFTA, MERCOSUR, CACM (Central American Common Market), CARICOM (Carribean Common Market), ASEAN, USIS (US–Israel free trade agreement) and AUSD (Australian–New-Zealand free trade agreement).

Our panel is unbalanced because the number of sovereign countries (as defined by the UNO) has increased greatly since 1950. This increase has come in two major waves: in the early sixties, a large number of formerly colonialized African and South East Asian countries have become independent. In the mid nineties, the break-ups of some formerly communist countries has added again to the number of dyads, N. Note that this number increases approximately quadratically in the number of countries C since $N = C\,(C-1)\,/2$. In 1950, $C = 76$ so that $N = 2850$ while in 1997 $C = 181$ so that $N = 16290$. Thus, over the last half-decade, the number of potential bilateral trading relationships has increased by more than 570%. Unfortunately, some of the relevant explanatory variables listed in table 1 are missing, so that our sample comprises 1,239 country pairs (out of 2,850 possible ones) in 1950 and 11,993 (out of 16,290) in 1997. On average, we have about 6,599 observations per year.

5.2 An illustration of the distance-puzzle

Before turning to a consistent estimation of the corner-solution model by means of a Tobit approach, we want to provide a concise illustration of what has been referred to as the distance-puzzle above. To do so, we estimate (9) with $T_{ijt}^{*} = T_{ijt} > 0$, sticking to the widely accepted method of Rose (2004) and many others, i.e., relying on pooled OLS. Our basic model is straightforwardly found from (9), where all covariates which vary only with time are captured by inclusion of time-varying intercepts, denoted by a vector $\mathbf{K}_t$. More specifically, these intercepts are meant to control in a non-parametric way for year-specific global events, such as oil price shocks, the global business cycle, or – of special importance in our context – the efficiency of the transport sector. To illustrate the time dependency of the distance coefficient, we interact the year-specific effects with distance:

$$\ln T_{ijt} = \boldsymbol{\delta}_t[(\ln D_{ij})\mathbf{K}_t] + \gamma \ln (Y_{it}Y_{jt}) + \boldsymbol{\pi} \ln \mathbf{S}_{ijt} + \mathbf{K}_t + \varepsilon_{ijt}, \qquad (20)$$

where ε_{ijt} is an error with the usual properties. The role of distance is reported through a vector of time-varying coefficients $\boldsymbol{\delta}_t$, and γ is the elasticity of joint GDP, which is now freed from unity, while $\boldsymbol{\pi}$ is a vector of elasticities with respect to structural characteristics, as noted in section 3 above. Note that with the presence of time-varying intercepts $\mathbf{K}_t$, there is no scope left to include pure time-variant variables, such as the price of petroleum or transport equipment, and world GDP. In line with Rose (2004), we assume that observations are independent across country-pairs (groups), but not necessarily independent within groups. We adjust the variance-covariance matrix in order to capture within-group correlation.

One might also believe that, despite the gravity model's impressive successl in explaining the magnitude of bilateral trade, there are still important country-pair specific determinants not explicitly covered by (20) above. The error term then becomes $\varepsilon_{ijt} = v_{ij} + u_{ijt}$, and the appropriate method of estimation is a random effects model, as for instance used in Brun et al. (2004). However, our prime interest lies in the behavior of $\boldsymbol{\delta}_t$ over time, hence it seems equally important to carefully control for year-specific effects. Our reading of the existing literature is that this has not always been done.

Figure 4 shows the time path of δ. As in Brun et al. (2004), we observe that the (absolute value) of the coefficient has almost doubled in the time span from 1950 to 1997. The black solid line in the figure is based on estimation of (20) by robust OLS. The dashed line represents an estimation where unobserved heterogeneity is taken into account, applying a random effects panel estimator. Finally, the grey solid line depicts the distance coefficient in (20), estimated separately for each year in the sample. Note that, unlike the two other procedures, this allows for time-variation in γ and $\boldsymbol{\pi}$. We control for a host of country-specific determinants of bilateral trade, some of which are also time-variant. All of them are standard in the gravity model. From figure 4, we conclude that the puzzle is robust to the obvious modifications of the estimation technique, and that the change in δ over time is non-trivial in size.

