Geographical Diversification of Developing Country Exports

Shepherd, Ben
Princeton University, GEM, Sciences Po

24. October 2008

Geographical Diversification of Developing Country Exports

Ben Shepherd*

Princeton University & GEM (Sciences Po)

This Version Dated: October 24, 2008

Abstract

This paper shows that export costs, tariffs, and international transport costs are all important determinants of geographical export diversification in a sample of 123 developing countries. A 10% reduction in any one of these factors produces a 5%-6% increase in the number of foreign markets entered. Moreover, there is evidence that these impacts differ significantly across countries and sectors: geographical export diversification is more sensitive to export costs and transport costs in more differentiated sectors, and to export costs in lower income countries. These results are generally robust to alternative specifications, and instrumental variables estimation.

JEL codes: F13; F15; O24.

Keywords: International trade; Trade policy; Trade and development; Extensive margin; Economic geography.

*Postdoctoral Research Associate, Niehaus Center for Globalization and Governance, Princeton University; and Research Associate, GEM, Sciences Po. Comments to bshepher@princeton.edu; bashepherd@gmail.com. I am grateful to the following for helpful input on previous drafts: Jens Arnold, Chad Bown, Matt Cole, Allen Dennis, Felix Eschenbach, Ana Fernandes, Bernard Hoekman, Leonardo Iacovone, Patrick Messerlin, and participants at the Fall 2008 Midwest International Economics Meetings.

1 Introduction

Developing country trade growth can take place in four dimensions: more trade in goods that existing trading partners already exchange (the intensive margin); introduction of new product varieties (the new products margin); an increase in the unit values of traded goods (the quality margin); and creation of trading relationships between new partners (the new markets margin). While there is a vast literature on the determinants of intensive margin trade growth (e.g., Anderson and Van Wincoop, 2003), and an emerging body of work on the new products margin (e.g., Hummels and Klenow, 2005; Broda and Weinstein, 2006) and the quality margin (Schott, 2004; Baldwin and Harrigan, 2007), there is almost no empirical work on the new markets margin. Yet recent findings suggests that geographical export diversification–or growth at the new markets margin–can be an important mechanism through which developing countries can become more integrated in the world trading system.[1] For example, Evenett and Venables (2002) report that around 1/3 of developing country export growth over the period 1970-1997 was due to the export of "old" goods to new markets. Using a different dataset and methodology, Brenton and Newfarmer (2007) suggest that the proportion was around 18% for the period 1995-2004. Although Besedes and Prusa (2007) argue that intensive margin growth may actually be more important than the extensive margin in a dynamic sense, Cadot et al. (2007) suggest that the relative importance of the intensive and extensive margins depends on the exporting country's income level: the extensive margin is generally more important for poorer countries. Finally, Amurgo-Pacheco and Pierola (2008) find that in terms of the extensive margin itself, new markets growth dominates new products growth in poorer countries.

This paper aims to fill the void that currently exists in relation to the determinants of trade growth at the new markets margin by examining the impact of three sets of factors: market size and development level in the exporting country; international trade costs (distance, tariffs); and export costs (border formalities, customs) in the exporting country. In line with the broader literature on the determinants of trade growth, I find evidence that the first set of factors impacts the new markets margin positively, but the remaining

[1] An additional welfare argument in favor of a certain level of product and market diversification could also be that risk averse governments seek to take advantage of the covariance of demand shocks to obtain a minimally volatile export portfolio for a given level of return. However, development of this kind of framework is outside the scope of the present paper. See Brainard and Cooper (1968).

factors have a negative impact. These results are highly robust to alternative specifications, estimation via instrumental variable techniques, and (geographical) first differencing. In the case of export costs, there is evidence that the magnitude of this negative impact is stronger in poorer countries. Moreover, both export costs and transport costs have stronger impacts in more differentiated sectors. In policy terms, these results are therefore of particular relevance to lower income countries engaged in industrialization, i.e., a shift towards differentiated manufactured goods, and away from homogeneous agricultural products.

