Trading Partners and Trading Volumes*

by

Elhanan Helpman
Harvard University and CIAR

Marc Melitz
Harvard University,NBER, and CEPR

and

Yona Rubinstein
Tel Aviv University

PRELIMINARY AND INCOMPLETE

March 31, 2005

*We thank Zvi Eckstein and Manuel Trajtenberg for comments. Helpman thanks the NSF for financial support.

1 Introduction

Estimation of international trade flows has a long tradition. Tinbergen (1962) pioneered the use of gravity equations in empirical specifications of bilateral trade flows, in which the volume of trade between two countries is proportional to the product of an index of their economic size, and the factor of proportionality depends on measures of "trade resistance" between them. Among the measures of trade resistance, he included geographic distance, a dummy for common borders, and dummies for Commonwealth and Benelux memberships. Tinbergen's specification has been widely used, simply because it provides a good fit to most data sets of regional and international trade flows. And over time, his approach has been furnished with theoretical underpinnings and better estimation techniques.[1]

While the accurate estimation of international trade flows is important for an understanding of the structure of world trade, the accuracy of such estimates and their interpretation have gained added significance as a result of their wide use in various branches of the empirical literature. These studies rely on measures of trade openness as instruments in the estimation of the impact of economic and political variables on economic success. Much of this work builds on Frankel and Romer (1999), who studied the impact of trade openness on income per capita in a large sample of countries. Their methodology consists of estimating a first-stage gravity equation of bilateral trade flows, which includes indexes of geographic characteristics (size of area, whether a country is landlocked, and whether the two countries have a common border) and bilateral distances. The predicted trade volume from this equation is then used as a measure of trade openness in a second-stage equation that estimates the impact of trade openness on income per capita. They found a large and significant effect.[2]

Hall and Jones (1999) used instrumental variables to estimate the impact of social infrastructure on income per capita. They combined an index of government anti-diversion policies and the fraction of years in which a country was open according to the Sachs and Warner (1995) index to measure social infrastructure.[3] Among the instruments they included the Frankel and Romer (1999) measure of trade openness. Evidently, the accuracy of the estimates from the Frankel–Romer first-stage equation affects the accuracy of the estimates in the second-stage equation, including the marginal impact of social infrastructure on income per capita.

Persson and Tabellini (2003) also used instrumental variables, but they used this method to estimate the impact of political institutions on productivity and growth. They found that in well-established democracies economic policies are more growth-oriented in presidential

[1] See, for example, Anderson (1979) , Helpman and Krugman (1985), Helpman (1987), Feenstra (2002), and Anderson and van Wincoop (2003).

[2] In the working paper that preceded the published version of their paper, Frankel and Romer (1996) used the same methodology to study the impact of openness on the rate of growth of income per capita. They found a strong positive effect.

[3] The index of government anti-diversion policies aggregates measures of law and order, bureaucratic quality, corruption, risk of expropriation, and government repudiation of contracts.

than in parliamentary systems, while in weak democracies economic policies are more growth-oriented in parliamentary systems. Similarly to Hall and Jones (1999), they used the Frankel–Romer instrument of trade openness to reach this conclusion. Therefore, in this case too, the quality of the first-stage gravity equation affects the quality of the second-stage estimates of the impact of political institutions on economic performance.

These examples illustrate the prominent role of the gravity equation in areas other than international trade. In the area of international trade this equation has dominated empirical research. It has been used to estimate the impact on trade flows of international borders, preferential trading blocs, currency unions, membership in the WTO, as well as the size of home-market effects.[4]

All the above mentioned studies estimate the gravity equation on samples of countries that have only positive trade flows between them. We argue in this paper that, by disregarding countries that do not trade with each other, these studies give up important information contained in the data, and they produce biased estimates as a result. We also argue that standard specifications of the gravity equation impose symmetry that is inconsistent with the data, and that this too biases the estimates. To correct these biases, we develop a theory that predicts positive as well as zero trade flows between countries, and use the theory to derive estimation procedures that exploit the information contained in data sets of trading and non-trading countries alike.[5]

The next section briefly reviews the evolution of the volume of trade among the 161 countries in our sample, and the composition of country pairs according to their trading status.[6] Three features stand out. First, about half of the country pairs do not trade with one-another.[7] Second, the rapid growth of world trade from 1970 to 1997 was predominantly due to the growth of the volume of trade among countries that traded with each other in 1970 rather than due to the expansion of trade among new trade partners. Third, the average volume of trade at the end of the period between pairs of countries that exported to one-another in 1970 was much larger than the average volume of trade at the end of the period of country pairs with a different trade status. Nevertheless, we show in Section 6 that the volume of trade between pairs of countries that traded with one-another was significantly influenced by the fraction of firms that engaged in foreign trade, and that this fraction varied

[4] See McCallum (1995) for the study that triggered an extensive debate on the role of international borders, as well as Wei (1996), Evans (2003), and Anderson and van Wincoop (2003). Feenstra (2003, chap. 5) provides an overview of this debate. Also see Frankel (1997) on preferential trading blocs, Rose (2000) and Tenreyro and Barro (2002) on currency unions, Rose (2004) on WTO membership, and Davis and Weinstein (2003) on the size of home-market effects.

[5] Anderson and van Wincoop (2004), Evenett and Venables (2002), and Haveman and Hummels (2004) all highlight the prevalence of zero bilateral trade flows and suggest theoretical interpretations for them. We provide a theoretical framework that jointly determines both the set of trading partners and their trade volumes, and we develop estimation procedures for this model.

[6] See appendix A for data sources and for the list of the 161 countries.

[7] We say that a country pair i and j does not trade with one-another if i does not export to j and j does not export to i.

systematically with country characteristics. Therefore the intensive margin of trade was substantially driven by variations in the fraction of trading firms, but not by new trading partners.

We develop in Section 3 the theoretical model that motivates our estimation procedures. This is a model of international trade in differentiated products in which firms face fixed and variable costs of exporting, along the lines suggested by Melitz (2003). Firms vary by productivity, and only the more productive firms find it profitable to export. Moreover, the profitability of exports varies by destination; it is higher to countries with higher demand levels, lower variable export costs, and lower fixed export costs. As a result, to every destination country i, there is a marginal exporter in country j that just breaks even by exporting to i. Country j firms with higher productivity than the marginal exporter have positive profits from exporting to i.

This model has a number of implications for trade flows. First, it allows all firms in a country j to choose not to export to a country i, because it is possible for no firm in j to have productivity above the threshold that makes exports to i profitable. The model is therefore able to predict zero exports from j to i for some country pairs. As a result, the model is consistent with zero trade flows in both directions between some countries, as well as zero exports from j to i but positive exports from i to j for some country pairs. Both types of trade patterns exist in the data. Second, the model predicts positive trade flows in both directions for some country pairs, which is also needed in order to explain the data. And finally, the model generates a gravity equation.

Our derivation of the gravity equation generalizes the Anderson and van Wincoop (2003) equation in two ways. First, it accounts for firm heterogeneity and fixed trade costs. Second, it accounts for asymmetries between the volume of exports from j to i and the volume of exports from i to j. Both are important for data analysis. We also develop a set of sufficient conditions under which more general forms of the Anderson-van Wincoop equations aggregate trade flows across heterogeneous firms facing both fixed and variable trade costs.

Section 4 develops the empirical framework for estimating the gravity equation derived in Section 3. We propose a two stage estimation procedure. The first stage consists of estimating a Probit equation that specifies the probability that country j exports to i as a function of observable variables. The specification of this equation is derived from the theoretical model and an explicit introduction of unobservable variations. Predicted components of this equation are then used in the second stage to estimate the gravity equation in log-linear form. We show that this procedure yields consistent estimates of the parameters of the gravity equation, such as the marginal impact of distance between countries on their exports to one-another.[8] It simultaneously corrects for two types of potential biases: a Heckman selection bias and a bias from potential asymmetries in the trade flows between pairs of

[8] We also show that consistency requires the use of separate country fixed effects for exporters and importers, as proposed by Feenstra (2002).

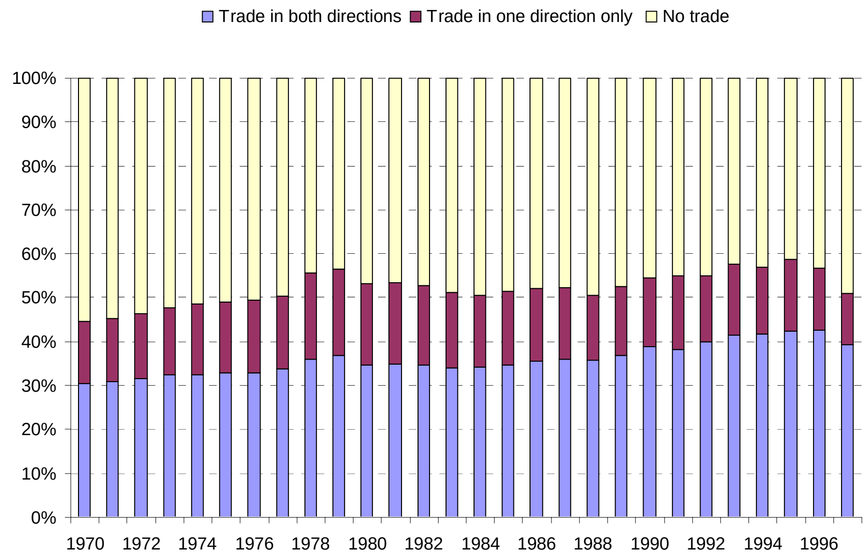

Figure 1: Distribution of country pairs among pairs trading in both directions, pairs trading in one direction only, and nontrading pairs: 12,880 pairs constructed form 161 countries, 1970-1997

countries. Since this procedure is easy to implement, it can be effectively used in many application, such as instrumental variables estimation of the impact of political variables on economic outcomes.