Before turning to a brief interpretation of this result, we test for the existence of a time trend in δ by parametrizing the time-distance interaction, whence the regression equation becomes

$$\ln T_{ijt} = \delta_0 \ln D_{ij} + \delta_1 (\ln D_{ij})t + \tau t + \gamma \ln (Y_{it} Y_{jt}) + \boldsymbol{\pi} \ln \mathbf{S}_{ijt} + \mathbf{K}_t + \varepsilon_{ijt}. \qquad (21)$$

Clearly, with the presence of a time trend τt, the matrix of time-specific effects, $\mathbf{K}_t$, captures effects that are different from those in equation (20) above. Table 2 shows the results of this regression. Although we use the same methodology and (almost) the same data as Rose (2004), the vector of controls $\mathbf{S}_{ijt}$ differs somewhat from his, because we explicitly control for the degree of trade friendliness of institutions, as dichotomized by the Sachs-Warner index. Moreover, we also control for armed conflicts between countries, and for whether countries are ravaged by civil wars. Note that we are at this stage still working with a sample that does not contain zero bilateral trade observations, so that the armed conflicts controls turn out to be statistically insignificant. However, once we turn to the Tobit approach below, this will change.

Figure 4

The elasticity of bilateral trade with respect to distance over time; 1950-1997
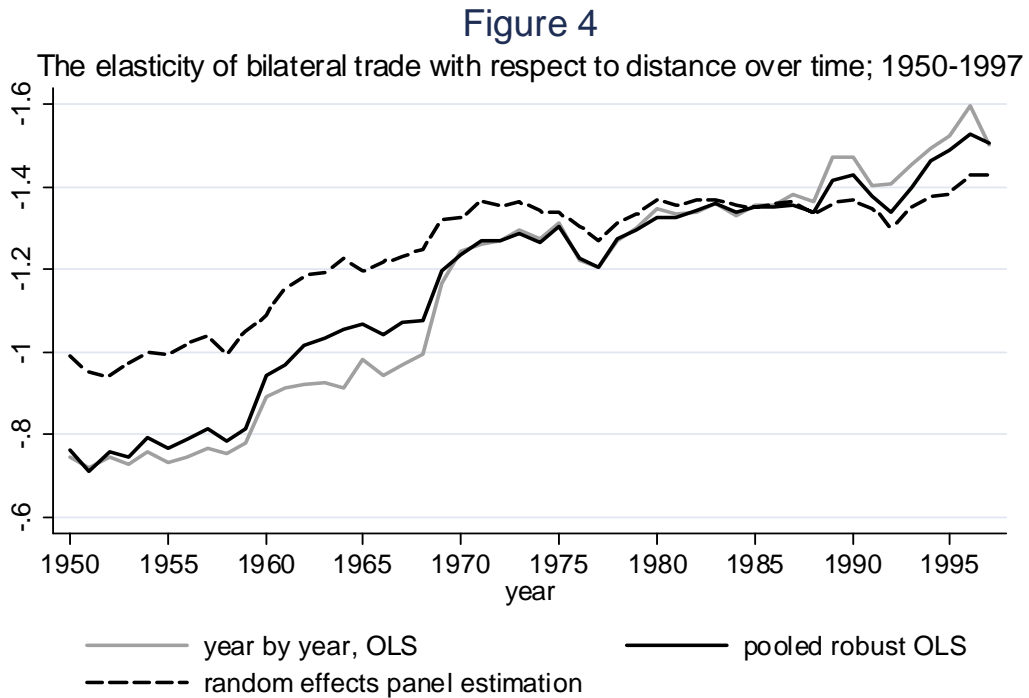
Column (1) in Table 2 reproduces a standard gravity equation, such as the one found in Rose (2004). While our sample is not exactly the same has Rose's, the point estimates of our regression and the overall fit (adjusted R^2) are very close to Rose's.[20] Note that

---

[20] We should add here that some of our additional regressors, such as the Sachs-Warner index, are not

we suppress year-specific effects in the output. Column (2) reports the result of the most parsimonious model incorporating the time-distance interaction. The coefficient δ_1 is very precisely estimated. Its absolute value grows from 0.8 in 1950 to 1.4 in 1997. Column (3) adds additional information on geographical barriers that may impede trade between two countries. Column (4) supplements the set of regressors by variables describing the stance of trade policy (the Sachs-Warner index, membership in regional trade agreements, membership in the WTO) and international relations (such as whether the countries are currently fighting a war, or suffer from a civil war).