What is the economic intuition behind these results? Firm heterogeneity and market-specific trade costs provide a powerful explanation for why countries export goods to some countries but not others. Only a relatively small proportion of firms in an economy export, and the set of foreign markets they enter is based on the costs of doing so. Only the most productive firms can enter the most costly (least accessible) foreign markets. The existence of a bilateral trading relationship at the country level therefore depends on whether or not there is at least one firm with sufficiently high productivity (low marginal cost) to export profitably to the foreign market. Factors that shift the equilibrium cost cutoff for a given country pair upwards can thereby increase the probability that bilateral trade is observed between that country pair. This is the process of trade growth at the new markets margin that is central to this paper. Theory suggests that the range of factors that can shift cost cutoffs might include trade costs, market size, and technology. I find support for these predictions in the data.

This paper's results complement those of Evenett and Venables (2002), and Eaton et al. (2005), the two main previous contributions to deal explicitly with trade growth at the new markets margin.[2] Evenett and Venables (2002) examine the export growth of 23 developing countries to 93 foreign markets over the period 1970-1997. They work at the 3 digit level of the SITC classification. Conducting logit regressions separately for each 3 digit product, they find that the probability of exporting to a given destination is generally decreasing in distance, but increasing in market size. Exporting to proximate markets is found to be a significant predictor of geographical diversification, which Evenett and Venables (2002) argue could be consistent with learning effects. They also find some evidence that a common border and

[2]The sample selection gravity model developed by Helpman et al. (2008) allows for trade expansion at the new markets and new products margins, in addition to the intensive margin. However, the authors' empirical work does not distinguish between the first two effects, and therefore does not make it possible to examine whether particular factors impact one or the other margin more strongly. The same is true of the Tobit model used by Amurgo-Pacheco and Pierola (2008). Besedes and Prusa (2007) focus on the duration of trading relationships, not on geographical diversification as such.

common language increase the probability of observing trade for a given country dyad.

Eaton et al. (2005) use a database of French firms to analyze the determinants of export behavior. They find that bigger firms (i.e., those with higher levels of sales in France) tend to export to a larger number of foreign markets. However, they do not directly examine the empirical impact of policy-related factors such as trade costs.

The paper proceeds as follows. In the next section, I set out the hypotheses to be tested in the remainder of the paper, and motivate them by reference to recent theoretical work. Section 3 presents the dataset, empirical model, and results. Section 4 concludes, and discusses some directions for future research in this area.

2 Theoretical Motivation

This section motivates the empirical work in the remainder of the paper by relating it to a class of trade models with heterogeneous firms and market specific trade costs. I do not set out a full model, but rely instead on existing theoretical results due to Helpman et al. (2008). The comparative statics of their model's equilibrium suggest that trade expansion at the new markets margin should depend on fixed and variable trade costs, the size of the exporting country's home market, and the exporting country's technology level. In the remainder of this section, I develop the intuition behind these results, which I demonstrate more formally in the Appendix.

The model in Helpman et al. (2008) postulates a world of J countries. Identical consumers in each country have Dixit-Stiglitz preferences over a continuum of varieties with elasticity of substitution ε. On the production side, each firm produces a unit of its distinct variety using inputs costing $c_j a$, where c_j is a country-level index of factor prices, and a is an inverse measure of firm productivity. Since higher c_j means a more expensive input bundle, it can be seen as an inverse index of country productivity. In addition to standard iceberg costs τ_{ij} affecting exports from country j to country i, firms must also pay a fixed cost $c_j f_{ij}$ associated with each bilateral route. When selling in the domestic market, $\tau_{jj} = 1$ and $f_{jj} = 0$.