It is interesting to note that despite the fact that our theoretical model has firm heterogeneity, we do not need firm-level data to estimate the gravity equation. This stems from the fact that the features of marginal exporters can be identified from the variation in the characteristics of the destination countries. That is, for every country j, its exports to different countries vary by the characteristics of the importers. As a result, there exist sufficient statistics, which can be computed from aggregate data, that predict the volume of exports of heterogeneous firms.[9]

Section 5 shows that variables that are commonly used in gravity equations also affect the probability that two countries trade with each other. This provides evidence for a potential bias in the standard estimates. The extent of this bias is then studied in Section 6.

[9] Eaton and Kortum (2002) apply a similar principle to determine an aggregate gravity equation across heterogeneous Ricardian sectors. As in our model, the predicted trade volume reflects an extensive margin (number of sectors/goods traded) and an intensive one (volume of trade per good/sector). However, Eaton and Kortum do not model fixed trade costs and the possibility of zero bilateral trade flows. Unlike our equations, theirs are subject to the criticism raised by Haveman and Hummels (2004). Bernard, Eaton, Jensen, and Kortum (2003) use direct information on U.S. plant-level sales, productivity, and export status to calibrate a model which is then used to simulate the extensive and intensive margins of bilateral trade flows.

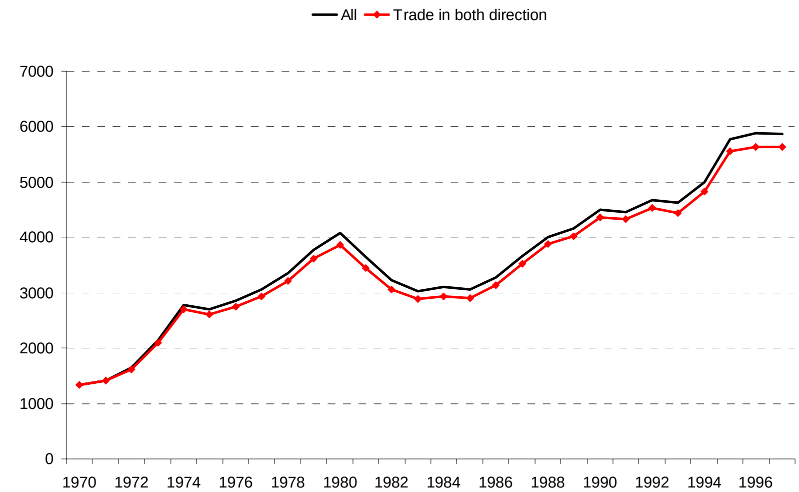

Figure 2: Aggregate volumes of exports, measured in billions of 2000 U.S. dollars, of all country pairs and of country pairs that traded in both directions in 1970, 1970-1997

2 A Glance at the Data

Figure 1 depicts the empirical extent of zero trade flows. In this figure, all possible country pairs are partitioned into three categories: the top portion represents the fraction of country pairs that do not trade with one-another; the bottom portion represents those that trade in both directions (they export to one-another); and the middle portion represents those that trade in one direction only (one country imports from, but does not export to, the other country). As is evident from the figure, by disregarding countries that do not trade with each other or trade only in one direction one disregards close to half of the observations. We show below that these observations contain useful information for estimating international trade flows.[10]

Figure 2 shows the evolution of the aggregate real volume of exports of all 161 countries in our sample, and of the aggregate real volume of exports of the subset of country pairs that exported to one-another in 1970. The difference between the two curves represents the volume of trade of country pairs that either did not trade in 1970 or traded in 1970 in one direction only. It is clear from this figure that the rapid growth of trade, at an annual rate of 7.5% on average, was mostly driven by the growth of trade between countries that traded with each other in both directions at the beginning of the period. In other words, the

[10] Silva and Tenreyro (2003) also argue that zero trade flows can be used in the estimation of the gravity equation, but they emphasize a heteroskedasticity bias that emanates from the log-linearization of the equation rather than the selection and asymmetry biases that we emphasize. Moreover, the Poisson method that they propose to use yields similar estimates on the sample of countries that have positive trade flows in both directions and the sample of countries that have positive and zero trade flows. We shall have more to say about their paper in Section 5.

contribution to the growth of trade of countries that started to trade after 1970 in either one or both directions, was relatively small.

Combining this evidence with the evidence from Figure 1, which shows a relatively slow growth of the fraction of trading country pairs, suggests that bilateral trading volumes of country pairs that traded with one-another in both directions at the beginning of the period must have been much larger than the bilateral trading volumes of country pairs that either did not trade with each other or traded in one direction only at the beginning of the period. Indeed, at the end of the period the average bilateral trade volume of country pairs of the former type was about 35 times larger than the average bilateral trade volume of country pairs of the latter type. This suggests that the rapid growth of world trade was an intensive margin phenomenon. That is, the enlargement of the set of trading countries did not contribute in a major way to the growth of world trade.[11]

3 Theory

Consider a world with J countries, indexed by $j = 1, 2, ..., J$. Every country consumes and produces a continuum of products. Country j's utility function is

$$u_j = \left[\int_{l \in B_j} x_j (l)^{\alpha} dl \right]^{\alpha}, \ \ 0 < \alpha < 1 \ ,$$

where $x_j(l)$ is its consumption of product l and B_j is the set of products available for consumption in country j. The parameter α determines the elasticity of substitution across products, which is $\varepsilon = 1/(1-\alpha)$. This elasticity is the same in every country.

Let Y_j be the income of country j, which equals its expenditure level. Then country j's demand for product l is

$$x_j(l) = \frac{\hat{p}_j(l)^{-\varepsilon} Y_j}{P_j^{1-\varepsilon}} \ , \tag{1}$$

where $\hat{p}_j(l)$ is the price of product l in country j and P_j is the country's ideal price index, given by

$$P_j = \left[\int_{l \in B_j} \hat{p}_j (l)^{1-\varepsilon} dl \right]^{1/(1-\varepsilon)} . \tag{2}$$

This specification implies that every product has a constant demand elasticity ε.

Some of the products consumed in country j are domestically produced while others are imported. Country j has a measure N_j of firms, each one producing a distinct product. The products produced by country-j firms are also distinct from the products produced by

[11] This contrasts with the sector-level evidence presented by Evenett and Venables (2002). They find a substantial increase in the number of trading partners at the 3-digit sector level for a selected group of 23 developing countries. We conjecture that their country sample is not representative and that most of their new trading pairs were originally trading in other sectors.

country-i firms for $i \neq j$. As a result, there are $\sum_{j=1}^{J} N_j$ products in the world economy.

A country-j firm produces one unit of output with a cost-minimizing combination of inputs that cost $c_j a$, where a measures the number of bundles of the country's inputs used by the firm per unit output and c_j measures the cost of this bundle. The cost c_j is country specific, reflecting differences across countries in factor prices, whereas a is firm-specific, reflecting productivity differences across firms in the same country. The inverse of a, $1/a$, represents the firm's productivity level.[12] We assume that a cumulative distribution function $G(a)$ with support $[a_L, a_H]$ describes the distribution of a across firms, where $a_H > a_L \geq 0$. This distribution function is the same in all countries.[13]

We assume that a producer bears only production costs when selling in the home market. That is, if a country-j producer with coefficient a sells in country j, the delivery cost of its product is $c_j a$. If, however, this same producer seeks to sell its product in country i, there are two additional costs it has to bear: a fixed cost of serving country i, which equals $c_j f_{ij}$, and a transport cost. As is customary, we adopt the 'melting iceberg' specification and assume that τ_{ij} units of a product have to be shipped from country j to i in order for one unit to arrive. We assume that $f_{jj} = 0$ for every j and $f_{ij} > 0$ for $i \neq j$, and $\tau_{jj} = 1$ for every j and $\tau_{ij} > 1$ for $i \neq j$. Note that the fixed cost coefficients f_{ij} and the transport cost coefficients τ_{ij} depend on the identity of the importing and exporting countries, but not on the identity of the exporting producer. In particular, they do not depend on the producer's productivity level.