All of these modifications do not change the time effect on δ significantly. Column (5) takes out the year-specific intercepts and replaces them with the price of petroleum and real world GDP. Again, this modification has only very minor effects on the significance and size of the estimates. Finally, column (6) moves away from pooled OLS and runs (21) using the random effects estimator. This modification increases the distance coefficient at the beginning of the sample, but reduces the rate at which δ changes over time. At the end of the sample, the coefficient is 1.5. While the distance-puzzle in terms of the time-variation of δ thus persists if we control for unobserved heterogeneity, the reason why the effect of distance as such is larger in this case seems obvious. One major unobserved variable, of course, is the quality of inter-country transport infrastructure. If this variable has a positive effect on bilateral trade flows but is negatively correlated with distance, then OLS under-estimates δ. If the negative correlation between infrastructure and distance weakens over time with the establishment of intercontinental flights, improved telecommunication, etc., then the bias in the estimated δ falls over time.

Having run a wide variety of sample sensitivity checks, trying an almost exhaustive permutation of the variables at the right-hand-side of our regression, we must thus conclude that the distance-puzzle seems very robust to changes in methodology, sample composition, or the specification of the gravity equation. All of this, however, holds for the case where we do not consistently differentiate between the two margins of world trade: the extensive and the intensive margin. Therefore, we now return to the corner-solutions model introduced above,

---

available for all of his observations.

and apply it to our data using the Tobit estimation approach.

5.3 Results from estimating a corner-solutions model

In section 2 above, we have emphasized the empirical importance of the dual margin of world trade, the extensive and the intensive margin. In section 3, we have seen that incorporating the dual margin in a unified gravity model, called the corner-solutions model, raises statistical problems with applying OLS or simple NLS procedures. In this section, we therefore resort to the Tobit approach which is well established as the appropriate procedure for censored regression problems. We present estimation results based on the above mentioned gravity specification, using the data set described above.

Table 3 reports the Tobit estimation results. For reference purposes we include column (3) of table 2, reporting the results obtained using only strictly positive trade observations and employing OLS estimation. Column (2) expands the sample to include zero trade observations. While OLS is not a consistent way to estimate the gravity equation when there is a large mass of zero trade observations, it nevertheless serves a useful purpose as a reference case. Clearly, including zero observations changes most coefficients dramatically. This is not surprising, given what we already know from section 2 above, and given the fact that the sample size is now almost twice as large as before. As in Silva and Tenreyro (2003), who use a Poisson model to estimate a non-linear gravity equation for the year 1990, the coefficient on the product of GDP is no longer close to unity. This is evidence against the canonical form of the gravity model which implies a coefficient of unity. But it is not something we want to explore any further, since there are well-established versions of the gravity equation which allow for a non-unitary income elasticity, as we have mentioned in section 3 above. Also, the coefficient on the product of land area turns positive, while it is negative in the standard model.[21]

---

[21] There is a plausible explanation to this result: If the income elasticity of bilateral trade is close to unity, country size as measured by GDP is irrelevant for trade. On face value, this is at odds with the stylized fact that larger countries tend to be more closed. However, land area is positively correlated to GDP and enters negatively in the standard regression. Thus, the results are compatible with a negative size/openness relation. If, however, the income elasticity is below unity, as in our results, larger countries trade less. In order to replicate the negative size/openness correlation, our model does not require a negative effect of geographical size on bilateral trade. In other words, collinearity between geographical size and GDP makes it difficult to

28

A more important point, however, relates to the coefficient on GATT- or WTO-membership. Rose (2004) has found it hard to conclude from the standard model that membership has a statistically significant and positive effect, as one might perhaps expect – and the WTO might hope. Column (2) of table 3 reveals a coefficient which is quite large and strongly positive. Turning to the distance-puzzle, we note that the total effect of distance on trade is not wildly different in column (2) from column (1): at the end of the time window, i.e. in 1997, time t is equal to 48, and the estimated elasticity of distance is -1.45 and -1.36, respectively. Setting aside the difference in overall levels, however, once zero trade observations are included, there is a strong *positive trend* in the (negative) coefficient of distance. Hence, there is some evidence already at this stage that the inhibiting force of distance has, in fact, *fallen* over time.[22]

Notice that the only difference between columns (2) and (1) in table 3 is that (2) also looks at the extensive margin. We know already from section 2 above that the extensive margin of globalization is important in terms of descriptive empirical evidence. What table 3 tells us in column (2) is that including this margin in gravity estimations is also important in terms of the distance-puzzle and with respect to WTO-membership. However, in view of section 3 above, we should be aware of the fact that column (2) does not yet incorporate the extensive margin in a methodologically sound way.

