THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. CXXIII May 2008 Issue 2

ESTIMATING TRADE FLOWS: TRADING PARTNERS AND TRADING VOLUMES*

Elhanan Helpman
Marc Melitz
Yona Rubinstein

We develop a simple model of international trade with heterogeneous firms that is consistent with a number of stylized features of the data. In particular, the model predicts positive as well as zero trade flows across pairs of countries, and it allows the number of exporting firms to vary across destination countries. As a result, the impact of trade frictions on trade flows can be decomposed into the intensive and extensive margins, where the former refers to the trade volume per exporter and the latter refers to the number of exporters. This model yields a generalized gravity equation that accounts for the self-selection of firms into export markets and their impact on trade volumes. We then develop a two-stage estimation procedure that uses an equation for selection into trade partners in the first stage and a trade flow equation in the second. We implement this procedure parametrically, semiparametrically, and nonparametrically, showing that in all three cases the estimated effects of trade frictions are similar. Importantly, our method provides estimates of the intensive and extensive margins of trade. We show that traditional estimates are biased and that most of the bias is due not to selection but rather due to the omission of the extensive margin. Moreover, the effect of the number of exporting firms varies across country pairs according to their characteristics. This variation is large and particularly so for trade between developed and less developed countries and between pairs of less developed countries.

I. Introduction

Estimation of international trade flows has a long tradition. Tinbergen (1962) pioneered the use of gravity equations in

* We thank Costas Arkolakis, Robert Barro, Moshe Buchinsky, Zvi Eckstein, Gene Grossman, Bo Honore, Larry Katz, Marcelo Moreira, Ariel Pakes, Jim Powell, Manuel Trajtenberg, Zhihong Yu, and three referees for comments. Dror Brenner and Brent Neiman provided superb research assistance. Helpman thanks the NSF for financial support. Melitz thanks the NSF and the Sloan Foundation for financial support and the International Economics Section at Princeton University for its hospitality.

The Quarterly Journal of Economics, May 2008

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 supplemented with theoretical underpinnings and better estimation techniques.[1] The gravity equation has dominated empirical research in international trade; it has been used to estimate the impact on trade flows of international borders, preferential trading blocs, currency unions, and membership in the WTO, as well as the size of home-market effects.[2]

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 nontrading countries alike.[3]

The next section briefly reviews the evolution of the volume of trade among the 158 countries in our sample and the composition of country pairs according to their trading status.[4] Three features stand out. First, about half of the country pairs do not trade with

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

2. See McCallum (1995) for the study that triggered an extensive debate on the role of international borders, as well as Wei (1996), and Anderson and van Wincoop (2003), and Evans (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 (2003) on currency unions, Rose (2004) on WTO membership, and Davis and Weinstein (2003) on the size of home-market effects.

3. Evenett and Venables (2002), Anderson and van Wincoop (2004), 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.

4. See Appendix I for data sources.

one another.[5] 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.[6] 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 that did not. Nevertheless, we show in Section VI 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 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.[7]

We develop in Section III 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 for exports to countries with higher demand levels, lower variable export costs, and lower fixed export costs. Positive trade flows from country j to country i thus aggregate exports over varying distributions of firms. Each distribution is bounded by a marginal exporter in j who just breaks even by exporting to i. Country j firms with higher productivity levels generate positive profits from exports to i.

This model has a number of implications for trade flows. First, no firm from country j may be productive enough to profitably export to country i. 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

5. We say that a country pair i and j do not trade with one another if i does not export to j and j does not export to i.

6. Felbermayr and Kohler (2006) report that prior to 1970 new trade flows contributed substantially to the growth of world trade.

7. The role of the number of exported products, as opposed to exports per product, has been found to be important in a number of studies. To illustrate, Hummels and Klenow (2005) find that 60% of the greater export of larger economies in their sample of 126 exporting countries is due to variation in the number of exported products, and Kehoe and Ruhl (2002) find that during episodes of trade liberalization in 18 countries a large fraction of trade expansion was driven by trade in goods that were not traded before.

from i to j for some country pairs. Both types of trade patterns exist in the data. Second, the model predicts positive—though asymmetric—trade flows in both directions for some country pairs, which are also needed 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 and thus predicts an extensive margin for trade flows. 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 IV develops the empirical framework for estimating the gravity equation derived in Section III. 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 sample selection bias and a bias from potential asymmetries in the trade flows between pairs of countries. The latter bias is due to an omitted variable that measures the impact of the number (fraction) of exporting firms, that is, the extensive margin of trade. Because this procedure is easy to implement, it can be effectively used in many applications.

Our theoretical model has firm heterogeneity, yet we do not need firm-level data to estimate the gravity equation. This property results from the fact that the characteristics of the marginal exporters to different destinations can be identified from the variation in features of the destination countries and of observable bilateral trade costs. As a result, there exist sufficient statistics,

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

which can be computed from aggregate data, that predict the selection of heterogeneous firms into export markets and their associated aggregate trade volumes.[9] This is an important advantage of our approach, which extracts from country-level data information that would normally require firm-level data. Although more firm-level data sets have become available over time, it is not yet possible to pool them together into a comprehensive data set that can be used for cross-country estimation purposes.

Section V 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 Sections VI and VII. In Section VI, we estimate the model on a partial sample of countries for which we have data on regulation of entry costs, which we use as the excluded variables in the two-stage estimation procedure. We argue that these variables satisfy the exclusion restrictions on theoretical grounds. In Section VII, we use this reduced sample to test for the validity of other potential excluded variables, which are available for virtually all country pairs, representing a substantial increase in sample size. We show that an index for common religion (across country pairs) satisfies the exclusion restrictions for this sample. We then reestimate our model on the full sample of countries using this common religion index as the excluded variable. This approximately doubles the number of usable observations. This substantial increase in sample size is the main motivation behind our construction of the religion variable in the first place.

In both Sections VI and VII, we implement three estimation methods, progressively relaxing some parameterization assumptions: nonlinear least squares, semiparametric, and nonparametric. The nonlinear least squares (NLS) version of the two-stage procedure uses functional forms derived from the theoretical model under the assumption that productivity follows a truncated Pareto distribution. We show that the corrections for the selection

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 et al. (2003) use direct information on U.S. plant-level sales, productivity, and export status to calibrate a model that is then used to simulate the extensive and intensive margins of bilateral trade flows.

and omitted variable biases have a measurable downward impact on the estimated coefficients. Moreover, the extent of this bias is not sensitive to the use of the alternative excluded variables. The nature and extent of this bias is further confirmed when we estimate the model in the other two alternative ways: first with a semiparametric method, where we replace the truncated Pareto distribution for firm productivity with a general distribution approximated by a polynomial fit, and second with a nonparametric method, which further relaxes the joint normality assumption for the unobserved trade costs. In both cases, we obtain results very similar to our fully parametrized NLS specification. An additional advantage of the latter two methods is that they can be easily implemented using OLS in the second stage.

A number of additional insights from our estimates are discussed in Section VIII. First, we show that most of the bias is due to the omitted correction for the extensive margin of trade and not due to the selection bias. In fact, the selection bias is economically negligible though statistically (strongly) significant. Second, we show that the asymmetric impact of the extensive margin of trade is important in explaining the asymmetries in trade flows observed in the data. Finally, we show that the biases not only are large, but also systematically vary with the characteristics of trade partners. For this purpose we perform a counterfactual exercise in which trade frictions are reduced. A reduction in these frictions induces trade among country pairs that did not trade before and raises trade volumes among country pairs with existing trade relations. When countries are partitioned by income (high versus low), we find that the impact of reduced trade frictions differs substantially across country pairs according to these income levels. The elasticity of trade with respect to such frictions can vary by a factor of three. That is, it can be three times larger for some country pairs than for others. This highlights both the size, and also the large variations in the biases across country pairs. Section IX concludes.

II. A Glance at the Data

Figure I 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);

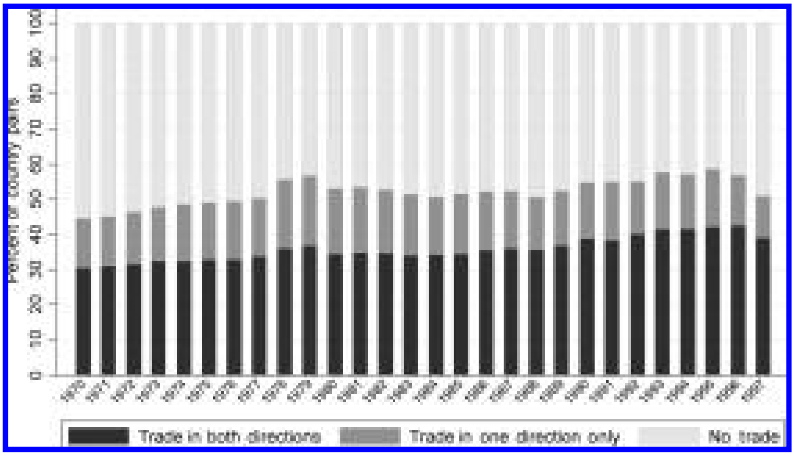

FIGURE I
Distribution of Country Pairs Based on Direction of Trade
Note. Constructed from 158 countries.

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 II shows the evolution of the aggregate real volume of exports of all 158 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 or traded in one direction only in 1970. 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 contribution to the

10. Silva and Tenreyro (2006) also argue that zero trade flows can be used in the estimation of the gravity equation, but they emphasize a heteroscedasticity 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. This finding is consistent with our finding that the selection bias is rather small.