Firms are heterogeneous in terms of productivity, with a drawn randomly from a truncated Pareto distribution with support $[a_L, a_H]$. As in Melitz (2003), selection into export markets is based on productivity: only those firms with sufficiently high productivity (low a) can overcome the zero profit threshold a_{ij} associated with exporting from country j to country i. In light of the multi-country nature of the model, however, the outcome of this process is more complex than in Melitz (2003). Instead of selecting into just two groups, firms select into export and non-export groups with respect to each foreign market. The zero profit thresholds for all $J(J-1)$ bilateral relationships can be used to define the set M_j of foreign markets entered by at least one firm from country j:

$$M_j = \left\{ a_{ij} \mid a_{ij} \geq a_L \right\} \tag{1}$$

Assume that if a firm's marginal cost draw a is less than a_{kj} then it enters all other markets in M_j with $a_{lj} \geq a_{kj}$.[3] Then it follows that M_j is coterminous with the set of markets to which non-zero export flows from j can be observed in aggregate trade data. Changes in M_j brought about by changes in any of the full set of a_{ij}'s therefore equate to the kind of trade growth at the new markets margin–or geographical export diversification–that can be observed in aggregate trade data.

Using results in Helpman et al. (2008), it can be shown (see Appendix) that the following comparative statics hold in equilibrium:

$$\frac{da_{ij}}{d\tau_{ij}} < 0 \tag{2}$$

$$\frac{da_{ij}}{df_{ij}} < 0 \tag{3}$$

$$\frac{da_{ij}}{dc_j} < 0 \tag{4}$$

$$\frac{da_{ij}}{dY_j} > 0 \tag{5}$$

Thus, the export cost cutoff falls as fixed and variable trade costs rise, but increases in line with home

[3] Although this mechanism is intuitively appealing, Eaton et al. (2005) find only relatively weak support for it among French exporting firms. A recent paper by Lawless and Whelan (2008) reports similar difficulties using Irish data. There could be many possible explanations for these findings, including the existence of firm-specific trade costs that would be consistent with departures from the strict market ordering postulated here. However, an expansion of the canonical heterogeneous firms model in this direction is outside the scope of this paper.

GDP and technology $\left(\frac{1}{c_j}\right)$. Given the link between changes in the a_{ij}'s and shifts in the membership of M_j, these comparative statics suggest that geographical diversification of exports should similarly be decreasing in fixed and variable trade costs, but increasing in home market size and technology. In the remainder of the paper, I take these predictions to the data.

3 Empirics

My empirical strategy is straightforward, and relies on cross-sectional and cross-country variation in the data to identify the impacts of trade costs, market size, and development level on geographical export diversification. Given the importance of geographical export diversification from a development point of view, I limit the sample to developing countries, defined as all countries except those in the World Bank's high income group.[4] As an observable proxy for the number of elements in M_j (the set of export markets entered by at least one firm from country j), I use a count of the number of foreign markets to which a given country has non-zero exports.[5] Export markets are counted at the HS 2-digit level to allow for possible cross-sectoral variation in the impact of trade costs and other factors.[6] Since the dependent variable is count data, my empirical work is based on a Poisson model with sectoral fixed effects. I find that trade costs have a consistently negative and significant impact on geographical export diversification, while the size and development level of the home economy tend to act in the opposite direction. These results are highly robust to alternative specifications, including the use of data on import procedures and time as instruments for export costs, and transformation of the model into (geographical) first differences to eliminate the sectoral fixed effects.

[4] In the context of robustness checks, I show that my main results continue to hold when the country sample is varied. Importantly, excluding high income countries makes it unlikely that the results reported here are being driven by entrepôt trade, since Hong Kong and Singapore are excluded from the baseline estimation sample.

[5] In additional results, available on request, I show that the paper's conclusions are not affected by excluding very small trade flows that might be subject to excessive statistical noise.

[6] In additional results, available on request, I show that the paper's results in terms of coefficient signs continue to hold if the data are aggregated across all sectors to yield one observation per exporting country. Given the greatly reduced sample size, however, export costs and tariffs are not statistically significant in the aggregate formulation. I present disaggregated results here so as to facilitate the examination of cross-sectoral heterogeneity (see below), and because one of the independent variables of interest (tariffs) varies at the sectoral level.