There is monopolistic competition in final products. Since every producer of a distinct product is of measure zero, the demand function (1) implies that a country-j producer with an input coefficient a maximizes profits by charging the mill price

$$p_j(a) = \frac{1}{\alpha} c_j a \ . \tag{3}$$

This is a standard markup pricing equation, with the markup being smaller the larger the demand elasticity of demand. It follows that if the country-j producer of product l has the input coefficient a and it sells its product in the home market, the home market consumer pays $\hat{p}_j(l) = c_j a/\alpha$. If, however, it sells the product in a foreign country i, the consumers in i are charged $\hat{p}_i(l) = \tau_{ij} c_j a/\alpha$. As a result, the producer's operating profits from selling in country i are

$$\pi_{ij}(a) = (1-\alpha)\left(\frac{\tau_{ij} c_j a}{\alpha P_i}\right)^{1-\varepsilon} Y_i - c_j f_{ij} \ .$$

Evidently, these operating profits are positive for sales in the domestic market, because $f_{jj} = 0$. Therefore all N_j producers sell in country j. But sales in country $i \neq j$ are

[12] See Melitz (2003) for a discussion of a general equilibrium model of trading countries in which firms are heterogeneous in productivity. We follow his specification.

[13] The as only capture relative productivity differences across firms in a country. Aggregate productivity differences across countries are subsumed in the c_js.

profitable only if $a \leq a_{ij}$, where a_{ij} is defined by $\pi_{ij}(a_{ij}) = 0$, or [14]

$$(1-\alpha)\left(\frac{\tau_{ij}c_j a_{ij}}{\alpha P_i}\right)^{1-\varepsilon} Y_i = c_j f_{ij} \ . \tag{4}$$

It follows that only a fraction $G(a_{ij})$ of country j's N_j firms export to country i. For this reason the set B_i of products that are available in country i is smaller than the set of products available in the world economy. In particular, no firm from country j exports to country i if a_{ij} is smaller than a_L, i.e., if the least productive firm that can profitably export to country i has a coefficient a that is below the support of $G(a)$. And all firms from country j export to country i if a_{ij} is larger than a_H.

We next characterize bilateral trade volumes. Let

$$V_{ij} = \begin{cases} \int_{a_L}^{a_{ij}} a^{1-\varepsilon} dG(a) & \text{for} \quad a_{ij} \geq a_L \\ 0 & \text{otherwise} \end{cases} \ . \tag{5}$$

Then the demand function (1) and the pricing equation (3) imply that the value of country i's imports from j is

$$M_{ij} = \left(\frac{c_j \tau_{ij}}{\alpha P_i}\right)^{1-\varepsilon} Y_i N_j V_{ij} \ . \tag{6}$$

This bilateral trade volume equals zero when $a_{ij} \leq a_L$, because under these circumstances $V_{ij} = 0$. Using the definition of V_{ij} and (2), we also obtain

$$P_i^{1-\varepsilon} = \sum_{j=1}^{J} \left(\frac{c_j \tau_{ij}}{\alpha}\right)^{1-\varepsilon} N_j V_{ij}. \tag{7}$$

Equations (4)-(7) provide a mapping from the income levels Y_i, the numbers of firms N_i, the unit costs c_i, the fixed costs f_{ij}, and the transport costs τ_{ij}, to the bilateral trade flows M_{ij}.

We show in Appendix B that, together with equality of income and expenditure, equations (4)-(7) can be used to derive a generalization of Anderson and van Wincoop's (2003) gravity equation that embodies third-country effects. Their equation applies when transport costs are symmetric, i.e., $\tau_{ij} = \tau_{ji}$ for all country pairs, and the variables V_{ij} can be multiplicatively decomposed into three components: one that depends only on importer characteristics, a second that depends only on exporter characteristics, and a third that depends on the country pair characteristics but is symmetric across country pairs, so that it is the same for i, j as for j, i. This decomposability holds in Anderson and van Wincoop's model. Importantly, however, there are other cases of interest, less restrictive than the Anderson and van Wincoop specification, that satisfy them too. Therefore, our equation applies under

[14] Note that $a_{ij} \to +\infty$ as $f_{ij} \to 0$.

wider circumstances, and in particular, when there is productivity heterogeneity across firms and firms bear fixed costs of exporting. Under these circumstances only a fraction of the firms export; those with the highest productivity. Finally, note that our formulation is more relevant for empirical analysis, because, unlike previous formulations, it enables bilateral trade flows to equal zero. This flexibility is important because, as we have explained in the introduction, there are many zero bilateral trade flows in the data.

In order to gain as much flexibility as possible in the empirical application, we develop in the next section an estimation procedure that builds directly on equations (4)-(7), which allow for asymmetric bilateral trade flows, including zeros.

4 Empirical Framework

We maintain the assumption of a Pareto distribution for productivity, $1/a$, but now assume that this distribution is truncated at an upper bound $1/a_L$. Thus, $G(a) = a^k / \left(a_H^k - a_L^k\right)$, and $a_H > a_L > 0$. In addition, we allow $a_{ij} < a_L$ for some i, j pairs. When this happens, no firm from country j is productive enough to export to country i, inducing zero exports from j to i, i.e., $V_{ij} = 0$ and $M_{ij} = 0$. However, firms from country j may export to other destinations and country i may import from other sources. In other words, this framework allows for asymmetric trade flows, $M_{ij} \neq M_{ji}$, which may also be unidirectional, with $M_{ji} > 0$ and $M_{ij} = 0$, or $M_{ji} = 0$ and $M_{ij} > 0$. Such unidirectional trading relationships are empirically common and can be predicted using our empirical method. Moreover, asymmetric trade frictions are not necessary to induce such asymmetric trade flows when productivity is drawn from a truncated Pareto distribution.

Our assumptions imply that V_{ij} can be expressed as (see (5)):

$$V_{ij} = \frac{k a_L^{k-\varepsilon+1}}{(k-\varepsilon+1)\left(a_H^k - a_L^k\right)} W_{ij},$$

where

$$W_{ij} = \max\left\{\left(\frac{a_{ij}}{a_L}\right)^{k-\varepsilon+1} - 1, 0\right\}, \tag{8}$$

and a_{ij} is determined by the zero profit condition (4). Note that both V_{ij} and W_{ij} are monotonic functions of the proportion of exporters from j to i, $G(a_{ij})$. The export volume from j to i, given by (6), can now be expressed in log-linear form as

$$m_{ij} = (\varepsilon - 1)\ln\alpha - (\varepsilon - 1)\ln c_j + n_j + (\varepsilon - 1)\, p_i + y_i + (1-\varepsilon)\ln\tau_{ij} + v_{ij},$$

where lowercase variables represent the natural logarithms of their respective uppercase variables. τ_{ij} captures variable trade costs; costs that affect the volume of firm-level exports. We assume that these costs are stochastic due to i.i.d. unmeasured trade frictions u_{ij}, which are country-pair specific. In particular, let $\tau_{ij}^{\varepsilon-1} \equiv D_{ij}^{\gamma} e^{-u_{ij}}$, where D_{ij} represents the (symmetric) distance between i and j, and $u_{ij} \sim N(0, \sigma_u^2)$.[15] Then the equation of the bilateral trade flows m_{ij} yields the following estimating equation:

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + w_{ij} + u_{ij}, \tag{9}$$

where $\chi_i = (\varepsilon - 1)\, p_i + y_i$ is a fixed effect of the importing country and $\lambda_j = -(\varepsilon - 1) \ln c_j + n_j$ is a fixed effect of the exporting country.[16]

The estimating equation (9) highlights several important differences with the gravity equation, as derived, for example, by Anderson and van Wincoop (2003). The most important difference is the addition in our formulation of the new variable w_{ij}, that controls for the fraction of firms (possibly zero) that export from j to i. This variable is a function of the cutoff a_{ij}, which is determined by other explanatory variables (see (4)). When w_{ij} is not included on the right-hand-side, the coefficient γ on distance (or any other coefficient on a potential trade barrier) can no longer be interpreted as the elasticity of a firm's trade with respect to distance (or other trade barriers), which is the way in which such trade barriers are almost always modeled in the literature that follows the "new" trade theory. Instead, the estimation of the standard gravity equation confounds the effects of trade barriers on firm-level trade with their effects on the proportion of exporting firms, which induces an upward bias in the estimated coefficient γ.

Another bias is introduced in the estimation of equation (9) when country pairs with zero trade flows are excluded. This selection effect induces a positive correlation between the unobserved u_{ij}s and the trade barrier d_{ij}s; country pairs with large observed trade barriers (high d_{ij}) that trade with each other are likely to have low unobserved trade barriers (high u_{ij}). Although this induces a downward bias in the trade barrier coefficient, our empirical results show that this effect is dominated by the upward bias generated by the endogenous number of exporters.