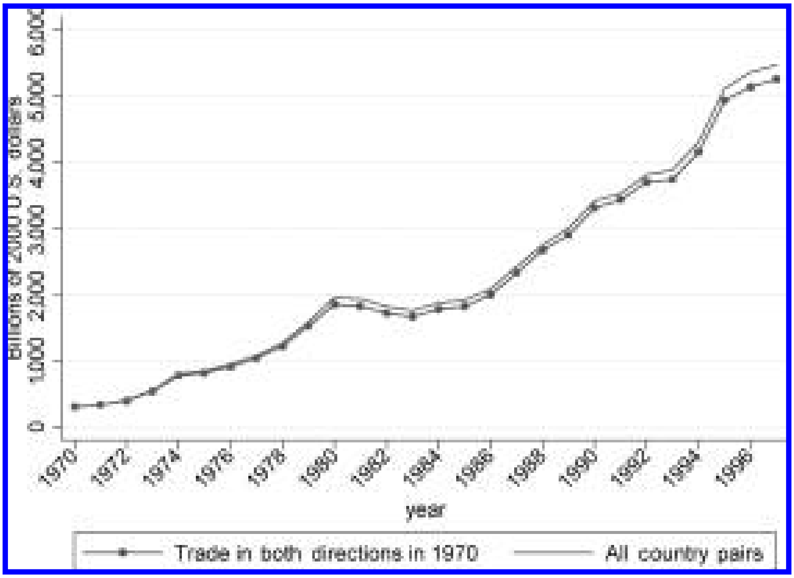

FIGURE II
Aggregate Volume of Exports of All Country Pairs and of Country Pairs That Traded in Both Directions in 1970

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 I, 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 enlargement of the set of trading countries did not contribute in a major way to the growth of world trade.[11]

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 three-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. And this also contrasts

III. THEORY

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

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

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

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

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

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

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 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

with the finding that changes in the number of trading products has a measurable impact on trade flows (see Kehoe and Ruhl [2002] and Hummels and Klenow [2005]).

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 $a_H > a_L > 0$. 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 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. Because 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 $p_j(a) = c_j a/\alpha$. This is a standard markup pricing equation, with a smaller markup associated with a larger elasticity of demand. If this country-j producer of a product l sells to consumers in country i, it then sets a delivered price (in country i) equal to

$$\check{p}_j\,(l) = \tau_{ij}\frac{c_j a}{\alpha}. \tag{3}$$

As a result, the associated operating profits from these sales to country i are

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

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 a's only capture relative productivity differences across firms in a country. Aggregate productivity differences across countries are subsumed in the c_j's.

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

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

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 available in country i is smaller than the total set of products produced in the world economy. In addition, it is possible for $G(a_{ij})$ to be zero: no firm from country j finds it profitable to export to country i. This happens whenever $a_{ij} \leq a_L$: the least productive firm that can profitably export to country i has a coefficient a below the support of $G(a)$. We explicitly consider these cases that explain zero bilateral trade volumes. If a_{ij} were larger than a_H, then all firms from country j would export to i. However, given the pervasive firm-level evidence on the coexistence of exporting and nonexporting firms, even within narrowly defined sectors, we disregard this possibility.

We next characterize bilateral trade volumes. Let

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

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

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

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

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

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

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 II that, together with equality of income and expenditure, equations (4)–(7) can be used to derive a generalized version of Anderson and van Wincoop's (2003) gravity equation with third-country effects. This generalization applies when transport costs are symmetric ($\tau_{ij} = \tau_{ji}\ \forall i, j$) and 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 for that country pair. This decomposability holds in Anderson and van Wincoop's model. Importantly, however, there are other cases of interest with positive fixed export costs and an extensive margin of trade that also satisfy the generalized gravity equation. Yet even this more generalized version of the gravity equation cannot explain the documented pattern of zero trade flows and the bilateral trade asymmetries (see Appendix II for details). Thus, 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.

IV. Empirical Framework

We begin by formulating a fully parametrized estimation procedure for this model, which delivers our benchmark results. We then progressively loosen these parametric restrictions and reestimate the model. In all cases, we obtain similar results that are consistent with the analysis of the baseline scenario.

In the baseline specification, we assume that firm productivity $1/a$ is Pareto distributed, truncated to the support $[a_L, a_H]$. Thus, we assume $G(a) = (a^k - a_L^k)/(a_H^k - a_L^k)$, $k > (\varepsilon - 1)$. As previously highlighted, we allow for $a_{ij} < a_L$ for some i–j pairs, inducing zero exports from j to i (i.e., $V_{ij} = 0$ and $M_{ij} = 0$). This framework also 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

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

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 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]

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}, which controls for the fraction of firms (possibly zero) that export from j to i. This

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 matrix 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.

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 into 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.

Last, we emphasize again that in our formulation, bilateral trade flows need not be balanced, even when all bilateral trade barriers are symmetric. First and foremost, w_{ij} can be asymmetric. We document later in Section VIII that such asymmetries are empirically important and substantial. Second, the importer fixed effects may differ from the exporter fixed effects for given countries. This substantiates the use of directional trade flows and separate fixed effects for the exporting and the importing countries.

IV.A. 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

$$Z_{ij} = \frac{(1-\alpha)\left(P_i \frac{\alpha}{c_j \tau_{ij}}\right)^{\varepsilon-1} Y_i a_L^{1-\varepsilon}}{c_j f_{ij}}. \tag{10}$$

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}; that is, $W_{ij} = Z_{ij}^{(k-\varepsilon+1)/(\varepsilon-1)} - 1$ (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 $f_{ij} \equiv \exp(\phi_{EX,j} + \phi_{IM,i} + \kappa\phi_{ij} - \nu_{ij})$, 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

$$z_{ij} = \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa\phi_{ij} + \eta_{ij}, \tag{11}$$

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), $\xi_j = -\varepsilon\ln c_j + \phi_{EX,j}$ is an exporter fixed effect, and $\zeta_i = (\varepsilon - 1)p_i + y_i - \phi_{IM,i}$ is an importer fixed effect. 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. Because we do not want to impose $\sigma_\eta^2 \equiv \sigma_u^2 + \sigma_\nu^2 = 1$, we divide (11) by the standard deviation σ_η and specify the Probit equation

$$\begin{aligned} \rho_{ij} &= \Pr(T_{ij} = 1 \mid \text{observed variables}) \\ &= \Phi\left(\gamma_0^* + \xi_j^* + \zeta_i^* - \gamma^* d_{ij} - \kappa^*\phi_{ij}\right), \end{aligned} \tag{12}$$

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 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}.

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 (12) 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 (12) then capture the combined effect of these barriers.

Let $\hat{\rho}_{ij}$ be the predicted probability of exports from j to i, using the estimates from the Probit equation (12), and let $\hat{z}^*_{ij} = \Phi^{-1}(\hat{\rho}_{ij})$ be the predicted value of the latent variable $z^*_{ij} \equiv z_{ij}/\sigma_\eta$. Then a consistent estimate for W_{ij} can be obtained from

$$W_{ij} = \max\left\{\left(Z^*_{ij}\right)^\delta - 1, 0\right\}, \tag{13}$$

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

IV.B. 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 $E[w_{ij} \mid ., T_{ij} = 1]$ and $E[u_{ij} \mid ., T_{ij} = 1]$. Both terms depend on $\bar{\eta}^*_{ij} \equiv E[\eta^*_{ij} \mid ., T_{ij} = 1]$. Moreover, $E[u_{ij} \mid ., T_{ij} = 1] = \text{corr}\,(u_{ij}, \eta_{ij})(\sigma_u/\sigma_\eta)\bar{\eta}^*_{ij}$. Since η^*_{ij} has a unit normal distribution, a consistent estimate $\hat{\bar{\eta}}^*_{ij}$ is obtained from the inverse Mills ratio, that is, $\hat{\bar{\eta}}^*_{ij} = \phi(\hat{z}^*_{ij})/\Phi(\hat{z}^*_{ij})$. Therefore $\hat{\bar{z}}^*_{ij} \equiv \hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}$ is a consistent estimate for $E[z^*_{ij} \mid ., T_{ij} = 1]$ and $\hat{\bar{w}}^*_{ij} \equiv \ln\{\exp[\delta(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij})] - 1\}$ is a consistent estimate for $E[w_{ij} \mid ., T_{ij} = 1]$ (see (13)). We therefore can estimate (9) using the transformation

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + \ln\left\{\exp\left[\delta\left(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}\right)\right] - 1\right\} + \beta_{u\eta}\hat{\bar{\eta}}^*_{ij} + e_{ij}, \tag{14}$$

where $\beta_{u\eta} \equiv \text{corr}\,(u_{ij}, \eta_{ij})(\sigma_u/\sigma_\eta)$ and e_{ij} is an i.i.d. error term satisfying $E[e_{ij} \mid ., T_{ij} = 1] = 0$. Because (14) is nonlinear in δ, we estimate it using nonlinear least squares.

The use of $\hat{\bar{\eta}}^*_{ij}$ to control for $E[u_{ij} \mid ., T_{ij} = 1]$ 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}. However, 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 was not correlated with the export decision. Thus, 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.[19] Our empirical results highlight the overwhelming contribution of this channel relative to the standard correction for sample selection, which ignores firm-level heterogeneity.

To summarize, our theoretical framework delivers two equations, (11) and (14), which can be estimated in two stages. Although the theoretical model allows for arbitrary variation in bilateral variable and fixed trade costs, for estimation purposes we restrict these variations to $\tau_{ij}^{\varepsilon-1} \equiv D_{ij}^{\gamma} e^{-u_{ij}}$ and $f_{ij} \equiv \exp(\phi_{EX,j} + \phi_{IM,i} + \kappa\phi_{ij} - \nu_{ij})$, respectively. These restrictions make it possible to identify γ and δ, which are important parameters, but they do not make it possible to infer every parameter of the model. For example, we cannot separately identify the elasticity of demand ε. Evidently, it is necessary to impose more restrictions in order to gain additional identification.[20]

Before describing the empirical results, we pause to note that our distributional assumptions on the joint normality of the unobserved trade costs and the Pareto distribution of firm-level productivity affect the functional form of the trade flow equation (14) via the functional form of the two additional controls for firm heterogeneity ($\hat{w}_{ij}^{*}$) and sample selection ($\hat{\bar{\eta}}_{ij}^{*}$). After presenting our main results, we will describe a number of alternative specifications that relax these assumptions, yet generate very similar estimates. They illustrate the robustness of the findings in our baseline specification.

V. 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 I provides a representative estimate of this sort for all bilateral trade flows reported in 1986 from a set of 158 countries (the full list is reported in Appendix I). Note that instead of constructing symmetric trade flows by combining exports and imports for each country pair, we use the unidirectional trade value

19. Eaton, Kortum, and Kramarz (2004) find that more French firms export to larger foreign markets, and Bernard, Jensen, and Schott (2005) find a similar pattern for U.S. firms. Our model is consistent with these findings.