3.1 Data

Data and sources are set out in full in Table 1, and descriptive statistics are in Table 2. Many of the data come from standard sources and do not require any particular discussion. However, two aspects are more novel and are discussed in detail here: export market counts, and direct measures of the cost of exporting. Due to limited availability of trade cost data, the analysis takes place using data for a single year only (2005).[7]

First, I define

$$m_j = |M_j| = \sum_{i=1}^{J} I\left(a_{ij}\right) \tag{6}$$

where $I\left(a_{ij}\right)$ is an indicator returning unity if $a_{ij} \geq a_L$, else zero. The variable m_j is thus a count of the number of markets to which exports from j are observed. I then operationalize m_j in terms of its empirical counterpart m_{es}, which varies by exporter (e) and sector (s). To do this, I use the BACI trade data included in CEPII's Market Access Map (MAcMap) database to produce counts of the number of export markets served by each country in each 2-digit HS sector. BACI is based on standard UN Comtrade data at the 6 digit HS level. The main difference for present purposes is that BACI uses a harmonization methodology to reconcile mirror flows, thereby providing more complete geographical coverage than if only a single direction of Comtrade statistics were to be used. BACI's approach to harmonization consists of computing a weighted average of mirror flows based on an estimated quality indicator for export and import declarations in each country. The methodology is set out in detail in Gaulier et al. (2007). It is advantageous to use import and export flows jointly in this way to produce measures of geographical export diversification because relying on just one set of flows could result in significant bias due to poor reporting practices in some countries.

Trade costs can cover numerous dimensions. Here, I focus on three of the most important. As is common in the gravity literature, I use international distance as a proxy for transport costs. Since data are by exporter (not bilateral), I take the average distance of each exporter from the rest of the world. The second dimension of trade costs captured here is effectively applied tariffs (i.e., including preferences).

[7] Although the export cost data discussed below are now available for two years, they exhibit almost no temporal variation at this stage. Combined with the limited availability of trade data for 2007 and 2008, this makes panel data estimates infeasible for the time being. As new data become available, it would clearly be desirable for future work to use panel data techniques to examine the robustness of the results reported here.

These are sourced from the MAcMap database for the year 2004, and aggregated to the HS 2 digit level using the reference groups methodology of Laborde et al. (2007). The essence of that aggregation approach is to limit endogeneity concerns by weighting tariffs according to the imports of a group of similar countries (not the importer itself). MAcMap tariffs include ad valorem and specific measures, with the latter converted to ad valorem terms using the reference group unit value. Again, I take the simple average across the rest of the world to obtain a figure for each exporter.

In addition to applied tariffs, I also use new data from the World Bank's *Doing Business* database to measure export costs. For the first time in 2006, the "Trading Across Borders" component of *Doing Business* captures the total official cost for exporting a standardized cargo of goods ("Export Cost"), excluding ocean transit and trade policy measures such as tariffs. The four main components of the costs that are captured are: costs related to the preparation of documents required for trading, such as a letter of credit, bill of lading, etc.; costs related to the transportation of goods to the relevant sea port; administrative costs related to customs clearance, technical controls, and inspections; and ports and terminal handling charges. The indicator thus provides a useful cross-section of information in relation to a country's approach to trade facilitation, and covers elements of variable and fixed costs. The data are collected from local freight forwarders, shipping lines, customs brokers, and port officials, based on a standard set of assumptions, including: the traded cargo travels in a 20ft full container load; the cargo is valued at $20,000; and the goods do not require any special phytosanitary, environmental, or safety standards beyond what is required internationally. These export operations cost as little as $300-$400 in Tonga, China, Israel, Singapore, and UAE, whereas they run at nearly ten times that level in Gabon and Tajikistan. On average, the cost is around $1278 per container (excluding OECD and EU countries). Closely related *Doing Business* data on the time taken at export and import have been used in empirical work by Djankov et al. (forthcoming), who find that such delays have a significant negative impact on bilateral trade.