Results from a sound estimation, based on the Tobit approach, are found in columns (3) through (6) of table 3. All estimates reported are *marginal* coefficients, and a very high of precision allows for comfortable statistical inference. First, column (3) shows the marginal coefficients of the latent linear model that underlies the Tobit model. These are obviously difficult to interpret. In columns (4) through (6), the corner-solution is appropriately taken into account. As we know from the previous section, the model is inherently non-linear, so that the marginal effects are not constant and need to be evaluated at some sample point. We have chosen to go the conventional way and use the means of all dependent variables to

---

isolate the effect of either of these variables.

[22] One should perhaps mention that there is anecdotal, and even more formal evidence on the evolution of distance related trade costs which is consistent with a positive trend in the distance coefficient, as apparent in column (1); see Hummels (1999). But extraneous information of this kind is inadequate to discriminate between columns (1) and (2), since, based on section 3 above, we have strong methodological reasons to question the validity of *both* column (1) *and* column (2).

do this evaluation.[23]

What we are looking for is a direct handle on our two margins of world trade. The *intensive margin* has been identified above as the marginal effect on the mean of trade, conditional on the gravity force and the existence of a trading relationship; see equation (14) above. The estimated coefficients are listed in column (4) of table 1. The *extensive margin*, on the other hand, is associated with the marginal effect on the probability of an existing trading relationship, conditional on gravity forces; see equation (15) above. The corresponding estimates are given in column (6) of table 1. The *overall* effect on the mean of trade, conditional on gravity alone, is given in column (5) in the middle, corresponding to equation (12) above. Clearly, columns (3) through (6) share the same t-values, and the signs of the coefficients have to be identical. Note, however, that the decomposition (16) is valid only for levels, and multiplying the conditional partial derivative and the probability of observing a positive trade volume does not deliver the unconditional partial derivative.

What are the core results of a consistent Tobit estimation, vis à vis OLS in column (2)? For the sake of time and space, we restrict ourselves to a brief focus on the role of distance. Looking at column (4), the *intensive margin*, we realize that OLS over-estimates the effect of distance towards the end of the sample. For instance, for 1997 we obtain an estimate of -1.45, compared to -1.26 for the Tobit result. On the other hand, it under-estimates the trade inhibiting force of distance in earlier periods of the sample. Thus, for 1950 we obtain a coefficient of -0.78 versus -1.4. This result is broadly in line with Silva and Tenreyro (2003), who also conclude for 1990 that the standard model overestimates the role of distance. Moreover, the product of GDPs now (i.e., with Tobit) appears with a very low marginal coefficient, indicating that countries with large GDPs trade less compared to countries with small GDPs, everything else equal. In absolute values, the coefficients estimated in column (4) are larger than those estimated in (2). This is true in particular for the trade policy proxies.

Turning to the probability that a strictly positive trade flow is observed between a country pair, i.e., the *extensive margin*, we observe from column (6) that distance plays an ever

---

[23]In computing the partial derivatives, sums of binary variables which take integer values over the interval $[0, 2]$, are treated as continous variables.

decreasing role. Thus, at mean distance, increasing distance by one percent lowers the probability of an existing trading relationship by 0.32 percentage points in 1950, while in 1997 this marginal effect has come down to 0.28 percentage points.

6   Conclusions

We have started out by drawing attention to a twofold dimension of growth in world trade. At any point in time, countries that already have established a trading relationship may intensify their bilateral trade, and where a bilateral trading relationship has not yet been established it may be taken up for the first time. We have argued that the latter dimension has found inadequate attention in theoretical and empirical research, particularly regarding the so-called gravity approach which attempts to explain the volume of bilateral trading relationships.