Lastly, we emphasize again that in our formulation bilateral trade flows need not be balanced, even when all bilateral trade barriers are symmetric. First, the variables w_{ij} can be asymmetric. Second, the fixed effects of importers may differ from the fixed effects of exporters. This substantiates the use of export flows and separate fixed effects as an exporter and as an importer, for every country.

[15] In the following derivations, we use distance as the only source of observable variable trade costs. It should nevertheless be clear how this approach generalizes to a vector of observable bilateral trade frictions paired with a vector of elasticities γ.

[16] We replace v_{ij} with w_{ij}, and therefore β_0 now also contains the log of the constant multiplier in V_{ij}. If tariffs are not directly controlled for, then the importer's fixed effect will subsume an average tariff level. Similarly, average export taxes will show up in the exporter's fixed effect.

Firm Selection Into Export Markets

The selection of firms into export markets, represented by the variable W_{ij}, is determined by the cutoff value of a_{ij}, which is implicitly defined by the zero profit condition (4). We define a related latent variable Z_{ij} as:

$$Z_{ij} = \frac{(1-\alpha)\left(P_i \frac{\alpha}{c_j \tau_{ij}}\right)^{\varepsilon-1} Y_i a_L^{1-\varepsilon}}{c_j f_{ij}}.$$

This is the ratio of variable export profits for the most productive firm (with productivity $1/a_L$) to the fixed export costs (common to all exporters) for exports from j to i. Positive exports are observed if and only if $Z_{ij} > 1$. In this case W_{ij} is a monotonic function of Z_{ij}, i.e., $W_{ij} = Z_{ij}^{(k-\varepsilon+1)/(\varepsilon-1)} - 1$ (see (4) and (8)). As with the variable trade costs τ_{ij}, we assume that the fixed export costs f_{ij} are stochastic due to unmeasured trade frictions ν_{ij} that are i.i.d., but may be correlated with the u_{ij}s. Let $f_{ij} \equiv \exp\left(\phi_{EX,j} + \phi_{IM,i} + \kappa\phi_{ij} - \nu_{ij}\right)$, where $\nu_{ij} \sim N(0, \sigma_\nu^2)$, $\phi_{IM,i}$ is a fixed trade barrier imposed by the importing country on all exporters, $\phi_{EX,j}$ is a measure of fixed export costs common across all export destinations, and ϕ_{ij} is an observed measure of any additional country-pair specific fixed trade costs.[17] Using this specification together with $(\varepsilon-1)\ln\tau_{ij} \equiv \gamma d_{ij} - u_{ij}$, the latent variable $z_{ij} \equiv \ln Z_{ij}$ can be expressed as

$$z_{ij} = \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\phi_{ij} + \eta_{ij}, \tag{10}$$

where $\eta_{ij} \equiv u_{ij} + \nu_{ij} \sim N(0, \sigma_u^2 + \sigma_\nu^2)$ is i.i.d. (yet correlated with the error term u_{ij} in the gravity equation), $\xi_j = -\varepsilon \ln c_j + \phi_{EX,j}$ are fixed effects of exporters, and $\zeta_i = (\varepsilon - 1)\, p_i + y_i - \phi_{IM,i}$ are fixed-effects of importers. Although z_{ij} is unobserved, we observe the presence of trade flows. Therefore $z_{ij} > 0$ when j exports to i and $z_{ij} = 0$ when it does not. Moreover, the value of z_{ij} affects the export volume.

Define the indicator variable T_{ij} to equal 1 when country j exports to i and 0 when it does not. Let ρ_{ij} be the probability that j exports to i, conditional on the observed variables. Since we do not want to impose $\sigma_\eta^2 \equiv \sigma_u^2 + \sigma_\nu^2 = 1$, we divide (10) by the standard deviation σ_η, and specify the following Probit equation:

$$\rho_{ij} = \Pr(T_{i,j} = 1 \mid \text{observed variables}) = \Phi\left(\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma^* d_{ij} - \kappa^*\phi_{ij}\right), \tag{11}$$

where $\Phi(\cdot)$ is the cdf of the unit-normal distribution, and every starred coefficient represents the original coefficient divided by σ_η.[18] Importantly, this selection equation has been derived

[17] As with variable trade costs, it should be clear how this derivation can be extended to a vector of observable fixed trade costs.

[18] By construction, the error term $\eta_{ij}^* \equiv \eta_{ij}/\sigma_\eta$ is distributed unit-normal. The Probit equation (11) distinguishes between observable trade barriers that affect variable trade costs (d_{ij}) and fixed trade costs (f_{ij}). In practice, some variables may affect both. Their coefficients in (11) then capture the combined effect of these barriers.

from a firm-level decision, and it therefore does not contain the unobserved and endogenous variable W_{ij} that is related to the fraction of exporting firms. Moreover, the Probit equation can be used to derive consistent estimates of W_{ij}.

Let $\hat{\rho}_{ij}$ be the predicted probability of exports from j to i, using the estimates from the Probit equation (11), and let $\hat{z}^*_{ij} = \Phi^{-1}\left(\hat{\rho}_{ij}\right)$ be the estimated latent variable $z^*_{ij} \equiv z_{ij}/\sigma_\eta$. Then, a consistent estimate for W_{ij} can be obtained from

$$W_{ij} = \max\left\{\left(Z^*_{ij}\right)^{\delta} - 1, 0\right\}, \tag{12}$$

where $\delta \equiv \sigma_\eta \left(k - \varepsilon + 1\right) / \left(\varepsilon - 1\right)$.

Consistent Estimation of the Log-Linear Equation

Consistent estimation of (9) requires controls for both the endogenous number of exporters (via w_{ij}) and the selection of country pairs into trading partners (which generates a correlation between the unobserved u_{ij} and the independent variables). We thus need estimates for $E\left[w_{ij} \mid ., T_{ij} = 1\right]$ and $E\left[u_{ij} \mid ., T_{ij} = 1\right]$. Both terms depend on $\bar{\eta}^*_{ij} \equiv E\left[\eta^*_{ij} \mid ., T_{ij} = 1\right]$. Moreover, $E\left[u_{ij} \mid ., T_{ij} = 1\right] = \operatorname{corr}\left(u_{ij}, \eta_{ij}\right) \frac{\sigma_u}{\sigma_\eta} \bar{\eta}^*_{ij}$. Since η^*_{ij} has a unit Normal distribution, a consistent estimate $\hat{\bar{\eta}}^*_{ij}$ is obtained from the inverse Mills ratio, i.e., $\hat{\bar{\eta}}^*_{ij} = \phi(\hat{z}^*_{ij})/\Phi(\hat{z}^*_{ij})$. Therefore $\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}$ is a consistent estimate for $E\left[z^*_{ij} \mid ., T_{ij} = 1\right]$ and $\hat{\bar{w}}^*_{ij} \equiv \ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}\right)\right] - 1\right\}$ is a consistent estimate for $E\left[w_{ij} \mid ., T_{ij} = 1\right]$ (see (12)). We therefore can estimate (9) using the transformation

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}\right)\right] - 1\right\} + \beta_{u\eta}\hat{\bar{\eta}}^*_{ij} + e_{ij}, \tag{13}$$

where $\beta_{u\eta} \equiv \operatorname{corr}\left(u_{ij}, \eta_{ij}\right) \frac{\sigma_u}{\sigma_\eta}$ and e_{ij} is an i.i.d. normally distributed error term satisfying $E\left[e_{ij} \mid ., T_{ij} = 1\right] = 0$. Since (13) is non-linear in δ, we estimate it using maximum likelihood (maintaining the normality assumption for e_{ij}).

The use of $\hat{\bar{\eta}}^*_{ij}$ to control for $E\left[u_{ij} \mid ., T_{ij} = 1\right]$ is the standard Heckman (1979) correction for sample selection. This addresses the biases generated by the unobserved country-pair level shocks u_{ij} and η_{ij}, but this does not correct for the biases generated by the underlying unobserved firm-level heterogeneity. The latter biases are corrected by the additional control $\hat{z}^*_{ij}$ (along with the functional form determined by our theoretical assumptions). Used alone, the standard Heckman (1979) correction would only be valid in a world without firm-level heterogeneity, or where such heterogeneity is not correlated with the export decision. Then, all firms are identically affected by trade barriers and country characteristics, and make the same export decisions — or make export decisions that are uncorrelated with trade barriers and country characteristics. This misses the potentially important effect of trade barriers and country characteristics on the share of exporting firms. In a world with firm-level heterogeneity, a larger fraction of firms export to more "attractive" export destinations.

Our empirical results highlight the overwhelming contribution of this channel relative to the standard correction for sample selection, which ignores firm-level heterogeneity.