3.2 Non-Parametric Results

Before moving to a fully-specified regression model, it is useful to take a first look at the interrelationships in the data using non-parametric techniques. To do this, I use a multivariate Lowess smoother

(Royston and Cox, 2005).[8] For ease of presentation, I do not use the sectorally disaggregated measure of geographical diversification m_{es}, but instead an aggregate count of the total number of foreign markets to which a given country exports in any sector ($m_e = \sum^s m_{es}$).

Figure 1 presents results using distance, tariffs, and Doing Business export costs as proxies for trade costs, GDP as a proxy for market size, and GDP per capita as a proxy for the exporting country's development level. Lowess smoothes provide clear evidence of a positive association between the number of export markets entered, and aggregate and per capita GDP. There is also a fairly clear negative relationship between tariffs and geographical export diversification. Results are less clear for the remaining trade cost variables, however. The smoothes for export costs and distance exhibit an inverted U-shape, although the curvature is quite weak and the peak occurs early enough in the sample that the dominant tendency in both cases is negative. Thus, although non-parametric methods provide some initial support for the hypotheses developed above, it is clearly necessary to investigate the data in more detail before reaching any strong conclusions.

3.3 Empirical Model, Estimation Results, and Robustness Checks

To proceed with the empirical analysis, it is assumed that the number of markets entered for each exporter-sector combination, m_{es}, can be adequately represented by a Poisson process. This is appropriate given that m_{es} represents strictly non-negative integer count data. The mean and variance of that process are equal to μ_{es}, and its density conditional on a set of independent variables $\mathbf{X}_{es}$ is given by:

$$f\left(m_{es} \mid \mathbf{X}_{es}\right) = \frac{\exp\left(-\mu_{es}\right) \mu_{es}^{m_{es}}}{m_{es}!} \tag{7}$$

Based on the theoretical results discussed above, a reduced-form specification for the conditional mean

[8] At each point in the sample space, Lowess runs a regression of the dependent variable on the independent variables using a subset (bandwidth = 0.8) of the available data. Observations further away from the central point in each subsample are downweighted. The smoothed plot is then constructed by joining the set of predicted values generated from the regressions. For an economic application of a very similar methodology, see Imbs and Wacziarg (2003).

function would be:

$$\mu_{es} = \delta_s \exp \left[\begin{array}{c} \beta_1 \ln(export_e) + \beta_2 \ln(dist_e) + \beta_3 \ln(1 + t_{es}) + \ldots \\ \ldots + \beta_4 \ln(gdp_e) + \beta_5 \ln(gdppc_e) \end{array} \right] \tag{8}$$

Export costs, distance, and tariffs capture the trade costs faced by exporters, while the exporting country's own GDP proxies the size of the home market. Per capita GDP in the exporting country is used as a proxy for the country-wide technology parameter c_j. Finally, the sector fixed effects δ_s control for unobservables that impact all exporters in a given sector in the same way. An important example of such a factor is worldwide demand by sector. The comparative statics presented above suggest that β_1, β_2, and β_3 should all be negative, while β_4 and β_5 should be positive.

Estimation of the fixed effects Poisson model in (7) and (8) is straightforward (Cameron and Trivedi, 2001). Results for the baseline specification appear in column 1 of Table 3. All parameters carry the expected signs and have sensible magnitudes: export costs, distance, and tariffs are all negatively associated with the number of export markets entered, while the two GDP variables exhibit a positive association. In terms of absolute value elasticities, the strongest trade cost impacts come from (in descending order) distance, tariffs, and then export costs. In terms of precision, all coefficients except tariffs are statistically significant at the 10% level.

Thus far, the data tend to support the core contentions of this paper. Concretely, 1% reductions in international transport costs, export costs, and importer tariffs are associated with increases of 0.3%, 0.2%, and 0.3% respectively in the number of export destinations serviced. One percent increases in aggregate and per capita GDP are associated with increased geographical diversification of 0.4% and 0.1% respectively. In the remainder of this section, a number of specification checks are applied to ensure that these results are robust.