20. See, for example, Eaton and Kortum (2002) and Anderson and van Wincoop (2003) for ways to estimate this elasticity.

TABLE I
BENCHMARK GRAVITY AND SELECTION INTO TRADING RELATIONSHIPS

	1986		1980s			
Variables	m_{ij}	(Probit) T_{ij}	m_{ij}	(Probit) T_{ij}	m_{ij}	(Probit) T_{ij}
Distance	−1.176**	−0.263**	−1.201**	−0.246**	−1.200**	−0.246**
	(0.031)	(0.012)	(0.024)	(0.008)	(0.024)	(0.008)
Land border	0.458**	−0.148**	0.366**	−0.146**	0.364**	−0.146**
	(0.147)	(0.047)	(0.131)	(0.032)	(0.131)	(0.032)
Island	−0.391**	−0.136**	−0.381**	−0.140**	−0.378**	−0.140**
	(0.121)	(0.032)	(0.096)	(0.022)	(0.096)	(0.022)
Landlock	−0.561**	−0.072	−0.582**	−0.087**	−0.581**	−0.087**
	(0.188)	(0.045)	(0.148)	(0.028)	(0.147)	(0.028)
Legal	0.486**	0.038**	0.406**	0.029**	0.407**	0.028**
	(0.050)	(0.014)	(0.040)	(0.009)	(0.040)	(0.009)
Language	0.176**	0.113**	0.207**	0.109**	0.203**	0.108**
	(0.061)	(0.016)	(0.047)	(0.011)	(0.047)	(0.011)
Colonial ties	1.299**	0.128	1.321**	0.114	1.326**	0.116
	(0.120)	(0.117)	(0.110)	(0.082)	(0.110)	(0.082)
Currency union	1.364**	0.190**	1.395**	0.206**	1.409**	0.206**
	(0.255)	(0.052)	(0.187)	(0.026)	(0.187)	(0.026)
FTA	0.759**	0.494**	0.996**	0.497**	0.976**	0.495**
	(0.222)	(0.020)	(0.213)	(0.018)	(0.214)	(0.018)
Religion	0.102	0.104**	−0.018	0.099**	−0.038	0.098**
	(0.096)	(0.025)	(0.076)	(0.016)	(0.077)	(0.016)
WTO (none)					−0.068	−0.056**
					(0.058)	(0.013)
WTO (both)					0.303**	0.093**
					(0.042)	(0.013)
Observations	11,146	24,649	110,697	248,060	110,697	248,060
R^2	0.709	0.587	0.682	0.551	0.682	0.551

Notes. Exporter, importer, and year fixed effects. Marginal effects at sample means and pseudo R^2 reported for Probit. Robust standard errors (clustering by country pair).
+ Significant at 10%.
*Significant at 5%.
**Significant at 1%.

and introduce both importing and exporting country fixed effects. With these fixed effects every country pair is represented twice: one time for exports from i to j and another time for exports from j to i.[21] Nevertheless, the results in Table I 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

21. Among the $158 \times 157 = 24{,}806$ possible bilateral trading relationships, there are only 11,146 (less than half) positive trade flows.

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, or one country has colonized the other. The probability that two randomly drawn persons, one from each country, share the same religion raises export volumes.[22] Details on the construction of all the variables are provided in Appendix I.

We next estimate a Probit equation for the presence of a trading relationship using the same explanatory variables as the initial gravity specification (the specification follows (12), with exporter and importer fixed effects). The marginal effects, evaluated at the sample means, are reported in column (2).[23] 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 similar to that of a common language, increasing the probability of trade by 10% for the "typical" country pair). 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 IV.

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

22. The common religion variable is not used in traditional gravity equations. We have constructed it especially for use in our two-stage estimation procedure, as explained in the following sections.

23. The sample size is reduced from $158 \times 157 = 24{,}806$ to 24,649 because Congo did not export to anyone in 1986, and an exporter fixed effect cannot be estimated.

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 (2007), 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.[24] Our theoretical framework provides a 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 an impact on the probability of trade similar to common language or colonial ties.[25]

In reporting results for the 1980s, we aim to show that our choice of 1986 for the cross-section study does not affect the estimates. In other words, there is nothing special about 1986. And moreover, because this is mostly a methodological paper, we do not think that the choice of year is particularly important. Yet 1986 has the added advantage that it allows us to compare our results with French firm-level export data by destination reported in Eaton, Kortum, and Kramarz (2004) (see below).

VI. Two-Stage Estimation

We now turn to the second stage estimation of the trade flow equation (14). As we describe in Section IV, this requires a first-stage Probit selection equation (12) such as that reported in Table I, which yields a predicted probability of export $\hat{\rho}_{ij}$ (and thus the additional $\hat{\bar{w}}^*_{ij}$ and $\hat{\bar{\eta}}^*_{ij}$ controls). Because we do not want the identification of our second stage estimates to rely on the normality assumption for the unobserved trade costs, we also need to select valid excluded variables for that second stage (we will also relax these distributional assumptions through the use of nonparametric methods). Our theoretical model suggests that trade barriers that affect fixed trade costs but do not affect variable (per-unit) trade costs satisfy this exclusion restriction. We now describe the construction of such variables.

24. 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.

25. When two countries *both* join the WTO, their probability of trade increases by 15%.

We start with country-level data on the regulation costs of firm entry, collected and analyzed by Djankov et al. (2002). These entry costs are measured via their effects on the number of days, the number of legal procedures, and the relative cost (as a percentage of GDP per capita) for an entrepreneur to legally start operating a business.[26] We surmise (and confirm empirically) that they also affect the costs faced by exporting firms to/from that country, and that these costs are magnified when both exporting and importing countries impose high regulatory hurdles. By their nature, these measures affect firm-level fixed rather than variable costs of trade. We therefore construct an indicator for high fixed-cost trading country pairs, consisting of country pairs in which both the importing and exporting countries have entry regulation measures above the cross-country median. One variable uses the sum of the number of days and procedures above the median (for both countries) whereas the other uses the sum of the relative costs above the median (again for both countries).[27] By construction, these bilateral variables reflect regulation costs that should not depend on a firm's volume of exports to a particular country, and therefore satisfy the requisite exclusion restrictions.[28]

Using these additional variables for our first stage estimation of selection into trading relationships entails a substantial drop in sample size. First, 42 of 158 countries do not have any available regulation cost data.[29] Second, among the remaining countries, 8 of them export to everyone, and Japan imports from

26. Unfortunately, historic data were not available. For this reason, we use the data for 1999. See Djankov et al. (2002) for details.

27. Recall that these relative costs are measured as a percentage of GDP per capita, so these cost measures can be compared across countries. We could also have separated the number of days and procedures into separate variables, but we found that the jointly defined indicator variable has substantially more explanatory power.

28. Variable (per-unit) export costs at the country level could potentially be correlated with the fixed regulation costs associated with trade. However, our first stage estimation also includes country fixed effects. These correlated country-level variable costs would then have to interact in the same pattern as the fixed costs across country pairs in order to generate a correlation at the country level that is left uncontrolled by the country fixed effects. This possibility is substantially more remote than the potential correlation at the country level.

29. These 42 countries are Afghanistan, Bahamas, Bahrain, Barbados, Belize, Bermuda, Brunei, Cayman Islands, Comoros, Cuba, Cyprus, Djibouti, Equatorial Guinea, French Guiana, Gabon, Gambia, Greenland, Guadeloupe, Guinea-Bissau, Guyana, Iceland, Iraq, Kiribati, North Korea, Liberia, Libya, Maldives, Malta, Mauritius, Myanmar, New Caledonia, Qatar, Reunion, Seychelles, Somalia, St. Kitts, Sudan, Suriname, Trinidad-Tobago, Turks Caicos, Western Sahara, and Zaire.

everyone.[30] Fixed exporter (and in the case of Japan, importer) effects can thus not be estimated, and all the observations with that particular exporter (or importer) are dropped. Third, the number of observations decreases with the square of the number of dropped countries. Jointly, these factors account for the halving of the available observations. This substantial decrease has led us to statistically test the validity of the exclusion restriction for additional bilateral trade barriers available for our full sample of countries (see following section). For now, the most relevant issue for our estimation purposes is that the additional cost variables have substantial explanatory power for the formation of trading relationships. This is strongly confirmed by the results in the first column of Table II. We reran the same Probit equation (based on (12)) as previously reported in Table I, adding our two cost measures. The results for all the explanatory variables from Table I are roughly similar, and the two cost variables are economically and statistically significant.

We next estimate our fully parametrized trade flow equation (14) using nonlinear least squares (NLS). We use the estimates of the Probit equation for the reduced sample to construct $\hat{\bar{\eta}}^*_{ij} = \phi(\hat{z}^*_{ij})/\Phi(\hat{z}^*_{ij})$ and $\hat{w}^*_{ij}(\delta) = \ln\{\exp[\delta(\hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij})] - 1\}$ for all country pairs with positive trade flows.[31] The former controls for the sample selection bias whereas the latter controls for unobserved firm heterogeneity, that is, the effect of trade frictions and country characteristics on the proportion of exporters. We first report the results from a benchmark gravity equation without these controls in the second column of Table II and then report our NLS results in the third column. The standard errors are bootstrapped based on sampling (500 times) all available countries with replacement and using all the potential country pairs from that country sample. Both the nonlinear coefficient δ for $\hat{w}^*_{ij}$ and the linear

30. These 8 countries are Japan, Hong Kong, France, Germany, Italy, the Netherlands, the United Kingdom, and Sweden.

31. Recall that $\hat{z}^*_{ij} = \Phi^{-1}(\hat{\rho}_{ij})$. The characteristics of our data induces a complication associated with this transformation: Our sample includes a relatively small number of country pairs whose characteristics are such that their probability of trade $\hat{\rho}_{ij}$ is indistinguishable from 1. We therefore cannot infer any differences in the $\hat{z}^*_{ij}$'s among this subgroup of country pairs based on their probability of trade (whose binary realization is the only relevant datum we observe). Hence, we assign the same $\hat{z}^*_{ij}$ to those country pairs with an estimated $\hat{\rho}_{ij} > .9999999$, equivalent to an estimated $\hat{\rho}_{ij}$ at this cutoff. This censoring affects 4.3% of the 6,602 country pairs with positive trade flows from the Probit estimation on the reduced sample (12,198 country pairs).