5 Traditional Estimates

Traditional estimates of the gravity equation use data on country pairs that trade in at least one direction. The first column in Table 1 provides a representative estimate of this sort, for 1986. Note that instead of constructing symmetric trade flows by combining exports and imports for each country pair, we use the unidirectional trade value and introduce both importing and exporting country fixed effect. With these fixed effects every country pair can be represented twice: one time for exports from i to j and another time for exports from j to i. Nevertheless, the results in Table 1 are similar to those obtained with symmetric trade flows and a unique country fixed effect. They show that country j exports more to country i when the two countries are closer to each other, they both belong to the same regional free trade agreement (FTA), they share a common language, they have a common land border, they are not islands, they share the same legal system, they share the same currency, and if one country has colonized the other. The probability that two randomly drawn persons, one from each country, share the same religion does not affect export volumes. Details on the construction of the variables are provided in the appendix.

Among the 158 countries with available data, there are 24,806 possible bilateral export relationships. However, only 11,146 of these relationships have non-zero exports. We then estimate a Probit equation for the presence of a trading relationship using the same explanatory variables as the initial gravity specification (the specification follows (11), with exporter and importer fixed effects). The results are reported in column 2, along with the marginal effects evaluated at the sample means. These results clearly show that the very same variables that impact export volumes from j to i also impact the *probability* that j exports to i. In almost all cases, the impact goes in the same direction. The effect of a common border is the only exception: it raises the volume of trade but reduces the probability of trading. We attribute this finding to the effect of territorial border conflicts that suppress trade between neighbors. In the absence of such conflicts, common land borders enhance trade. We also note that a common religion strongly affects the formation of trading relationships (its effect is almost as large as that for a common language), yet its effect on trade volumes is negligible. Overall, this evidence strongly suggests that disregarding the selection equation of trading partners biases the estimates of the export equation, as we have argued in Section 4.

These results, and their consequences, are not specific to 1986. We repeat the same regressions increasing the sample years to cover all of the 1980s, adding year fixed effects. The results in columns 3 and 4 are very similar to those in the first two columns. As expected, the standard errors are reduced (all standard errors are robust to clustering by country pairs). Adding the time variation also allows the identification of the effects of changing country characteristics. We use this additional source of variation to investigate the effects of WTO/GATT membership (hereafter summarized as WTO) on trade volumes as well as the formation of bilateral trade relationships. We thus repeat the same regressions for the 1980s, adding bilateral controls whenever both countries or neither country is a member of WTO. As emphasized by Subramanian and Wei (2003), the use of unidirectional trade data and separate exporter and importer fixed effects substantially increases the statistically significant positive effect of WTO membership on trade volumes.[19] Our theoretical framework provides the justification for this estimation strategy when bilateral trade flows are asymmetric. Furthermore, we also find that WTO membership has a very strong and significant effect on the formation of bilateral trading relationships. The coefficients in column 6 show that, for any country pair, joint WTO membership has a similar impact on the probability of trade as a common language or colonial ties.

6 Two-Stage Estimation

Now turn to the second-stage estimation of the trade flow equation, as proposed in Section 4. We have already run the first-stage Probit selection equation (11), which yields the predicted probability of export $\hat{\rho}_{ij}$ (see Table 1). We use the estimates of this equation to construct $\hat{\bar{\eta}}^*_{ij} = \phi(\hat{z}^*_{ij})/\Phi(\hat{z}^*_{ij})$ and $\hat{w}^*_{ij}(\delta) = \ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}\right)\right] - 1\right\}$.[20] The former controls for the sample selection bias while the latter controls for unobserved firm heterogeneity, i.e., the effect of trade frictions and country characteristics on the proportion of exporters. Our theoretical model suggests that potential trade barriers that only represent fixed trade costs should only be used as explanatory variables in the selection equation. Econometrically, this provides the needed exclusion restriction for identification of the second stage gravity equation for trade volumes. On both theoretical and empirical grounds (see the results in Table 1), we omit the common religion indicator from the second stage estimation.[21]

The results from the selection equation are reproduced in the initial columns of Table 2 for both 1986 and the 1980s. We also re-run the standard "benchmark" gravity equation omitting the religion control and report the results in the next columns (they are almost identical to those in Table 1). The following columns implement the second stage estimation by incorporating the controls for $\hat{w}^*_{ij}$ and $\hat{\bar{\eta}}^*_{ij}$. Both the non-linear coefficient δ for $\hat{w}^*_{ij}$ and the linear coefficient for $\hat{\bar{\eta}}^*_{ij}$ are precisely estimated. The remaining results for the linear coefficients clearly demonstrate the importance of unmeasured heterogeneity bias when estimating the effect of trade barriers: higher trade volumes are not just the direct consequence of lower trade barriers; they also represent a greater proportion of exporters to a particular

[19] Rose (2004) reports a significant though smaller effect of WTO membership on trade volumes using symmetric trade flow data and a unique set of country fixed effects.

[20] Recall that $\hat{z}^*_{ij} = \Phi^{-1}\left(\hat{\rho}_{ij}\right)$.

[21] Another source of identification comes from the opposite effect of a common border in the selection and trade volume equations.

destination. Consequently, the measures of the effects of trade frictions in the benchmark gravity equation are biased upwards as they confound the true effect of these frictions with their indirect effect on the proportion of exporting firms.[22] As highlighted in Table 2, these biases are substantial. The coefficient on distance drops roughly by a third, indicating a much smaller effect of distance on firm level (hence product level) trade.[23] The effects of a currency union and colonial ties on firm or product level trade are also reduced by a similar proportion. The biases for the effects of FTAs and WTO membership are even more severe as their coefficients drop roughly in half, though they both remain economically and statistically significant. The measured effect of a common language is even more affected as it becomes insignificant (and precisely estimated around zero). This suggests that a common language predominantly reduces the fixed costs of trade: it has a great influence on a firm's choice of export location, but not on its export volume, once that decision is made.

Decomposing the Biases

Our second stage estimation addresses two different sources of bias for standard gravity equations: a selection bias that arises from the pairing of countries into exporter-importer relationships, and an unobserved heterogeneity bias that results from the variation in the fraction of firms that export from a source to a destination country. To examine the relative importance of these biases, we now estimate two specifications of the second-stage export equation, one controlling for unobserved heterogeneity only, the other controlling for selection only.

The results for 1986 are reported in Table 3. The first two columns report the standard gravity "benchmark" equation and our second stage estimation from Table 2. The differences in the estimated coefficients of these two equations represent the joint outcome of the two biases. As we discussed, all the coefficients, with the exception of the land border effect, are lower in absolute value in the second column. We then implement a simple linear correction for unobserved heterogeneity by adding $\hat{z}^*_{ij} = \Phi^{-1}(\hat{\rho}_{ij})$ as an additional regressor to the standard gravity specification (here, we do not correct for the sample selection bias via $\hat{\bar{\eta}}^*_{ij}$). The results reported in the third column clearly show that this unobserved heterogeneity (the proportion of exporting firms) addresses almost all the biases in the standard gravity equation. The coefficients and standard errors for all the observed trade barriers are very similar to those obtained in our second stage non-linear estimation.

In the fourth column, we correct only for the selection bias (the standard two-stage Heckman selection procedure) by introducing the Mills ratio $\hat{\bar{\eta}}^*_{ij}$ as an additional regressor to the benchmark specification. Although the estimated coefficient on $\hat{\bar{\eta}}^*_{ij}$ is positive and significant,

[22]The effect of a land border is an exception here since it negatively affects the probability of trade.

[23]Several studies have documented that the effect of distance in gravity models is overstated since distance is correlated with other trade frictions (such as lack of information). The same issue applies here, and would even further reduce the directly measured effect of distance.

the remaining coefficients are very similar to those obtained in the benchmark specification of column 1. Thus, the bias corrections implemented in our second stage estimation are dominated by the influence of unobserved firm heterogeneity rather than sample selection. This finding suggests that while aggregate country-pair shocks do have a significant effect on trade patterns, they only negligibly affect the responsiveness of trade volumes to observed trade barriers.[24] The results in column 3 clearly show that this is not the case for the effects of unobserved heterogeneity: the latter would affect trade volumes even were all country pairs trading with one-another, since it operates independently of the selection effect. Neglecting to control for this unobserved heterogeneity induces most of the biases exhibited in the standard gravity specification.

Evidence on Asymmetric Trade Relationships

As was previously mentioned, our model predicts asymmetric trade flows between countries. These asymmetries can be extreme, with trade predicted in only one direction, as also reflected in the data. More nuanced, trade can be positive in both directions, but with a net trade imbalance. Figure 3 graphically represents the extent of the predicted trade asymmetries by plotting the predicted probability of export between country pairs ($\hat{\rho}_{ij}$ versus $\hat{\rho}_{ji}$). The predicted asymmetries are clearly large, as measured by the distance from the diagonal for a substantial proportion of country pairs. Do these predicted asymmetries have explanatory power for the direction of trade flows and net bilateral trade balances? The answer is an overwhelming yes, as evidenced by the results reported in Table 4. The first part of the table shows the results of the OLS regression of $T_{ij} - T_{ji}$ on $\hat{\rho}_{ij} - \hat{\rho}_{ji}$ (based on the Probit results for 1986). Note that the regressand, $T_{ij} - T_{ji}$, takes on the values -1, 0, 1, depending on the direction of trade between i and j (it is 0 if trade flows in both directions or if the countries do not trade at all). The magnitude of the regressor $\hat{\rho}_{ij} - \hat{\rho}_{ji}$ measures the model's prediction for an asymmetric trading relationship, while its sign predicts the direction of the asymmetry. Table 4 shows that the predicted asymmetries have a substantial amount of explanatory power; the regressor coefficient is significant at any conventional level and explains on its own 23% of the variation in the direction of trade.[25] We emphasize that the regressor is constructed only from the predicted probability of export $\hat{\rho}_{ij}$, which is a function only of country level variables (the fixed effects) and *symmetric* bilateral measures.