Our aim in this paper is to explore this "unchartered territory". In a first step, we have provided systematic empirical evidence on the relative importance of the two dimensions, which we have called the extensive and the intensive margin of world trade. Towards this end, we have introduced a vintage-notion for a trading relationship. Using this tool in dissecting the evolution of world trade from 1950 to 1997, we have found significant action on both margins, with the extensive margin still far from being fully exploited even towards the end of the 1990s. Moreover, we find evidence for episodes where significantly more action took place on the extensive than the intensive margin, and vice versa. We have taken this as an incentive to develop a unified framework that allows us to simultaneously capture both margins of globalization in a consistent way.

This incentive is reinforced by recent evidence from the gravity-based empirical literature, indicating that distance as a trade-inhibiting force has gained in importance through recent history. This runs counter to the wide-spread consensus that technological improvements in transport and communication should have reduced the barrier-effect of distance. Using an extensive data set covering trade between 1950 and 1997, we find that the distance-puzzle is very robust with respect to methods and specifications for the estimated gravity equation, as long as one adheres to the traditional approach where only country pairs with strictly positive trade are included. In doing so, however, one looks only at the intensive margin of world

trade, thus systematically ignoring all evidence from the extensive margin which relates to the establishment of new trading relationships.

Going one step further, we take up the obvious question of whether including the extensive margin makes a difference. Arguing that the extensive margin may be treated on an equal footing with the intensive margin of world trade within a generalized gravity model of bilateral trade, we develop a simple model which incorporates a rudimentary treatment of the transport sector. Including the extensive margin of world trade proves difficult, however, due to econometric problems associated with censored regression. We provide a comprehensive discussion of these issues, framed in the so-called corner-solutions model of world trade, whereby zero trade is an endogenous outcome of the trade-generating and trade-inhibiting forces at work. Writing down a statistical model for the corner-solutions gravity equation, the extensive and intensive margins of world trade appear as well defined marginal effects on the conditional mean of bilateral trade. Moreover, the model clearly points to the possibility of biased estimates in conventional, "intensive-margin-only" models. In other words, the model questions the use of traditional estimation techniques for a gravity equation explaining bilateral trade.

As a final step, we therefore provide estimates obtained relying on the Tobit approach towards consistent estimation of such a corner-solutions version of the gravity model, again using our extensive data set covering world trade between 1950 and 1997. Comparing our results with those of the existing literature, we find that the distance-puzzle does indeed disappear. We also find that GATT- or WTO-membership comes out more supportive of trade, than was recently concluded from gravity-type empirical studies.

There are various ways in which the results found here may be improved upon. First, distance-related trade costs are likely to be non-linear in distance. More specifically, it seems questionable that the oft-quoted technological improvements in transport and communication, an alleged driving force of globalization, 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. This points to the need for a less stepmotherly treatment of transport and trade costs

32

in structural models of trade, as recently proposed by Deardorff (2004).

Meanwhile, the results reported here 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. 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 trading relationships become utilized, still leaves much ground to be covered, and much gains from trade to be reaped.

Table 1
Summary Statistics

Variable	1950		1997		1950-1997	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Ln of bilateral trade (in real US$)	5.77	4.20	3.03	3.51	3.36	3.80
Ln of distance (in km)	8.43	0.93	8.45	0.82	8.47	0.82
#0,1,2: sum landlocked	0.19	0.41	0.41	0.57	0.33	0.53
Ln product of land surface (in sqkm)	25.50	2.19	23.29	3.40	23.94	3.20
Ln product of arable surface (in sqkm)	4.80	1.31	4.26	1.81	4.20	1.79
Ln product of real GDPs (in PPP)	47.61	2.00	47.95	2.71	47.58	2.60
Ln structural difference	14.47	2.79	16.72	2.59	15.99	2.76
Ln multilat. resist. index, country i	0.60	0.15	0.34	0.27	0.40	0.26
Ln multilat. resist. index, country j	2.00	1.17	1.80	1.42	2.05	1.25
Both countries in same RTA	0.00	0.00	0.49	0.50	0.40	0.49
#0,1,2: sum Sachs-Warner index	0.24	0.45	1.24	0.69	0.64	0.69
#0,1,2: sum GATT/WTO membership	0.88	0.69	1.60	0.57	1.26	0.69
Number of observations	1239		11993		6599	
Number of strictly positive observations	857		6513		3516	

<div align="center">

Table 2
Regression results: The distance puzzle in linear specifications
Dependent variable: ln of real bilateral trade