TABLE II
BASELINE RESULTS

	1986 reduced sample					
		m_{ij}				
					Indicator variables	
Variables	(Probit) T_{ij}	Benchmark	NLS	Polynomial	50 bins	100 bins
Distance	−0.213**	−1.167**	−0.813**	−0.847**	−0.755**	−0.789**
	(0.016)	(0.040)	(0.049)	(0.052)	(0.070)	(0.088)
Land border	−0.087	0.627**	0.871**	0.845**	0.892**	0.863**
	(0.072)	(0.165)	(0.170)	(0.166)	(0.170)	(0.170)
Island	−0.173*	−0.553*	−0.203	−0.218	−0.161	−0.197
	(0.078)	(0.269)	(0.290)	(0.258)	(0.259)	(0.258)
Landlock	−0.053	−0.432*	−0.347*	−0.362+	−0.352+	−0.353+
	(0.050)	(0.189)	(0.175)	(0.187)	(0.187)	(0.187)
Legal	0.049**	0.535**	0.431**	0.434**	0.407**	0.418**
	(0.019)	(0.064)	(0.065)	(0.064)	(0.065)	(0.065)
Language	0.101**	0.147+	−0.030	−0.017	−0.061	−0.036
	(0.021)	(0.075)	(0.087)	(0.077)	(0.079)	(0.083)
Colonial ties	−0.009	0.909**	0.847**	0.848**	0.853**	0.838**
	(0.130)	(0.158)	(0.257)	(0.148)	(0.152)	(0.153)
Currency union	0.216**	1.534**	1.077**	1.150**	1.045**	1.107**
	(0.038)	(0.334)	(0.360)	(0.333)	(0.337)	(0.346)
FTA	0.343**	0.976**	0.124	0.241	−0.141	0.065
	(0.009)	(0.247)	(0.227)	(0.197)	(0.250)	(0.348)
Religion	0.141**	0.281*	0.120	0.139	0.073	0.100
	(0.034)	(0.120)	(0.136)	(0.120)	(0.124)	(0.128)
Regulation	−0.108**	−0.146				
costs	(0.036)	(0.100)				
R. costs (days	−0.061*	−0.216+				
& proc.)	(0.031)	(0.124)				
δ (from $\hat{\bar{w}}^*_{ij}$)			0.840**			
			(0.043)			
$\hat{\bar{\eta}}^*_{ij}$			0.240*	0.882**		
			(0.099)	(0.209)		
$\hat{\bar{z}}^*_{ij}$				3.261**		
				(0.540)		
$\hat{\bar{z}}^{*2}_{ij}$				−0.712**		
				(0.170)		
$\hat{\bar{z}}^{*3}_{ij}$				0.060**		
				(0.017)		
Observations	12,198	6,602	6,602	6,602	6,602	6,602
R^2	0.573	0.693		0.701	0.704	0.706

Notes. Exporter and importer fixed effects. Marginal effects at sample means and pseudo R^2 reported for Probit. Regulation costs are excluded variables in all second stage specifications. Bootstrapped standard errors for NLS; robust standard errors (clustering by country pair) elsewhere.
+Significant at 10%.
*Significant at 5%.
**Significant at 1%.

coefficient for $\hat{\bar{\eta}}^*_{ij}$ are precisely estimated. The remaining results for the linear coefficients clearly demonstrate the importance of an unmeasured heterogeneity bias in the estimated effects 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 destination. Consequently, the measures of the effects of trade frictions in the benchmark gravity equation are biased upward, as they confound the true effect of these frictions with their indirect effect on the proportion of exporting firms.[32] As highlighted in Table II, 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.[33] The effects of a currency union and colonial ties on firm or product level trade are also substantially reduced. The bias for the effect of FTAs is even more severe, as its coefficient drops by almost an order of magnitude and becomes insignificant. The measured effect of a common language is also strongly affected; it becomes insignificant (the benchmark coefficient is significant at the 5.2% level) and is precisely estimated around zero. Similarly for common religion: it becomes insignificant. This suggests that FTAs, a common language, and a common religion predominantly reduce the fixed costs of trade: they have a great influence on a firm's choice of export location, but not on its export volume once the exporting decision has been made.

We now progressively relax the parameterization assumptions that determined our functional forms. First, we relax the assumption governing the distribution of firm heterogeneity, and hence the form of the control function $\hat{w}^*_{ij}(\delta)$ for $\hat{\bar{z}}^*_{ij}$ in the trade flow equation (14). That is, we drop the Pareto assumption for $G(.)$ and revert to the general specification for V_{ij} in (5). Using (4) and (10), $v_{ij} \equiv \upsilon(z_{ij})$ is now an arbitrary (increasing) function of z_{ij}. We then directly control for $E[V_{ij} \mid ., T_{ij} = 1]$ using $\upsilon(\hat{\bar{z}}^*_{ij})$, which we approximate with a polynomial in $\hat{\bar{z}}^*_{ij}$. This replaces $\hat{w}^*_{ij} \equiv \ln\{\exp[\delta(\hat{\bar{z}}^*_{ij})] - 1\}$ in (14).[34] As the nonlinearity induced by $\hat{w}^*_{ij}$ is eliminated, we now estimate the second stage using OLS.

32. The effect of a land border is an exception because it negatively affects the probability of trade.

33. Several studies have documented that the effect of distance in gravity models is overstated because 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.

34. Recall that w_{ij} and v_{ij} differ only by a constant term.

In practice, we have found no noticeable changes from expanding $\upsilon(\hat{\bar{z}}^*_{ij})$ beyond a cubic polynomial. The results from this second stage estimation (the first stage Probit remains unchanged) are reported in the fourth column of Table II. These results are very similar to the NLS estimates.[35] In other words, the Pareto distribution does not appear to unduly constrain our baseline specification.

We further relax the joint normality assumption for the unobserved trade costs, and hence the Mills ratio functional form for the selection correction. This naturally precludes the separation of the effects of the latter from the firm heterogeneity effects. However, we can still jointly control for these effects with a flexible nonparametric functional form and thus obtain our key results for the intensive-margin contribution of the various trade barriers. The first stage estimation remains the same except that we now can use any cumulative distribution function instead of the normal distribution. We have experimented with the Logit and t-distributions with various low degrees of freedom and found that the resulting predicted probabilities $\hat{\rho}_{ij}$ are strikingly similar. For this reason, we no longer use the normality assumption to recover the $\hat{\bar{z}}^*_{ij}$ and $\hat{\bar{\eta}}^*_{ij}$. Instead, we work directly with the predicted probabilities $\hat{\rho}_{ij}$.

In order to approximate as flexibly as possible an arbitrary functional form of the $\hat{\rho}_{ij}$s, we use a large set of indicator variables. We partition the obtained $\hat{\rho}_{ij}$s into a number of bins with equal observations and assign an indicator variable to every bin. We then replace the $\hat{\bar{w}}^*_{ij}$ and $\hat{\bar{\eta}}^*_{ij}$ controls from the NLS estimation or the $\hat{\bar{z}}^*_{ij}$ and $\hat{\bar{\eta}}^*_{ij}$ controls from the polynomial estimation with this set of indicator variables. We report results with both 50 and 100 bins, to ensure a large degree of flexibility.[36] The results are in the last two columns of Table II. Here, we use the predicted probabilities from the baseline Probit, but these results are virtually unchanged when switching to a Logit or a t-distribution in the first stage. Evidently, all three estimation methods yield very similar results.

35. Here, we report the robust standard errors controlling for clustering at the country-pair level but do not correct for the generated regressors in the second stage. We experimented with bootstrapping the standard errors, as performed for the NLS specification, but this barely affected any of them. No coefficient significance test (at the 1%, 5%, or 10% level) was affected.

36. As with the polynomial approximation, this specification is now linear, and we thus use OLS.

VII. An Alternative Excluded Variable

Although the use of regulation cost variables has advantages, it also has a drawback: it substantially reduces the number of usable observations, as we explained in Section V (from 24,649 to 12,198 for the first stage, and from 11,156 to 6,602 for the second stage). For this reason it is desirable to find at least one other variable that satisfies the exclusion restrictions, which can be used for estimation with the full sample of countries. We argue in this section that our religion variable is suitable for this purpose.[37]

Once we have reliable excluded variables, such as our regulation cost variables, we can test whether any additional variable satisfies the exclusion restrictions. The key is for this variable to be correlated with the z_{ij}'s but not be correlated with the residual of the second stage equation that has been estimated with the reliable excluded variables (the reliable excluded variables are believed to satisfy the exclusion restrictions on theoretical grounds). In our case this means that the residuals from the trade flow equation should be uncorrelated with this variable. We argue that our common religion variable satisfies these requirements.

That the religion variable satisfies the first requirement is evident from the Probit equation, in which religion has a positive and significant effect on the probability of exporting (see Tables I and II). A simple test of the second requirement is provided in Table II. As is evident from the standard errors, one cannot reject the hypothesis that the coefficient on religion equals zero in each and every case. In other words, religion is not correlated with the second stage residuals.[38] To further enhance confidence in common religion as the excluded variable, we rerun all the second stage specifications from Table II, dropping the regulation cost variables and using religion as the excluded variable. The results of this estimation procedure applied to the reduced sample are reported in the left-hand-side panel of Table III. Evidently, they are very similar to the results in Table II.

37. We also experimented using the common language variable as the excluded variable. We obtained results almost identical to those using religion as the excluded variable.

38. We also performed a chi-squared test with one overidentification restriction (see Wooldridge [2002]) using all three excluded variables (the two regulations of entry costs variables and common religion). However, since the second stage residuals are no longer normally distributed after correcting for sample selection, this test is only asymptotically valid. Still, in all specifications, we cannot reject the hypothesis that all three variables are uncorrelated with the second stage residuals.