The second part of Table 4 shows the results of the OLS regression of net bilateral trade $m_{ij} - m_{ji}$ (the percentage difference between exports and imports) on $\hat{\bar{w}}^*_{ij} - \hat{\bar{w}}^*_{ji}$ (only for those country pairs trading in both directions). This regressor captures differences in the pro-

[24] This finding also highlights the important information conveyed by the non-trading country pairs. If such zero trade values were just the outcome of censoring, then a Tobit specification would provide the best fit to the data. This is just a more restrictive version of the selection model, which is rejected by the data in favor of the specification incorporating firm heterogeneity.

[25] This understates the variable's explanatory power as it is continuous and predicting a discrete variable.

portion of exporting firms. Combined with the country fixed effects, these variables capture differences in the number of exporting firms from one country to the other. Again, we find that this single regressor is a strong predictor of net bilateral trade. On its own, it explains 16% of the variance in net trade, and along with the country fixed effects it explains 30% of that variance.

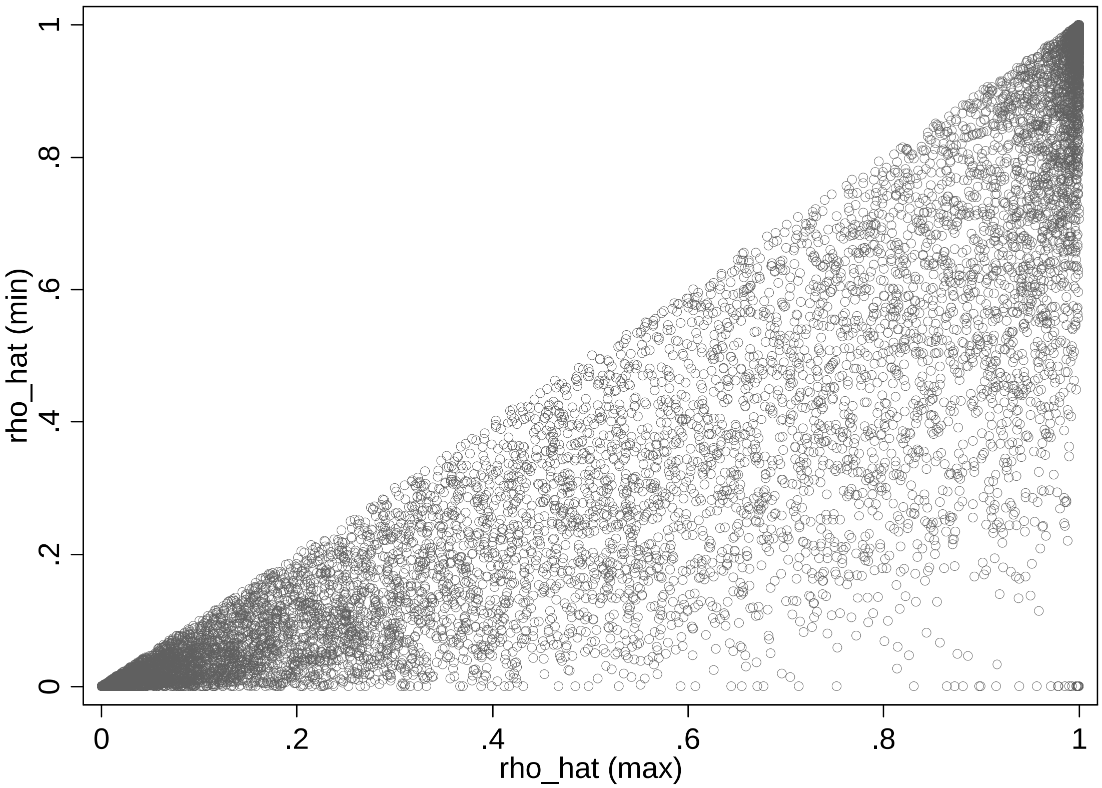

Figure 3: Predicted Asymmetries: $\min(\hat{\rho}_{ij}, \hat{\rho}_{ji})$ versus $\max(\hat{\rho}_{ij}, \hat{\rho}_{ji})$

Appendix A

We describe in this appendix our data sources.

Trade data

The bilateral trade flows are from Feenstra's "World Trade Flows, 1970-1992" and "World Trade Flows, 1980-1997". These data include 183 "country titles" over the period 1970 to 1997. In some cases Feenstra grouped several countries into a single title. We excluded 12 such titles, which we found difficult to identify with a particular country. This left usable data for bilateral trade flows among 161 countries. The list of these countries is provided at the end of this appendix.

For these 161 countries, we constructed a matrix of trade flows, measured in U.S. dollars. This matrix represents $161 \times 160 = 25,760$ trade flows, consisting of exports from country j to country i. Many of these export flows are zeros.

Country-level data

Population and real GDP per capita have been obtained from three standard sources: the Penn World Tables 6.1, the World Bank, and the IMF. *[specify variables and sources]*

We used the CIA's World Factbook to construct a number of variables, which can be classified as follows:[26]

1. **Geography** Latitude, longitude, and whether a country is landlocked or an island.

2. **Institutions** Legal origin, colonial origin, GATT/WTO membership.

3. **Culture** Primary language and religion. The later is represented by a vector, consisting of the fractions of people belonging to various religions, such as Catholic, Muslim, Protestant, and other.

We also used data from Rose (2000) to identify whether a country belongs to a currency union.

Using these data, we constructed country-pair specific variables, such as the distance between countries i and j, whether they share a border, the same legal system, the same colonial origin, or membership in the GATT/WTO.

[26] See http://www.cia.gov/cia/publications/factbook/docs/profileguide.html.

Appendix B

We derive in this appendix a gravity equation with third-country effects, which generalizes Anderson and van Wincoop's (2003) equation, and we show that their equation applies whenever $\tau_{ij} = \tau_{ji}$ for every country pair and V_{ij} can be decomposed in a particular way. We then discuss some limitations of their formulation.

Equality of income and expenditure implies $Y_i = \sum_{j=1}^{J} M_{ji}$. That is, country i's exports to all countries, including sales to home residents M_{ii}, equals the value of country i's output. Equation (6) then implies

$$Y_j = \left(\frac{c_j}{\alpha}\right)^{1-\varepsilon} N_j \sum_h \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} Y_h V_{hj} \ . \tag{B1}$$

Using this expression we can rewrite the bilateral trade volume (6) as

$$M_{ij} = \frac{Y_i Y_j}{Y} \frac{\left(\frac{\tau_{ij}}{P_i}\right)^{1-\varepsilon} V_{ij}}{\sum_{h=1}^{J} \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} V_{hj} s_h} \ , \tag{B2}$$

where $Y = \sum_{j=1}^{J} Y_j$ is world income and $s_h = Y_h / Y$ is the share of country h in world income.

We next show that if V_{ij} is decomposable in a particular way, and transport costs are symmetric (i.e., $\tau_{ij} = \tau_{ji}$ for all i and j), then (B2) yields the generalized gravity equation that has been derived by Anderson and van Wincoop (2003). Their specification satisfies these conditions. Importantly, however, there are other cases of interest, less restrictive than the Anderson and van Wincoop specification, that satisfy them too. Therefore, our derivation of the gravity equation shows that it applies under wider circumstances, and in particular, when there is productivity heterogeneity across firms and firms bear fixed costs of exporting. Under these circumstances only a fraction of the firms export; those with the highest productivity. Finally, note that our general formulation — without decomposability — is more relevant for empirical analysis, because, unlike previous formulations, it enables bilateral trade flows to equal zero. This flexibility is important because, as we have explained in the introduction, there are many zero bilateral trade flows in the data.

Consider the following

Decomposability Assumption V_{ij} is decomposable as follows:

$$V_{ij} = \left(\varphi_{IM,i} \varphi_{EX,j} \varphi_{ij}\right)^{1-\varepsilon} ,$$

where $\varphi_{IM,i}$ depends only on the parameters of the importing country, $\varphi_{EX,j}$ depends only on the parameters of the exporting country, and $\varphi_{ij} = \varphi_{ji}$ for all i, j.

In this decomposition, only the symmetric terms φ_{ij} depend on the joint identity of the importing and exporting countries, whereas all other parameters do not.