</div>

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) RE	(6) RE
Ln of distance	-0.7951 (0.0371)	-0.7958 (0.0372)	-0.7813 (0.0383)	-0.7786 (0.0385)	-1.0349 (0.0234)	-1.0279 (0.0238)
Distance time interaction	-0.0115 (0.0010)	-0.0115 (0.0010)	-0.0139 (0.0010)	-0.0139 (0.0010)	-0.0071 (0.0003)	-0.0071 (0.0003)
Time	0.0191 (0.0079)	0.0103 (0.0098)	0.0119 (0.0079)	0.0103 (0.0099)	-0.0091 (0.0025)	-0.0646 (0.0033)
Ln of product of real GDPs	0.8176 (0.0075)	0.8178 (0.0076)	0.9574 (0.0102)	0.9555 (0.0101)	0.7724 (0.0055)	0.7778 (0.0056)
Ln mult. resist. (ct. i)	1.3839 (0.0755)	1.3683 (0.0754)	0.5037 (0.0744)	0.4786 (0.0741)	1.0377 (0.0641)	0.8991 (0.0650)
Ln mult. resist. (ct. j)	-0.1604 (0.0168)	-0.1634 (0.0168)	-0.0691 (0.0156)	-0.0720 (0.0156)	-0.0854 (0.0126)	-0.0971 (0.0128)
Ln of real price of crude oil		-0.0534 (0.0145)		0.0259 (0.0151)		-0.0004 (0.0070)
Ln of real world gdp		0.0843 (0.1292)		-0.0613 (0.1256)		1.0448 (0.0498)
#0,1,2: sum island			0.1780 (0.0416)	0.1751 (0.0415)	0.0994 (0.0362)	0.0754 (0.0367)
#0,1,2: sum landlocked			-0.1742 (0.0328)	-0.1716 (0.0328)	-0.3699 (0.0307)	-0.3478 (0.0312)
Ln product of land surface (in sqkm.)			-0.1656 (0.0094)	-0.1640 (0.0093)	-0.1011 (0.0069)	-0.1057 (0.0070)
Ln product of arable land surface (in sqkm.)			-0.2277 (0.0126)	-0.2272 (0.0126)	-0.0794 (0.0101)	-0.0850 (0.0102)
Ln structural difference			-0.0098 (0.0022)	-0.0090 (0.0022)	-0.0013 (0.0010)	-0.0006 (0.0010)
Both countries member of same RTA			0.1279 (0.0344)	0.1306 (0.0334)	0.0161 (0.0120)	0.0531 (0.0115)
#0,1,2: Sachs-Warner open countries			0.5054 (0.0239)	0.5228 (0.0232)	0.1831 (0.0074)	0.2308 (0.0072)
#0,1,2: sum GATT/WTO members			0.0340 (0.0252)	0.0352 (0.0251)	0.0429 (0.0083)	0.0547 (0.0083)
Year specific effects ($\mathbf{K}_t$)	yes	no	yes	no	yes	no
Observations	168747	168747	168747	168747	168747	168747
R-squared	0.65	0.64	0.69	0.69	0.68	0.69

Notes: Robust standard errors (clustering by country-pairs) in parentheses. Year-specific intercepts and constants are not reported. RE means random effects estimator. In columns (5) and (6) the R-squared refers to 'between' R-squared.

Table 3
Regression results: The corner solutions model
Dependent variable: ln of real bilateral trade