3.4 Additional Controls

One important aspect of model identification in this case is the exclusion of additional country-level variables that might be driving geographical export diversification. (The sector fixed effects take care of all external influences at the sector level, which do not vary by country.) With this in mind, columns 3-5 of Table 3 include additional exporter control variables. In column 2, I include the percentage of industry in the exporting country's total value added, and the government effectiveness indicator from Kaufmann et al. (2008). These variables are intended as proxies for the exporting country's level of economic and industrial development, both of which could impact geographical export diversification. As can be seen from the table, estimated coefficients remain close in magnitude and sign to the baseline. There is some loss of precision, however: distance is only 15% significant (prob. = 0.133), while GDP per capita loses significance at all conventional levels.

Columns 3 through 5 include (alternately) a number of additional factors that have been found to be relevant to trade growth in the gravity model literature: language dummies; colonization dummies; and the exporting country's geography (internal distance and landlocked status). The intuition is that having an "international" language among a country's official languages may lessen the information costs associated with export market entry, while colonial links might result in beneficial market access or long-standing supply arrangements that might boost trade. Similarly, larger internal distances and being landlocked result in higher trade and transit costs.

The coefficients on aggregate and per capita GDP are relatively unchanged in magnitude and significance with the addition of these controls. Only in the geography regression does the coefficient on GDP per capita lose significance. However, the coefficients on the other variables undergo more substantial changes. The trade cost variables continue to have negative signs in all three regressions. In general, though, the coefficient estimates lose precision. None of the trade costs variables are 10% significant in the language model, while in the colony model only export costs are marginally significant at the 10% level (prob. = 0.104). In the geography model, distance is 10% significant and export costs are 25% significant.

The reason for this general loss in precision is that export costs and distance are both relatively strongly correlated with the new control variables. Although variance inflation factors are relatively low (less than two for the language regression), they increase by up to 60% (distance) once language dummies are added. Similarly, the correlation matrix of parameter estimates indicates significant correlation between trade costs and language, up to 60%.

Column 6 takes an even more extreme approach to controlling for exporting country characteristics by including country fixed effects. Since many of the variables of primary interest for this paper vary at the country-level, I replace them with dummy variables set equal to unity whenever the underlying variable exceeds the sample median. Thus, countries with "high" levels of trade costs–greater than the median–are coded as one, and the remainder are coded as zero. As Table 3 shows, all independent variables except tariffs have the expected signs, larger absolute value magnitudes than in the baseline regression, and are statistically significant at the 1% level.

Although results are not completely uniform across the different sets of control variables considered above, their general direction is clear: trade costs are negatively associated with geographical export diversification, but market size and development level are positively associated with the number of foreign markets entered. Although correlation with additional controls makes it difficult to obtain precise coefficient estimates, distance and Doing Business export costs are at least marginally significant at the 10% level in three of the five robustness regressions, and distance is significant in two out of five. Aggregate GDP is significant in all five robustness regressions, while per capita GDP is significant in two out of five cases.

3.5 Cross-Country Heterogeneity

An additional dimension in which it is important to check the robustness of the baseline results is the possibility of heterogeneous effects across countries. I examine this question in two ways. First, I expand the baseline sample to include all countries, developing and developed, for which data are available (Table 3, column 7). For almost all estimated coefficients, results are very close to the baseline in terms of sign, magnitude, and statistical significance. The only real exception is the tariff coefficient, which drops significantly in absolute value and remains statistically insignificant.

In column 8 of Table 3, I retain the full country sample and introduce interaction terms between GDP per capita and export costs, distance, and tariffs. The intuition is that changes in trade costs might have different impacts on geographical export diversification according to a country's level of technology and economic development. (The comparative statics in the Appendix provide formal motivation for this approach.)