To illustrate circumstances in which the decomposability assumption is satisfied, first consider a situation where the fixed costs f_{ij} are very small, so that $a_{ij} > a_H$ for all i, j. That is, the lowest productivity level that makes exporting profitable, $1/a_{ij}$, is lower than the lowest productivity level in the support of $G(\cdot)$, $1/a_H$. Under these circumstances all firms export and V_{ij} is the same for every country pair i, j.[27] Alternatively, suppose that productivity $1/a$ has a Pareto distribution with shape k and $a_L = 0$. That is, $G(a) = (a/a_H)^k$ for $0 \leq a \leq a_H$. Moreover, let either f_{ij} depend only on the identity of the exporter, so that $f_{ij} = f_j$, or let the fixed costs be symmetric, so that $f_{ij} = f_{ji}$. Then V_{ij} satisfies the decomposability assumption and in every country j only a fraction of firms export to country i.[28]

Using the decomposability property and symmetry requirements $\tau_{ij} = \tau_{ji}$ and $\varphi_{ij} = \varphi_{ji}$, we obtain[29]

$$\frac{M_{ij}}{Y} = s_i s_j \left(\frac{\tau_{ij}\varphi_{ij}}{Q_i Q_j}\right)^{1-\varepsilon}, \tag{B3}$$

where the values of Q_j are solved from

$$Q_j^{1-\varepsilon} = \sum_h \left(\frac{\tau_{jh}\varphi_{jh}}{Q_h}\right)^{1-\varepsilon} s_h. \tag{B4}$$

This is essentially the Anderson and van Wincoop (2003) system. Evidently, the solution of

[27] More precisely, $V_{ij} = \int_{a_L}^{a_H} a^{1-\varepsilon} dG(a)$.

[28] Under these conditions $V_{ij} = k (a_{ij})^{k-\varepsilon+1} / (a_H)^k (k - \varepsilon + 1)$ and either $a_{ij} = [c_j f_j / (1-\alpha)]^{1/(1-\varepsilon)} / (\tau_{ij} c_j / \alpha P_i)$, so that f_j becomes part of $v_{EX,j}$ whereas τ_{ij} becomes part of ϕ_{ij}, or $a_{ij} = [c_j f_{ij} / (1-\alpha)]^{1/(1-\varepsilon)} / (\tau_{ij} c_j / \alpha P_i)$, so that f_{ij} and τ_{ij} become part of ϕ_{ij}.

[29] Decomposability allows us to rewrite (B2) as

$$M_{ij} = \frac{Y_i Y_j}{Y} \left(\frac{\tau_{ij}\varphi_{ij}}{Q_i \hat{Q}_j}\right)^{1-\varepsilon}, \tag{F1}$$

where $Q_i = P_i / \varphi_{IM,i}$ and

$$\hat{Q}_j^{1-\varepsilon} = \sum_h \left(\frac{\tau_{hj}\varphi_{hj}}{Q_h}\right)^{1-\varepsilon} s_h\ . \tag{F2}$$

In addition, (7) and (B1) imply

$$Q_i^{1-\varepsilon} = \sum_h \left(\frac{c_h \tau_{ih}\varphi_{ih}}{\alpha}\right)^{1-\varepsilon} N_h \left(\varphi_{EX,h}\right)^{1-\varepsilon}\ ,$$

$$s_j = \left(\frac{c_j}{\alpha}\right)^{1-\varepsilon} N_j \left(\varphi_{EX,h}\right)^{1-\varepsilon} \hat{Q}_j^{1-\varepsilon}.$$

Therefore

$$Q_j^{1-\varepsilon} = \sum_h \left(\frac{\tau_{jh}\varphi_{jh}}{\hat{Q}_h}\right)^{1-\varepsilon} s_h\ . \tag{F3}$$

Equations (F2) and (F3) together with symmetry conditions $\tau_{ij} = \tau_{ji}$ and $\varphi_{ij} = \varphi_{ji}$ then imply that $Q_j = \hat{Q}_j$ for every j. As a result (F1) and (F2) yield the equations in the text.

the Q_js depends only on income shares and transport costs, and possibly on a constant in V_{ij} that is embodied in the φ_{ij}s. However, an upward shift of this constant raises proportionately the product $Q_i Q_j$, and therefore has no effect on M_{ij}. Therefore, imports of country i from j as a share of world income, which equal imports of country j from i as a share of world income, depend only on the structure of trade costs and the size distribution of countries. Bilateral imports as a fraction of world income are proportional to the product of the two countries' shares in world income, with the factor of proportionality depending on the structure of trading costs and the worldwide distribution of relative country size.

The decomposability assumption is too restrictive, however. It implies that if imports of country i from j equal zero, i.e., $V_{ij} = 0$, then either $\varphi_{IM,i}$ is infinite or $\varphi_{EX,j}$ is infinite, because $\varepsilon > 1$. In the former case imports of country i equal zero from all countries, while in the latter case exports of country j equal zero to all countries. In other words, some countries do not import at all while other countries do not export at all; but it is not possible for a country to import from some other countries but not from all of them or for a county to export to some other countries but not to all of them. These restrictions are not consistent with the data. As we have explained in the introduction, *most* countries trade only with a fraction of the countries in the world economy; neither with all of them nor with none of them. To explain these patterns, we need a flexible model that allows for zero bilateral trade flows. Such a model should help in explaining which countries trade with each other and the resulting volumes of bilateral trade flows. Indeed, the logic of our theoretical model suggests that the decision to export to a foreign country is *not* independent of the volume of exports. For this reason the decision to export should be analyzed in conjunction with the decision on the export volume. Moreover, unlike (B3) and (B4), a suitable model should allow country j's exports to i to differ from country i's exports to j. Unlike standard estimation procedures of the gravity equations, a model of this sort will enable estimation that takes advantage of all the observations in the data, not only observations of country pairs that have positive two-way bilateral trade flows. For these reasons we use the less restrictive equations (4)-(7) for estimation purposes.

References

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

[2] Anderson, James E. and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, Vol. 93, pp. 170-92.

[3] Anderson, James E. and Eric van Wincoop (2004), "Trade Costs," NBER Working Paper No. 10480.

[4] Bernard, Andrew. B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum (2003): "Plants and Productivity in International Trade," *American Economic Review*, Vol. 93, pp. 1268-1290.

[5] Davis, Donald R. and David E. Weinstein (2003), "Market Access, Economic Geography and Comparative Advantage: An Empirical Test," *Journal of International Economics*, Vol. 59, pp. 1-23.

[6] Eaton, Jonathan and Samuel S. Kortum (2002), "Technology, Geography, and Trade," *Econometrica*, Vol. 70, pp. 1741-1779.

[7] Evans, Carolyn L. (2003), "The Economic Significance of National Border Effects," *American Economic Review*, Vol. 93, pp. 1291-1312.

[8] Evenett, Simon J. and Anthony J. Venables (2002), "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows," Mimeo.

[9] Feenstra, Robert C. (2002), "Border Effects and the Gravity Equation: Consistent Methods for Estimation," *Scottish Journal of Political Economy*, Vol. 49, pp. 491-506.

[10] Feenstra, Robert C. (2003), *Advanced International Trade* (Princeton: Princeton University Press).

[11] Frankel, Jeffrey A. and David Romer (1996), "Trade and Growth: An Empirical Investigation," NBER Working Paper No. 5476.

[12] Frankel, Jeffrey A. and David Romer (1999), "Does Trade Cause Growth?" *American Economic Review*, Vol. 89, pp. 379-99.

[13] Hall, Robert E. and Charles I. Jones (1999), "Why do Some Countries Produce so Much More Output per Worker than Others?," *Quarterly Journal of Economics*, Vol. 114, pp. 83-116.

[14] Haveman, Jon and David Hummels (2004), “Alternative Hypotheses and the Volume of Trade: The Gravity Equation and the Extent of Specialization,” *Canadian Journal of Economics*, Vol. 37, 199-218.

[15] Helpman, Elhanan (1987), "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries," *Journal of the Japanese and International Economics*, Vol. 1, pp. 62-81.

[16] Helpman, Elhanan and Paul R. Krugman (1985), *Market Structure and Foreign Trade* (Cambridge, MA: The MIT Press).

[17] McCallum, John (1995), "National Borders Matter: Canada–U.S. Regional Trade Patterns," *American Economic Review*, Vol. 85, pp. 615-23.

[18] Melitz, Marc J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, Vol. 71, pp. 1695-1725.

[19] Persson, Torsten and Guido Tabellini (2003), *The Economic Effects of Constitutions* (Cambridge, MA: The MIT Press).

[20] Rose, Andrew K. (2000), "One Money One Market: Estimating the Effect of Common Currencies on Trade," *Economic Policy*, Vol. 15, pp. 7-46.

[21] Rose, Andrew K. (2004), "Do We Really Know that the WTO Increases Trade?," *American Economic Review*, Vol. 94, pp. 98-114.

[22] Sachs, Jeffrey D. and Andrew Warner (1995), “Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity*, No. 1, pp. 1-118.