TABLE III
ALTERNATE EXCLUDED VARIABLES

	1986 reduced sample				1986 full sample			
			Indicator variables				Indicator variables	
Variables	NLS	Polynomial	50 bins	100 bins	NLS	Polynomial	50 bins	100 bins
Distance	−0.822**	−0.853**	−0.751**	−0.731**	−0.798**	−0.862**	−0.671**	−0.623**
	(0.048)	(0.051)	(0.069)	(0.089)	(0.039)	(0.041)	(0.059)	(0.076)
Land border	0.878**	0.855**	0.903**	0.907**	0.834**	0.786**	0.894**	0.924**
	(0.169)	(0.164)	(0.166)	(0.167)	(0.132)	(0.144)	(0.147)	(0.150)
Island	−0.204	−0.219	−0.171	−0.142	−0.169	−0.200+	−0.091	−0.074
	(0.291)	(0.258)	(0.265)	(0.266)	(0.120)	(0.118)	(0.119)	(0.121)
Landlock	−0.348*	−0.361+	−0.347+	−0.344+	−0.447**	−0.482**	−0.437*	−0.439*
	(0.176)	(0.188)	(0.190)	(0.192)	(0.172)	(0.186)	(0.186)	(0.186)
Legal	0.439**	0.442**	0.424**	0.418**	0.387**	0.385**	0.350**	0.345**
	(0.066)	(0.064)	(0.065)	(0.066)	(0.048)	(0.049)	(0.050)	(0.050)
Language	−0.018	−0.005	−0.060	−0.068	0.023	0.045	−0.044	−0.062
	(0.085)	(0.077)	(0.079)	(0.085)	(0.062)	(0.061)	(0.064)	(0.068)
Colonial ties	0.835**	0.839**	0.837**	0.830**	1.001**	1.038**	0.960**	0.929**
	(0.251)	(0.147)	(0.149)	(0.148)	(0.204)	(0.116)	(0.117)	(0.119)
Currency union	1.034**	1.102**	1.021**	0.984**	1.023**	1.106**	0.977**	0.960**
	(0.361)	(0.334)	(0.341)	(0.353)	(0.273)	(0.261)	(0.265)	(0.270)

TABLE III
(CONTINUED)

	1986 reduced sample				1986 full sample			
			Indicator variables				Indicator variables	
Variables	NLS	Polynomial	50 bins	100 bins	NLS	Polynomial	50 bins	100 bins
FTA	0.154	0.254	−0.161	−0.200	0.380*	0.457**	0.050	−0.091
	(0.225)	(0.199)	(0.250)	(0.337)	(0.182)	(0.162)	(0.165)	(0.210)
δ (from $\hat{\hat{w}}_{ij}^*$)	0.827**				0.871**			
	(0.043)				(0.028)			
$\hat{\hat{\eta}}_{ij}^*$	0.198*	0.823**			0.372**	1.131**		
	(0.099)	(0.211)			(0.069)	(0.138)		
$\hat{\hat{z}}_{ij}^*$		3.229**				3.602**		
		(0.538)				(0.386)		
$\hat{\hat{z}}_{ij}^{*2}$		−0.709**				−0.782**		
		(0.169)				(0.123)		
$\hat{\hat{z}}_{ij}^{*3}$s		0.061**				0.064**		
		(0.017)				(0.013)		
Observations	6,602	6,602	6,602	6,602	11,146	11,146	11,146	11,146
R^2		0.700	0.702	0.705		0.721	0.722	0.723

Notes. m_{ij} is dependent variable throughout. Exporter and importer fixed effects. Religion is excluded variable in all second stage specifications. Bootstrapped standard errors for NLS; robust standard errors (clustering by country pair) elsewhere.
$^+$Significant at 10%.
*Significant at 5%.
**Significant at 1%.

Once religion is accepted as a legitimate excluded variable, we can use it to estimate the model on the full sample of countries instead of the smaller sample that has been used so far, thereby roughly doubling the number of usable observations. The results of this estimation are reported in the right-hand-side panel of Table III for all three estimation methods (the benchmark gravity and Probit selection results were already reported in Table I). The magnitudes of these coefficients remain comparable despite the substantial increase in sample size.[39] Similar to the reduced sample estimates, we find that heterogeneity matters; higher trade volumes are driven both by direct effects of lower trade barriers and by greater proportions of exporters. As a result, estimated trade frictions in the benchmark gravity equation are biased upward, confounding the true effects with the indirect effects on the fraction of exporting firms.[40] As is evident from Table III, these biases are substantial; the coefficients on distance, currency union, colonial ties, language, and FTAs drop significantly.[41]

The substantial increase in country coverage allows us to study how these biases vary with the characteristics of the country pairs, which we explore in our counterfactual analysis in the following section. For now, we also take advantage of the inclusion of French exports in the full sample to compare our estimates for the extensive margin of French exports to the direct measure of the number of French exporting firms across destinations reported by Eaton, Kortum, and Kramarz (2004) for 1986. In our model, w_{ij} is an increasing function of the fraction of firms exporting from country j to country i. Our estimates of w_{ij} for $j = France$ should therefore be positively correlated with the number of French exporters to country i reported by Eaton, Kortum, and Kramarz (2004). We check this using our estimates for both $\hat{\bar{w}}^*_{ij}$ from the NLS specification and for $\hat{\upsilon}(\hat{\bar{z}}^*_{ij})$ from the polynomial approximation. These correlations are extremely high in both cases: 77% for $\hat{\bar{w}}^*_{ij}$ and 78% for $\hat{\upsilon}(\hat{\bar{z}}^*_{ij})$.[42]

39. The effects of FTAs are estimated to be significantly higher in the NLS and polynomial approximation specifications, though still substantially lower than in the benchmark estimates.

40. The effect of a land border is again an exception, because it negatively affects the probability of trade.

41. In the working paper version (Helpman, Melitz, and Rubinstein 2007), we also report results for the 1980s. They show that 1986 is not exceptional in terms of the full sample estimates. The coefficient for joint membership in the WTO drops substantially, but remains statistically and economically significant.

42. Because w_{ij} is a logged value, we compute the correlation using the logarithm of the number of exporting firms.

VIII. Additional Insights

We now use the full sample estimates from the previous section to examine several aspects of the results in further detail.

VIII.A. Decomposing the Biases

Our second stage estimate 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 equation, one controlling for unobserved heterogeneity only, the other controlling for selection only.

The results are reported in Table IV. The first two columns report the benchmark equation and our second stage NLS estimates from the full sample from Tables I and III. 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 only 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 $\hat{\bar{\eta}}^*_{ij}$).[43] 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 nonlinear estimation.

In the fourth column, we correct only for the selection bias (the standard two-stage Heckman (1979) 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, the remaining coefficients are very similar to those obtained in the benchmark specification of column (1).[44]

43. In this exercise we want to ensure a simple monotonic transformation of $\hat{z}^*_{ij}$, so we do not add any higher order terms.

44. This is consistent with the finding of Silva and Tenreyro (2006), whose application of a Poisson estimation method on a sample that consists of positive trade flows and a sample that includes zeros as well yields similar results.

TABLE IV
BIAS DECOMPOSITION

	1986 full sample			
Variables	Benchmark	NLS	Firm heterogeneity	Heckman selection
Distance	−1.176**	−0.798**	−0.769**	−1.214**
	(0.031)	(0.039)	(0.038)	(0.031)
Land border	0.458**	0.834**	0.855**	0.436**
	(0.147)	(0.132)	(0.142)	(0.149)
Island	−0.391**	−0.169	−0.164	−0.425**
	(0.121)	(0.120)	(0.118)	(0.120)
Landlock	−0.561**	−0.447**	−0.433*	−0.565**
	(0.188)	(0.172)	(0.187)	(0.187)
Legal	0.486**	0.387**	0.381**	0.488**
	(0.050)	(0.048)	(0.049)	(0.050)
Language	0.176**	0.023	0.023	0.223**
	(0.061)	(0.062)	(0.060)	(0.061)
Colonial ties	1.299**	1.001**	0.979**	1.311**
	(0.120)	(0.204)	(0.119)	(0.123)
Currency union	1.364**	1.023**	0.996**	1.391**
	(0.255)	(0.273)	(0.260)	(0.257)
FTA	0.759**	0.380*	0.314+	0.737**
	(0.222)	(0.182)	(0.168)	(0.235)
Religion	0.102			
	(0.096)			
δ (from $\hat{w}^*_{ij}$)		0.871**		
		(0.028)		
$\hat{\bar{\eta}}^*_{ij}$		0.372**		0.265**
		(0.069)		(0.070)
$\hat{\bar{z}}^*_{ij}$			0.892**	
			(0.051)	
Observations	11,146	11,146	11,146	11,146
R^2	0.709		0.716	0.710

Notes. m_{ij} is dependent variable throughout. Exporter and importer fixed effects. Religion is excluded variable in all second stage specifications. Bootstrapped standard errors for NLS; robust standard errors (clustering by country pair) elsewhere.
+Significant at 10%.
*Significant at 5%.
**Significant at 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 although aggregate country-pair shocks do have a significant effect on trade patterns, they only negligibly affect the

TABLE V
ASYMMETRIES

	1986 full sample				
		$m_{ij} - m_{ji}$			
Variable	$T_{ij} - T_{ji}$	NLS		Polynomial	
$\hat{\rho}_{ij} - \hat{\rho}_{ji}$	0.999**				
	(0.0169)				
$\hat{\bar{w}}^*_{ij} - \hat{\bar{w}}^*_{ji}$		1.187**	1.251**		
		(0.042)	(0.266)		
$\hat{v}(\hat{\bar{z}}^*_{ij}) - \hat{v}(\hat{\bar{z}}^*_{ji})$				1.012**	0.703**
				(0.035)	(0.143)
Country fixed effects	No	No	Yes	No	Yes
Observations	12,246	4,517	4,517	4,517	4,517
R^2	0.219	0.153	0.324	0.157	0.325

**Significant at 1%.

responsiveness of trade volumes to observed trade barriers.[45] 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, because 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.

VIII.B. 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. 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 V. 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).[46] Note that the regressand, $T_{ij} - T_{ji}$, takes on the values -1, 0, and 1, depending on the direction of

45. This finding also highlights the important information conveyed by the nontrading 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.