	(1) OLS $E(T\|\mathbf{X}, \mathbf{T} > \mathbf{0})$	(2) OLS $E(T\|\mathbf{X})$	(3) TOBIT $E(T^*\|\mathbf{X})$	(4) TOBIT $E(T\|\mathbf{X}, \mathbf{T} > \mathbf{0})$	(5) TOBIT $E(T\|\mathbf{X})$	(6) TOBIT $P(T > 0\|\mathbf{X})$
Ln of distance	-0.7814 (0.0383)	-2.3455 (0.0521)	-3.0621 (0.0810)	-1.4428 (0.0382)	-2.0563 (0.0544)	-0.3223 (0.0085)
Distance time interaction	-0.0139 (0.0010)	0.0205 (0.0013)	0.0078 (0.0019)	0.0037 (0.0009)	0.0052 (0.0013)	0.0008 (0.0002)
Time	0.0119 (0.0079)	-0.2463 (0.0108)	-0.1688 (0.0161)	-0.0795 (0.0076)	-0.1134 (0.0108)	-0.0178 (0.0017)
#0,1,2: sum island	0.1781 (0.0416)	0.5484 (0.0468)	0.8143 (0.0753)	0.3837 (0.0355)	0.5468 (0.0506)	0.0857 (0.0079)
#0,1,2: sum landlocked	-0.1743 (0.0328)	-0.6490 (0.0351)	-1.1999 (0.0655)	-0.5654 (0.0309)	-0.8058 (0.0440)	-0.1263 (0.0069)
Ln product of land surface (in sqkm.)	-0.1656 (0.0094)	0.2092 (0.0086)	0.4819 (0.0161)	0.2271 (0.0076)	0.3236 (0.0108)	0.0507 (0.0017)
Ln product of arable land surface (in sqkm.)	-0.2277 (0.0126)	0.1899 (0.0119)	0.3772 (0.0230)	0.1777 (0.0108)	0.2533 (0.0154)	0.0397 (0.0024)
Ln of product of real GDPs	0.9575 (0.0102)	0.4304 (0.0084)	0.4938 (0.0157)	0.2327 (0.0074)	0.3316 (0.0105)	0.0520 (0.0017)
Ln structural difference	-0.0098 (0.0022)	-0.0596 (0.0025)	-0.0532 (0.0041)	-0.0251 (0.0019)	-0.0357 (0.0028)	-0.0056 (0.0004)
Ln mult. resist. (ct. i)	0.5034 (0.0744)	2.5168 (0.0853)	4.2432 (0.1471)	1.9994 (0.0693)	2.8495 (0.0998)	0.4466 (0.0155)
Ln mult. resist. (ct. j)	-0.0690 (0.0156)	-0.0236 (0.0145)	-0.0956 (0.0271)	-0.0451 (0.0128)	-0.0642 (0.0182)	-0.0101 (0.0029)
Both countries member of same RTA	0.1279 (0.0344)	0.9432 (0.0409)	1.4809 (0.0645)	0.7167 (0.0312)	1.0165 (0.0443)	0.1517 (0.0066)
#0,1,2: Sachs-Warner open countries	0.5052 (0.0239)	1.2536 (0.0283)	1.9673 (0.0466)	0.9270 (0.0220)	1.3211 (0.0313)	0.2071 (0.0049)
#0,1,2: sum GATT/WTO members	0.0341 (0.0252)	0.2667 (0.0270)	0.6872 (0.0465)	0.3238 (0.0219)	0.4615 (0.0312)	0.0723 (0.0049)
Both countries opposed in an interstate conflict	-0.7324 (0.7380)	0.6779 (0.6054)	-0.3602 (0.8109)	-0.1648 (0.3710)	-0.2350 (0.5290)	-0.0387 (0.0871)
Observations	168747	316744	316744	316744	316744	316744
R-squared	0.6900	0.6200				

Notes: Robust standard errors (clustering by country-pairs) in parentheses. Year specific intercepts and constants are not reported. Marginal effects are evaluated at sample means.

References

[1] Anderson, James E. 1979. A Theoretical Foundation for the Gravity Equation. *American Economic Review* **69(1)**, 106-16.

[2] Anderson, James E. and Eric van Wincoop. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review* **93(1)**, 170-192.

[3] Anderson, James E. and Eric van Wincoop. 2004. *Trade Costs.* NBER Working Paper 10480.

[4] Bernard, Andrew B., J. Bradford Jensen and Peter K. Schott. 2003. *Falling Trade Costs, Heterogeneous Firms, and Industry Dynamics.* NBER Working Paper 9639.

[5] Buch, Claudia, Jörn Kleinert and Farid Toubal. 2004. The Distance Puzzle: On the Interpretation of the Distance Coefficient in Gravity Equations. *Economics Letters* **83 (3),** 293-298.

[6] Brun, Jean-Francois, Céline Carrère, Patrick Guillaumont and Jaime de Melo. 2004. Has Distance Died? Evidence from a Panel Gravity Model. *World Bank Economic Review*, forthcoming.

[7] CIA. 2003. *The World Factbook 2003*, http://www.cia.gov/cia/publications/factbook/.