[23] Silva, J.M.C. Santos and Silvana Tenreyro, "Gravity-Defying Trade," mimeo, 2003.

[24] Subramanian, Arvind and Shang-Jin Wei (2003), “The WTO Promotes Trade, Strongly But Unevenly,” NBER Working Paper 10024.

[25] Tinbergen, Jan (1962), *Shaping the World Economy* (New York: The Twentieth Century Fund).

[26] Tenreyro, Silvana and Robert Barro (2003), "Economic Effects of Currency Unions," National Bureau of Economic Research, Working Paper 9435.

[27] Wei, Shang-Jin (1996), "Intra-national Versus International Trade: How Stubborn are Nations in Global Integration?" NBER, Working Paper No. 5531.

	1986			1980s					
	1	2		3	4		5	6	
	m_ij	T_ij (Probit)		m_ij	T_ij (Probit)		m_ij	T_ij (Probit)	
Variables		Coeff.	dF/dX		Coeff.	dF/dX		Coeff.	dF/dX
Distance	-1.176	-0.660	-0.263	-1.201	-0.618	-0.246	-1.200	-0.618	-0.246
	(0.031)**	(0.029)**	(0.012)**	(0.024)**	(0.021)**	(0.008)**	(0.024)**	(0.021)**	(0.008)**
Land border	0.458	-0.382	-0.148	0.366	-0.380	-0.146	0.364	-0.380	-0.146
	(0.147)**	(0.129)*	(0.047)*	(0.131)**	(0.089)**	(0.032)**	(0.131)**	(0.089)**	(0.032)**
Island	-0.391	-0.345	-0.136	-0.381	-0.355	-0.140	-0.378	-0.355	-0.140
	(0.121)**	(0.082)**	(0.032)**	(0.096)**	(0.056)**	(0.022)**	(0.096)**	(0.056)**	(0.022)**
Landlock	-0.561	-0.181	-0.072	-0.582	-0.220	-0.087	-0.581	-0.221	-0.087
	(0.188)**	(0.114)	(0.045)	(0.148)**	(0.071)**	(0.028)**	(0.147)**	(0.071)**	(0.028)**
Legal	0.486	0.096	0.038	0.406	0.072	0.029	0.407	0.071	0.028
	(0.050)**	(0.034)*	(0.014)*	(0.040)**	(0.022)**	(0.009)**	(0.040)**	(0.022)**	(0.009)**
Language	0.176	0.284	0.113	0.207	0.275	0.109	0.203	0.273	0.108
	(0.061)**	(0.042)**	(0.016)**	(0.047)**	(0.027)**	(0.011)**	(0.047)**	(0.027)**	(0.011)**
Religion	0.102	0.261	0.104	-0.018	0.249	0.099	-0.038	0.245	0.098
	(0.096)	(0.063)**	(0.025)**	(0.076)	(0.040)**	(0.016)**	(0.077)	(0.040)**	(0.016)**
Colonial Ties	1.299	0.325	0.128	1.321	0.288	0.114	1.326	0.293	0.116
	(0.120)**	(0.305)	(0.117)	(0.110)**	(0.209)	(0.082)	(0.110)**	(0.211)	(0.082)
Currency Union	1.364	0.492	0.190	1.395	0.530	0.206	1.409	0.531	0.206
	(0.255)**	(0.143)**	(0.052)**	(0.187)**	(0.071)**	(0.026)**	(0.187)**	(0.071)**	(0.026)**
FTA	0.759	1.985	0.494	0.996	1.854	0.497	0.976	1.842	0.495
	(0.222)**	(0.315)**	(0.020)**	(0.213)**	(0.207)**	(0.018)**	(0.214)**	(0.207)**	(0.018)**
WTO (none)							-0.068	-0.143	-0.056
							(0.058)	(0.033)**	(0.013)**
WTO (both)							0.303	0.234	0.093
							(0.042)**	(0.032)**	(0.013)**
Observations	11,146	24,649	24,649	110,697	248,060	248,060	110,697	248,060	248,060
R-Squared	0.709	0.587	0.587	0.682	0.551	0.551	0.682	0.551	0.551

Notes:
Exporter, Importer, and year fixed effects
Robust standard errors (clustering by country pair)
* significant at 5%; ** significant at 1%

Table 1

	1986			1980s		
	T_ij	m_ij		T_ij	m_ij	
Variables	(Probit)	Benchmark	ML	(Probit)	Benchmark	ML
Distance	-0.660 (0.029)**	-1.181 (0.031)**	-0.801 (0.030)**	-0.618 (0.021)**	-1.198 (0.024)**	-0.822 (0.024)**
Land border	-0.382 (0.129)*	0.468 (0.146)**	0.831 (0.139)**	-0.380 (0.089)**	0.360 (0.131)**	0.702 (0.123)**
Island	-0.345 (0.082)**	-0.387 (0.120)**	-0.171 (0.117)	-0.355 (0.056)**	-0.379 (0.096)**	-0.143 (0.094)
Landlock	-0.181 (0.114)	-0.556 (0.188)**	-0.448 (0.187)*	-0.221 (0.071)**	-0.582 (0.147)**	-0.440 (0.147)**
Legal	0.096 (0.034)*	0.490 (0.050)**	0.388 (0.049)**	0.071 (0.022)**	0.406 (0.040)**	0.327 (0.039)**
Language	0.284 (0.042)**	0.187 (0.061)*	0.024 (0.06)	0.273 (0.027)**	0.198 (0.047)**	0.033 (0.046)
Religion	0.261 (0.063)**	--	--	0.245 (0.040)**	--	--
Colonial Ties	0.325 (0.305)	1.299 (0.121)**	1.003 (0.114)**	0.293 (0.211)	1.326 (0.110)**	1.061 (0.106)**
Currency Union	0.492 (0.143)**	1.356 (0.256)**	1.026 (0.258)**	0.531 (0.071)**	1.412 (0.187)**	1.034 (0.191)**
FTA	1.985 (0.315)**	0.756 (0.222)**	0.386 (0.171)*	1.842 (0.207)**	0.978 (0.214)**	0.519 (0.148)**
WTO (none)	--	--	--	-0.143 (0.033)**	-0.070 (0.058)	0.001 (0.058)
WTO (both)	--	--	--	0.234 (0.032)**	0.302 (0.042)**	0.143 (0.042)**
delta (from w_hat)	--	--	0.716 (0.060)**	--	--	0.794 (0.067)**
eta_hat	--	--	0.399 (0.063)**	--	--	0.270 (0.049)**
Observations	24,649	11,146	11,146	248,060	110,697	110,697
R-Squared	0.587	0.709	--	0.551	0.682	--

Notes:
Exporter, Importer, and year fixed effects
Robust standard errors (clustering by country pair)
* significant at 5%; ** significant at 1%

Table 2

	Dependent variable: m_ij			
Variables	Benchmark	ML	Firm Heterogeneity	Heckman Selection
Distance	-1.181	-0.801	-0.824	-1.214
	(0.031)**	(0.030)**	(0.036)**	(0.031)**
Land border	0.468	0.831	0.807	0.436
	(0.146)**	(0.139)**	(0.139)**	(0.149)**
Island	-0.387	-0.171	-0.148	-0.425
	(0.120)**	(0.117)	(0.119)	(0.120)**
Landlock	-0.556	-0.448	-0.450	-0.565
	(0.188)**	(0.187)*	(0.190)*	(0.187)**
Legal	0.490	0.388	0.420	0.488
	(0.050)**	(0.049)**	(0.050)**	(0.050)**
Language	0.187	0.024	-0.008	0.223
	(0.061)**	(0.06)	(0.061)	(0.061)**
Colonial Ties	1.299	1.003	1.051	1.311
	(0.121)**	(0.114)**	(0.114)**	(0.123)**
Currency Union	1.356	1.026	1.028	1.391
	(0.256)**	(0.258)**	(0.256)**	(0.257)**
FTA	0.756	0.386	0.502	0.737
	(0.222)**	(0.171)*	(0.160)**	(0.235)**
delta (from w_hat)	--	0.716	--	--
		(0.060)**		
eta_hat	--	0.399	--	0.265
		(0.063)**		(0.070)**
z_hat	--	--	0.611	--
			(0.043)**	
Observations	11,146	11,146	11,146	11,146
R-squared	0.709	--	0.713	0.710

Notes:
All data for 1986
Exporter and Importer fixed effects
Robust standard errors (clustering by country pair)
* significant at 5%; ** significant at 1%

Table 3

Variable	T_ij - T_ji
rho_hat_ij - rho_hat_ji	0.994 (0.023)**
Country Fixed Effects	No
Observations	12403
R-Square	0.228

Variable	m_ij - m_ji	
w_hat_ij - w_hat_ji	2.073 (0.079)**	1.820 (0.320)**
Country Fixed Effects	No	Yes
Observations	4652	4652
R-Square	0.156	0.299

Notes:
All data for 1986
* significant at 5%; ** significant at 1%

Table 4