46. Recall that T_{ij} is the indicator variable for positive trade from j to i.

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 V 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 22% of the variation in the direction of trade.[47] 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 V focuses on bilateral trade flow asymmetries between country pairs trading in both directions. It shows the results of the OLS regression of net bilateral trade $m_{ij} - m_{ji}$ (the percentage difference between exports and imports) on alternatively $\hat{\bar{w}}^*_{ij} - \hat{\bar{w}}^*_{ji}$ (for the NLS specification) or $\hat{\nu}(\hat{\bar{z}}^*_{ij}) - \hat{\nu}(\hat{\bar{z}}^*_{ji})$ (for the polynomial approximation). That regressor captures differences in the proportion 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 (using either specification) is a strong predictor of net bilateral trade. On its own, it explains 15%–16% of the variance in net trade, and along with the country fixed effects it explains 32%–33% of that variance.[48]

VIII.C. Counterfactuals

We have just shown how the fitted values for ρ_{ij} and w^*_{ij} can explain a large portion of the variation in the direction of trade and in its extensive margin. We next show how to use these fitted values to make predictions about the response of trade to changes in trade costs. For every change in the bilateral trade costs d_{ij}, our model predicts the new pattern of trade, that is, who trades with whom, and in which direction. In addition, for country pairs that trade with each other, the model predicts the resulting changes in the composition of trade flows between the extensive and intensive margins. These counterfactual predictions can be measured, and

47. This understates the variable's explanatory power, because it is continuous and it predicts a discrete variable.

48. Anderson and van Wincoop (2003) account for asymmetric bilateral trade flows with asymmetric variable bilateral trade costs. In a more general model one can have both asymmetric bilateral trade costs and asymmetric extensive margins of trade.

we illustrate their quantitative impact for a reduction in trade costs associated with distance.

In response to a drop in distance-related trade costs, some countries start trading with one another. Trade rises for country pairs that traded before the drop in trade costs, and we report how the increase in trade can be decomposed into the intensive and extensive margin. We find that the extensive margin is especially important in shaping the response of trade flows across country pairs because it generates substantial heterogeneity across country pairs. This richness contrasts sharply with the uniform response implied by the baseline gravity model, which does not account for the extensive margin of trade (nor does it account for the creation of new trading relationships).

The computation of these responses involves some technical details that are explained in Appendix III. Here we report the results of a particular counterfactual experiment involving a decrease in the trade costs associated with distance. That is, we investigate the response of trade for any given country pair, assuming that the distance between these two country pairs decreases by a given percentage. We first focus on country pairs observed trading and focus on the elasticity of the overall trade response for each country pair: $|\hat{m}'_{ij} - m_{ij}|/|d'_{ij} - d_{ij}|$, where d_{ij} now specifically references the bilateral distance variable.[49] Since our model predicts different response elasticities with the magnitude of the trade decrease, we report these elasticities for the case of a 10% distance decrease ($d'_{ij} - d_{ij} = \log 0.9$), although any percentage decrease under 20% would yield virtually identical results.[50]

As was previously mentioned, the elasticities vary widely across different country pairs. In order to highlight how these elasticities vary along one important country pair dimension—country income—we report summary statistics across three groups of country pairs: North–North, North–South, and South–South, sorted by GDP per capita.[51] These statistics appear in Table IV for

49. To avoid any confusion when discussing "larger" versus "smaller" elasticities, we express the elasticities in absolute value. Naturally, for the case of trade costs, these elasticities are all negative.

50. Larger decreases in trade costs would produce larger elasticities but with similar qualitative patterns across country pairs.

51. We use 1986 U.S. \$15,000 as the cutoff GDP per capita between North and South. The former group is composed of nineteen countries: Australia, Austria, Belgium-Luxemburg, Canada, Denmark, Finland, France, Germany, Hong Kong, Iceland, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, and the United States.

TABLE VI
SUMMARY STATISTICS OF THE TRADE ELASTICITY RESPONSE ACROSS COUNTRY PAIRS

Country pairs group	Number of country pairs	Nonlinear least squares				Polynomial aproximation			
		Mean	S. D.	Min	Max	Mean	S. D.	Min	Max
NN	342	1.292	0.034	1.283	1.642	1.290	0.107	1.141	2.222
NS	4,626	1.404	0.152	1.283	2.949	1.526	0.386	1.141	2.895
SS	6,178	1.698	0.303	1.283	3.777	2.130	0.443	1.141	2.995
Overall	11,146	1.563	0.289	1.283	3.777	1.854	0.517	1.141	2.995

both our NLS and polynomial approximation specifications. Importantly, we emphasize that all the heterogeneity in the elasticity response is driven by the extensive margin because the elasticity response at the intensive margin is fixed at 0.799 (NLS) and 0.862 (polynomial approximation). Because this extensive margin response depends fundamentally on the functional forms for $\hat{\bar{w}}_{ij}^{*\prime}$ or $\hat{v}(\hat{\bar{z}}_{ij}^{*\prime})$ in terms of $\hat{\bar{z}}_{ij}^{*\prime}$, we report the elasticities for both cases. Although the shape of the functional form for $\hat{\bar{w}}_{ij}^{*\prime}$ is in part determined by our theoretical modeling assumptions (see (13)), the shape of the $\hat{v}(\hat{\bar{z}}_{ij}^{*\prime})$ is entirely data-driven. Reassuringly, both functions have very similar shapes over the range of $\hat{\bar{z}}_{ij}^{*\prime}$, and the counterfactual distributions of the response elasticity are similar.

The heterogeneous trade responses reported in Table IV show that these elasticities vary between 1.283 and 3.777 for the NLS estimates and between 1.141 and 2.995 for the semiparametric estimates, large variations indeed.[52] We visually depict these distributions across country pairs group in Figure III. The charts clearly document how the range and distribution of elasticities vary with country income: the elasticities are highest for South–South trade, lower for North–South trade, and lowest for North–North trade. Thus, when trade costs related to distance fall, our estimates predict that the response of the extensive margin of trade is larger for less developed countries.[53]

52. Of course, departing from the log-linear specification for distance would yield different elasticities for different changes in trade costs related to distance. Our main point is that, *given* a log-linear specification for distance in both stages, our model still predicts substantial differences in the response elasticity, driven by the characteristics of the country pairs that jointly determine the extensive margin of trade.

53. In Helpman, Melitz, and Rubinstein (2007) we also report how many of the countries that do not trade initially and which pairs start trading when the trade costs fall. These results suggest that large changes in trade-related costs are needed to induce nontrading country pairs, involving at least one Southern

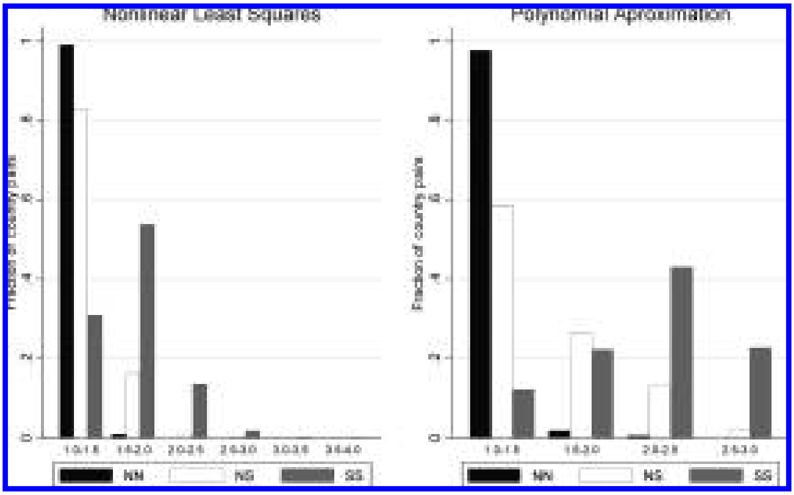

FIGURE III
The Distribution of the Distance Elasticity across Country Pairs

IX. CONCLUDING COMMENTS

Empirical explanations of international trade flows have a long tradition. The gravity equation with various measures of trade resistance plays a key role in this literature. Indeed, estimates of the impact of trade resistance measures provide important information about the roles played by common currencies, free trade areas, membership in the WTO, and other features of trading countries. For this reason, it is important to obtain reliable estimates of the effects of those trade barriers/enhancers on international trade flows.

We develop in this paper an estimation procedure that corrects certain biases embodied in the standard gravity estimation of trade flows. Our approach is driven by theoretical as well as econometric considerations. On the theoretical side, we develop a simple model that is capable of explaining empirical phenomena, such as zero trade flows between certain pairs of countries and larger numbers of exporters to larger destination markets. We then derive from this theory a two-equation system that can be estimated with standard data sets. Importantly, this system

country, to trade. Moreover, they are in line with the evidence presented in Figures I and II that almost all of the increase in world trade flows in the last thirty years has occurred among countries with trading relationships in 1970.

enables one to decompose the impact on trade volumes of all trade resistance measures into their intensive (trade volume per firm) and extensive (number of exporting firms) margin components. We show how to obtain estimates of this decomposition without having firm-level data, but rather country-level data that are normally used to estimate trade flows. The ability to obtain such a decomposition is important because, in practice, a substantial proportion of trade adjustment takes place at the extensive margin, and it is not possible to obtain consistent firm-level data with export destinations for a large number of countries (which would be needed for a direct estimation of the extensive margin component).

Our empirical analysis has been confined to country-level trade flows, where about half of the observations are zeros. Naturally, the problem of zeros is even more severe at the industry level. That is, in data sets of sectoral trade flows the fraction of zeros is much larger. Importantly, our estimation method can be implemented on such data sets as well. Manova (2006) is an example of this, highlighting the important contribution of the extensive margin of trade in explaining the impact of financial frictions on sectoral trade flows.

A variety of robustness checks show that the resulting estimates are not sensitive to the estimation method (parametric, semiparametric, or nonparametric) nor to the excluded variables from the first stage of our two-stage estimation procedure. Moreover, these estimates suggest that the biases embodied in the commonly used approach are substantial and that they are mostly due to the omitted control for the extensive margin of trade. Especially important is our finding that the bias not only is large, but also substantially varies across country pairs with different characteristics. In particular, the response of the trade flow between one pair of countries to a given reduction in distance-related trade frictions (such as transport costs) can be as much as three times larger than the response of the trade flow between another pair of countries. We show how these large variations across country pairs in the response to a given trade friction reduction are driven by variation in the extensive margin responses.

Finally, we note that our estimation procedure is easy to implement. In addition, it is flexible, because it allows the use of parametric, semiparametric, and nonparametric specifications. In other words, the procedure provides the researcher with flexibility and convenience in individual applications.