[8] Coe, David T., Arvind Subramanian and Natalia T. Tamirisa (with Rikhil Bhavnani). 2002. *The Missing Globalization Puzzle.* IMF Working Paper 02/171.

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

[10] Deardorff Alan V. 2004. Local Comparative Advantage: Trade Costs and the Pattern of Trade. University of Michigan, Ann Arbor: *Research Seminar in International Economics Discussion Paper* No. 500.

[11] Egger, Peter and Michael Pfaffermayr. 2004. *Distance, Trade and FDI: A Hausman-Taylor SUR Approach.* Journal of Applied Econometrics 19, 227–246.

[12] Eichengreen, Berry and Douglas A. Irwin. 1998. The Role of History in Bilateral Trade Flows. In: Frankel, Jeffrey A. (ed.), *The Regionalization of the World Economy.* Chicago and London: University of Chicago Press.

[13] Evenett, Simon J. and Wolfgang Keller. 2002. On Theories Explaining the Success of the Gravity Equation. *Journal of Political Economy* **110(2)**, 281-316.

[14] Evenett, Simon J. and Anthony J. Venables 2002. *Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows,* mimeo.

[15] Feenstra, Robert. 2004. *Advanced International Trade – Theory and Evidence*, Princeton and Oxford: Princeton University Press.

[16] Feenstra, Robert, James A. Markusen and Andrew K. Rose. 2001. Using the Gravity Equation to Differentiate Among Alternative Theories of Trade. *Canadian Journal of Economics* **34(2)**, 430-447.

[17] Frankel, Jeffrey A. 1997. *Regional Trading Blocs in the World Economic System.* Washington, D.C.: Institute for International Economics.

[18] Gleditsch, Kristian Skrede. 2002. Expanded Trade and GDP Data. *Journal of Conflict Resolution* **46(5)**, 712-724.

[19] Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. Armed Conflict 1946–2001: A New Data Set. *Journal of Peace Research* **39(5),** 615–637.

[20] Greene, William. 2003. *Econometric Analysis*, 5th ed., New York: Prentice Hall.

[21] Harrigan, James. 2004. *Specialization and the Volume of Production: Do the Data Obey the Laws?* In Choi, Kwan and James Harrigan (eds.), Handbook of International Trade, Basil Blackwell, ch. 4

[22] Haveman, Jon and David L. Hummels. 2004. Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization. *Canadian Journal of Economics* **37(1),** 199-218.

[23] Helpman, Elhanan and Paul R. Krugman. 1985. *Market Structure and Foreign Trade.* Cambridge, Mass.: MIT Press.

[24] Heston, Alan, Robert Summers and Bettina Aten. 2002. *Penn World Table Version 6.1.* Center for International Comparisons at the University of Pennsylvania (CICUP).

[25] Hummels, David. 1999. *Have International Transportation Costs Declined?* University of Chicago.

[26] Irwin, Douglas A. and Mark Terviö. 2002 Does trade raise income?: Evidence from the twentieth century. *Journal of International Economics* 58, 1-18.

[27] Leamer, Edward E. and James Levinsohn. 1995. International Trade Theory: The Evidence. In: Grossman, Gene and Kenneth Rogoff (eds.). *Handbook of International Economics* **Vol. III**, 1339-1394.

[28] Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade.* Cambridge, Mass.: MIT Press.

[29] Rose, Andrew. 2004. Do We Really Know That the WTO Increases Trade? *American Economic Review* **94(1)**, 98-114.

[30] Rozanski, Jerzy and Alexander Yeats. 1994. On the (In)accuracy of Economic Observations: An Assessment of Trends in the Reliability of International Trade Statistics. *Journal of Development Economics* **44(1)**, 103-130.

[31] Silva, Santos J.M.C. and Silvana Tenreyro. 2003. *Gravity-Defying Trade*. mimeo, Federal Reserve Bankd of Boston, Boston.

[32] Trefler, Daniel. 1995. The Case of the Missing Trade and Other Mysteries. *American Economic Review* 85(5), 1029-1046.

[33] Wacziarg, Romain T. and Karen Horn Welch. 2003. *Trade Liberalization and Growth: New Evidence*. Stanford University.

[34] Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA.: MIT Press.

CESifo Working Paper Series

(for full list see www.cesifo.de)

_____