APPENDIX I

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 twelve such country titles and three proper countries for which data other than trade flows were missing. This left usable data for bilateral trade flows among 158 countries. The list of these countries is provided in Table A.1.

For the 158 countries we constructed a matrix of trade flows, measured in constant 2000 U.S. dollars, using the U.S. CPI. This matrix represents $158 \times 157 = 24{,}806$ observations, 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 two standard sources: the Penn World Tables 6.1 and the World Bank's World Development Indicators.

We used the CIA's World Factbook to construct a number of variables, which can be classified as follows:[54]

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.

We also used data from Rose (2000) and Glick and Rose (2002), as presented on Andrew Rose's Web site, to identify whether a country pair belongs to the same currency union or the same FTA. And we used data from Rose (2004) to identify whether a country is a member of the GATT/WTO.

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 (see below).

The construction of the regulation costs of firm entry are described in the main text. As previously mentioned, cost data on the number of days, number of legal procedures, and relative cost (as percent of GDP per capita) are reported in Djankov et al. (2002).

54. See http://www.cia.gov/cia/publications/factbook/docs/profileguide.html.

TABLE A.1
LIST OF COUNTRIES

Afghanistan	Ecuador	Korea DPR	Romania
Albania	Egypt	Korea Rep.	Rwanda
Algeria	El Salvador	Kuwait	Saudi Arabia
Angola	Eq. Guinea	Laos	Senegal
Argentina	Ethiopia	Lebanon	Seychelles
Australia	Fiji	Liberia	Sierra Leone
Austria	Finland	Libya	Singapore
Bahamas	Fm. Czechoslovakia	Madagascar	Solomon Islds.
Bahrain	Fm. USSR	Malawi	Somalia
Bangladesh	Fm. Yugoslavia	Malaysia	South Africa
Barbados	France	Maldives	Spain
Belgium–Lux.	French Guiana	Mali	Sri Lanka
Belize	Gabon	Malta	St Kitts Nevis
Benin	Gambia	Mauritania	Sudan
Bermuda	Germany	Mauritius	Surinam
Bhutan	Ghana	Mexico	Sweden
Bolivia	Greece	Mongolia	Switzerland
Brazil	Greenland	Morocco	Syria
Brunei	Guadeloupe	Mozambique	Taiwan
Bulgaria	Guatemala	Myanmar	Thailand
Burkina Faso	Guinea	Nepal	Togo
Burundi	Guinea-Bissau	Netherlands	Trinidad-Tobago
Cambodia	Guyana	Neth. Antilles	Tunisia
Cameroon	Haiti	New Caledonia	Turkey
Canada	Honduras	New Zealand	Turks and Caicos
Cayman Islds	Hong Kong	Nicaragua	Uganda
Central Africa	Hungary	Niger	United Kingdom
Chad	Iceland	Nigeria	United Arab Em.
Chile	India	Norway	United Rep. Tanzania
China	Indonesia	Oman	
Colombia	Iran	Pakistan	United States
Comoros Islds.	Iraq	Panama	Uruguay
Congo	Ireland	Papua N.Guinea	Venezuela
Costa Rica	Israel	Paraguay	Vietnam
Cote D'Ivoire	Italy	Peru	Western Sahara
Cuba	Jamaica	Philippines	Yemen
Cyprus	Japan	Poland	Zaire
Denmark	Jordan	Portugal	Zambia
Djibouti	Kenya	Qatar	Zimbabwe
Dominican Rep.	Kiribati	Reunion	

Main Variables:

1. *Distance*: the distance (in km) between importer's i and exporter's j capitals (in logs).

2. *Common border*: a binary variable that equals one if importer i and exporter j are neighbors that meet a common physical boundary, and zero otherwise.
3. *Island*: a binary variable that equals one if both importer i and exporter j are islands, and zero otherwise.
4. *Landlocked*: a binary variable that equals one if both exporting country j and importing country i have no coastline or direct access to sea, and zero otherwise.
5. *Colonial ties*: a binary variable that equals one if importing country i ever colonized exporting country j or vice versa, and zero otherwise.
6. *Currency union*: a binary variable that equals one if importing country i and exporting country j use the same currency or if within the country pair money was interchangeable at a 1:1 exchange rate for an extended period of time (see Rose [2000, 2004] and Glick and Rose [2002]), and zero otherwise.
7. *Legal system*: a binary variable that equals one if the importing country i and exporting country j share the same legal origin, and zero otherwise.
8. *Religion*: (% Protestants in country i $\cdot$ % Protestants in country j) + (% Catholics in country i $\cdot$ % Catholics in country j) + (% Muslims in country i $\cdot$ % Muslims in country j).
9. *FTA*: a binary variable that equals one if exporting country j and importing country i belong to a common regional trade agreement, and zero otherwise.
10. *WTO*: a vector of two dummy variables: the first binary variable equals one if both exporting country j and importing country i *do not* belong to the GATT/WTO, and zero otherwise; the second binary variable equals one if both countries belong to the GATT/WTO, and zero otherwise.
11. *Entry costs*: a binary indicator that equals one if the sum of the number of days and procedures to form a business is above the median for both the importing country i and exporting country j, or if the relative cost (as percent of GDP per capita) of forming a business is above the median in the exporting country j and the importing country i, and zero otherwise.

APPENDIX II

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, the value of 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

$$Y_j = \left(\frac{c_j}{\alpha}\right)^{1-\varepsilon} N_j \sum_h \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} Y_h V_{hj}. \tag{15}$$

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

$$M_{ij} = \frac{Y_i Y_j}{Y} \frac{\left(\frac{\tau_{ij}}{P_i}\right)^{1-\varepsilon} V_{ij}}{\sum_{h=1}^{J} \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} V_{hj} s_h}, \tag{16}$$

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 (16) 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; 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.[55] 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 $0 \le a \le a_H$. 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.[56]

Using the decomposability property and symmetry requirements $\tau_{ij} = \tau_{ji}$ and $\varphi_{ij} = \varphi_{ji}$, we obtain[57]

$$\frac{M_{ij}}{Y} = s_i s_j \left(\frac{\tau_{ij}\varphi_{ij}}{Q_i Q_j} \right)^{1-\varepsilon}, \tag{17}$$

where the values of Q_j are solved from

$$Q_j^{1-\varepsilon} = \sum_h \left(\frac{\tau_{jh}\varphi_{jh}}{Q_h} \right)^{1-\varepsilon} s_h. \tag{18}$$

This is essentially the Anderson and van Wincoop (2003) system. Evidently, the solution of the Q_js depends only on income shares

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

56. 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 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}.

57. Decomposability allows us to rewrite (16) as

$$M_{ij} = \frac{Y_i Y_j}{Y} \left(\frac{\tau_{ij}\varphi_{ij}}{Q_i \hat{Q}_j} \right)^{1-\varepsilon}, \tag{20}$$

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 the product $Q_i Q_j$ proportionately 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, that is, $V_{ij} = 0$, then one of the φs ($\varphi_{IM,i}$, $\varphi_{EX,j}$, or φ_{ij}) must be infinite, because $\varepsilon > 1$. In other words, some trade costs, either at the country or bilateral level, must be infinite in order to explain zero trade flows. Our framework, which does not rely on this decomposability assumption, is much more general, as it can explain the prevalent zero trade flows based on finite trade costs (which can then be estimated). Furthermore, the gravity specification (17) based on the decomposability assumption cannot explain the asymmetries in bilateral trade flows (which must then stem from country fixed effects). In the case of zero bilateral trade in only one direction, this would impose either that the importer does not import from any other country or that the exporter does not export to any other

---

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

$$\hat{Q}_j^{1-\varepsilon} = \sum_h \left(\frac{\tau_{hj}\varphi_{hj}}{Q_h}\right)^{1-\varepsilon} s_h \,. \tag{21}$$

In addition, (7) and (15) imply

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

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

Therefore,

$$Q_j^{1-\varepsilon} = \sum_h \left(\frac{\tau_{jh}\varphi_{jh}}{\hat{Q}_h}\right)^{1-\varepsilon} s_h \,. \tag{22}$$

Equations (21) and (22) 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, (20) and (21) yield the equations in the text.

country. This is clearly inconsistent 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. In the case of positive trade flows in both directions, (17) imposes that all bilateral trade asymmetries stem from the country fixed-effects. This is also inconsistent with the observed pattern of trade, as documented in the second panel of Table V. Furthermore, that table documents that those asymmetries are highly correlated with the asymmetric pattern of zero trade flows (which would be inconsistent with (17)). Indeed, this is the main logic behind our more general theoretical model and empirical implementation: the decision to export to a foreign country is *not* independent of the volume of exports, and thus that the pattern of trading partners and trading volumes must jointly be analyzed. For these reasons we use the less restrictive equations (4)–(7) for estimation purposes.

APPENDIX III

We explain in this Appendix the computation of the counterfactuals in Section VIII. To this end, consider an observed change in the bilateral trade costs from d_{ij} to d'_{ij}.[58] The new predicted estimates of the probability of trade $\hat{\rho}'_{ij}$ and $\hat{z}^{*\prime}_{ij} = \Phi^{-1}(\hat{\rho}'_{ij})$ are obtained in a straightforward way from the first stage estimated Probit equation by replacing d_{ij} with d'_{ij}. We next need to obtain a consistent estimate of $z^{*\prime}_{ij}$ conditional on the observed trade status of j and i (trade or no trade) when trade costs are d_{ij}, given that we do not observe the trade status under the new trade costs d'_{ij}. This will replace $\hat{\bar{z}}^{*}_{ij}$ in our equations. Originally we were only concerned with computing $\hat{\bar{z}}^{*}_{ij}$ for country pairs with active trade, that is, with $T_{ij} = 1$. But now we also need to consider country pairs that do not trade under costs d_{ij} but might trade under costs d'_{ij}. For this reason we need to examine two cases.

III. A. Country Pairs Observed Trading

First, we note that the unobserved trade costs η^{*}_{ij} are not affected by the change in trade costs d_{ij}.[59] If we knew whether a

58. As in our previous derivations, d_{ij} can represent any given observable variable trade cost.

59. That is, we seek a ceteris paribus counterfactual prediction for a direct change in d_{ij}.

country pair traded under d'_{ij}, say T'_{ij}, then we could construct a new estimate for η^*_{ij}, say $\eta^{*\prime}_{ij}$, conditional on both T_{ij} and T'_{ij}. Absent this additional information, our best estimate for η^*_{ij} is conditional on T_{ij} and is still given by $\hat{\bar{\eta}}^*_{ij} = E[\eta^*_{ij} \mid ., T_{ij} = 1] = \phi(\hat{z}^*_{ij})/\Phi(\hat{z}^*_{ij})$. Thus, when $T_{ij} = 1$, our best estimate for $\hat{z}^{*\prime}_{ij}$ is given by

$$\hat{\bar{z}}^{*\prime}_{ij} = E\left[z^{*\prime}_{ij} \mid ., T_{ij} = 1\right] = \hat{z}^{*\prime}_{ij} + \phi\left(\hat{z}^*_{ij}\right)/\Phi\left(\hat{z}^*_{ij}\right).$$

Again, note that the new distance cost d'_{ij} is used to compute the new $\hat{z}^{*\prime}_{ij}$ but not the bias correction for η^*_{ij}. If $\hat{\bar{z}}^{*\prime}_{ij} < 0$, then we predict that j no longer exports to i. Because $\hat{\bar{z}}^*_{ij} > 0$, this can only happen when $d'_{ij} > d_{ij}$ (a scenario we will not explicitly consider). If $\hat{\bar{z}}^{*\prime}_{ij} > 0$, then we predict that the country pair continues to trade (this must be the case when $d'_{ij} < d_{ij}$). This new value of $\hat{\bar{z}}^{*\prime}_{ij}$ can then be used in conjunction with the second stage estimates to predict the response of trade flows at the extensive margin. In the case of the NLS estimation, this is $\hat{\bar{w}}^{*\prime}_{ij} = \ln\{\exp[\delta(\hat{\bar{z}}^{*\prime}_{ij})] - 1\}$ (and $\hat{\nu}(\hat{\bar{z}}^{*\prime}_{ij})$ for the polynomial approximation). The overall predicted trade response $\hat{m}'_{ij}$ is given by the fitted value from the estimated second stage equation (14) using the new values for $\hat{\bar{z}}^{*\prime}_{ij}$ and d'_{ij}:

$$\hat{m}'_{ij} = \hat{\beta}_0 + \hat{\lambda}_j + \hat{\chi}_i + \hat{\gamma} d'_{ij} + \hat{\bar{w}}^{*\prime}_{ij} + \hat{\beta}_{u\eta}\hat{\bar{\eta}}^*_{ij}. \tag{19}$$

In the case of the polynomial approximation, $\hat{\beta}_0 + \hat{\bar{w}}^{*\prime}_{ij}$ is replaced by $\hat{\nu}(\hat{\bar{z}}^{*\prime}_{ij})$.

III. B. Country Pairs Not Observed Trading

We now show how our model can be used to determine which nontrading country pairs are predicted to start trading under costs d'_{ij}, and the associated new predicted trade flow. The first stage yields a predicted $\hat{\rho}'_{ij}$ and $\hat{z}^{*\prime}_{ij}$ for all country pairs under d'_{ij}, including the nontrading country pairs. We now need to obtain a consistent estimate for $z^{*\prime}_{ij}$ for these country pairs, conditional on $T_{ij} = 0$. We start by expanding the definition for $\bar{\eta}^*_{ij}$ to include the country pairs that do not trade: $\bar{\eta}^*_{ij} = E[\eta^*_{ij} \mid ., T_{ij}]$. (This was previously defined only when $T_{ij} = 1$.) When $T_{ij} = 0$, this is given by

$$\bar{\eta}^*_{ij} = E\left[\eta^*_{ij} \mid ., T_{ij} = 0\right] = E\left[\eta^*_{ij} \mid ., \eta^*_{ij} < -z^*_{ij}\right] = \frac{-\phi(z^*_{ij})}{1 - \Phi(z^*_{ij})},$$

because η^*_{ij} is distributed standard normal. Hence, $\hat{\bar{\eta}}^*_{ij}$, our consistent estimate for $E[\eta^*_{ij} \mid ., T_{ij}]$, is constructed as

$$\hat{\bar{\eta}}^*_{ij} = \begin{cases} \dfrac{-\phi\left(\hat{z}^*_{ij}\right)}{1-\Phi\left(\hat{z}^*_{ij}\right)} & \text{if } T_{ij} = 0, \\ \dfrac{\phi\left(\hat{z}^*_{ij}\right)}{\Phi\left(\hat{z}^*_{ij}\right)} & \text{if } T_{ij} = 1. \end{cases}$$

Using this new expanded definition for $\hat{\bar{\eta}}^*_{ij}$, our previous definition for $\hat{\bar{z}}^*_{ij} = \hat{z}^*_{ij} + \hat{\bar{\eta}}^*_{ij}$ now provides a consistent estimate for $E[z^*_{ij} \mid T_{ij}]$, which now includes the case for country pairs with $T_{ij} = 0$. Note that, by construction, $\hat{\bar{z}}^*_{ij}$ must be negative whenever $T_{ij} = 0$ (recall that $\hat{\bar{z}}^*_{ij} > 0$ whenever $T_{ij} = 1$).

When trade costs change to d'_{ij}, we obtain a new $\hat{\bar{z}}^{*\prime}_{ij}$ for country pairs with $T_{ij} = 0$ in a way similar to that for $T_{ij} = 1$: $\hat{\bar{z}}^{*\prime}_{ij} = \hat{z}^{*\prime}_{ij} + \hat{\bar{\eta}}^*_{ij}$, where we do not adjust $\hat{\bar{\eta}}^*_{ij}$ for the new value of the trade costs.[60] Whenever $\hat{\bar{z}}^*_{ij} > 0$, our model predicts that j exports to i under the trade costs d'_{ij}. For these country pairs, the new predicted trade flow $\hat{m}'_{ij}$ can be predicted in a similar way to that for all the other trading country pairs using (19) along with the newly constructed $\hat{\bar{z}}^*_{ij}$.

HARVARD UNIVERSITY AND CIFAR
PRINCETON UNIVERSITY, CENTRE FOR ECONOMIC POLICY RESEARCH, AND NATIONAL BUREAU OF ECONOMIC RESEARCH
BROWN UNIVERSITY

REFERENCES

Anderson, James A., "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69 (1979), 106–116.
Anderson, James E., and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93 (2003), 170–192.
——,"Trade Costs," *Journal of Economic Literature*, 42 (2004), 691–751.
Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 93 (2003), 1268–1290.
Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott, "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. that Trade Goods," NBER Working Paper 11404, 2005.
Davis, Donald R., and David E. Weinstein, "Market Access, Economic Geography and Comparative Advantage: An Empirical Test," *Journal of International Economics*, 59 (2003), 1–23.

60. As before, we do not observe a new T'_{ij} under d'_{ij}.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, "The Regulation of Entry," *Quarterly Journal of Economics,* 117 (2002), 1–37.
Eaton, Jonathan, and Samuel S. Kortum, "Technology, Geography, and Trade," *Econometrica*, 70 (2002), 1741–1779.
Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "Dissecting Trade: Firms, Industries, and Export Destination," *American Economic Review* (Papers and Proceedings), 94 (2004), 150–154.
Evans, Carolyn L., "The Economic Significance of National Border Effects," *American Economic Review*, 93 (2003), 1291–1312.
Evenett, Simon J., and Anthony J. Venables, "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows," mimeo, London School of Economics, 2002.
Feenstra, Robert C., "Border Effects and the Gravity Equation: Consistent Methods for Estimation," *Scottish Journal of Political Economy*, 49 (2002), 491–506.
——, *Advanced International Trade* (Princeton, NJ: Princeton University Press, 2003).
Felbermayr, Gabriel J., and Wilhelm Kohler, "Exploring the Intensive and Extensive Margins of World Trade," *Review of World Economics*, 142 (2006), 642–674.
Frankel, Jeffrey A., *Regional Trading Blocks in the World Economic System* (Washington, DC: Institute of International Economics, 1997).
Glick, Reuven, and Andrew K. Rose, "Does a Currency Union Affect Trade? The Time Series Evidence," *European Economic Review*, 46 (2002), 1125–1151.
Haveman, Jon, and David Hummels, "Alternative Hypotheses and the Volume of Trade: The Gravity Equation and the Extent of Specialization," *Canadian Journal of Economics*, 37 (2004), 199–218.
Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1979), 153–161.
Helpman, Elhanan, "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries," *Journal of the Japanese and International Economics*, 1 (1987), 62–81.
Helpman, Elhanan, and Paul R. Krugman, *Market Structure and Foreign Trade* (Cambridge, MA: The MIT Press, 1985).
Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," NBER Working Paper No. 12927, 2007.
Hummels, David, and Peter J. Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review*, 95 (2005), 704–719.
Kehoe, Timothy J., and Kim J. Ruhl, "How Important Is the New Goods Margin in International Trade?" mimeo, University of Minnesota, 2002.
Manova, Kalina, "Credit Constraints, Heterogeneous Firms and International Trade," mimeo, Harvard University, 2006.
McCallum, John, "National Borders Matter: Canada–U.S. Regional Trade Patterns," *American Economic Review*, 85 (1995), 615–623.
Melitz, Marc J., "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71 (2003), 1695–1725.
Rose, Andrew K., "One Money One Market: Estimating the Effect of Common Currencies on Trade," *Economic Policy*, 15 (2000), 7–46.
——, "Do We Really Know That the WTO Increases Trade?" *American Economic Review*, 94 (2004), 98–114.
Silva, J. M. C. Santos, and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 88 (2006), 641–658.
Subramanian, Arvind, and Shang-Jin Wei, "The WTO Promotes Trade, Strongly But Unevenly," *Journal of International Economics*, 72 (2007), 151–175.
Tenreyro, Silvana, and Robert Barro, "Economic Effects of Currency Unions," National Bureau of Economic Research, Working Paper 9435, 2003.
Tinbergen, Jan, *Shaping the World Economy* (New York: The Twentieth Century Fund, 1962).
Wei, Shang-Jin, "Intranational Versus International Trade: How Stubborn are Nations in Global Integration?" NBER, Working Paper No. 5531, 1996.
Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: The MIT Press, 2002).