Estimation results are easiest to interpret for export costs. The coefficient on the variable in levels is negative, statistically significant (5%), and considerably larger in absolute value than in the baseline model. In addition, the coefficient on the interaction term is positive and 5% significant. A test of the null hypotheses that the two coefficients sum to zero is rejected at the 5% level ($\chi^2 = 5.89$, prob. = 0.015). Together, these results suggest that geographical export diversification is particularly sensitive to reductions in export costs in poorer countries, and that this sensitivity declines as country income level increases (see Figure 2). In Burundi, the poorest country in the sample, the elasticity is estimated to be -0.45, whereas in the high income countries it is very close to zero, and even slightly positive (up to 0.08).

For distance and tariffs, results are more difficult to interpret. In both cases, the coefficient on the variable in levels is positive, and the interaction term is negative. However, neither coefficient is significant for distance, whereas both are 1% significant for tariffs. The null hypothesis that the two coefficients sum to zero is rejected for tariffs ($\chi^2 = 9.32$, prob. = 0.002) but not for distance ($\chi^2 = 2.16$, prob. = 0.142). At least in the case of tariffs, these results would therefore appear to suggest that higher tariffs might be associated with greater geographical export diversification, but that the effect is declining in the exporting country's income level (see Figure 2). It is important to note, however, that this estimated positive effect reaches zero at a relatively low level of income, roughly that of Senegal, a least developed country. From that point onwards, the elasticity is negative.

What might be driving this seemingly counter-intuitive result for countries at very low income levels? One contributing factor is the prevalence of non-tariff barriers in many sectors of export interest to countries at the very lowest income levels, such as agricultural products. Due to lack of cross-country data, it is not currently possible to control for the effects of measures such as quotas and sanitary or phytosanitary standards, which might be expected to have particularly severe impacts on the poorest countries. Since these measures are likely to be correlated with tariffs and geographical export diversification, their absence from the regression might bias the tariff coefficient.

A second factor is likely to be trade preferences extended by the major rich country markets, in particular the increasingly common granting of duty-free access to least developed countries. Although the tariff data used to construct the variable included in the regressions take full account of preferential rates, the data series is still an average across all potential importers. There is thus a greater likelihood that results in relation to tariffs are more influenced by measurement error or noise than is the case for export costs.

3.6 Cross-Sectoral Heterogeneity

It is plausible that the inclusion of sectoral dummy variables does not fully take account of the potential for the determinants of geographical export diversification to work differently in different sectors. This would be the case, for instance, if sector characteristics interact with trade costs and other country-level data. One important example of such an interaction is the sectoral elasticity of substitution. The comparative statics set out in the Appendix clearly suggest that the elasticity of geographical export diversification with respect to trade costs might vary from sector to sector based on differences in substitutability. Chaney (2008) reaches a similar conclusion.

In column 9 of Table 3, I investigate this possibility directly by interacting export costs, distance, and tariffs with estimates of the sectoral elasticity of substitution taken from Broda and Weinstein (2006).[9] For Doing Business export costs, the basic series and the interaction term are both 5% significant but carry opposite signs. This suggests that a fall in export costs has a larger impact on geographical export diversification in sectors that are more strongly differentiated. A similar effect is apparent in the case of distance, although only the variable in levels is statistically significant. Interestingly, the position is reversed in the case of tariffs: only the interaction term is statistically significant (and negative), while the levels term is positive and statistically insignificant. Although there is again evidence that the elasticity of substitution plays a damping effect in the case of tariffs, the configuration of signs is unexpected, perhaps due to the data issues discussed above. It is important to note, however, that a test of the null hypothesis that the levels term and the interaction term sum to zero is rejected in the case of export costs ($\chi^2 = 5.88$, prob. = 0.015) and distance ($\chi^2 = 3.35$, prob. = 0.067), but not in the case of tariffs ($\chi^2 = 1.98$, prob. = 0.16).

[9] Broda and Weinstein (2006) work with highly disaggregated data. I take the simple average of the US elasticity of substitution to aggregate their results to the 2-digit HS level.

3.7 Endogeneity Issues

From an identification perspective, it is important to exclude the impact of reverse or circular causality in addition to ruling out the influence of external factors. I deal with endogeneity in two ways. First, I re-estimate the model using instrumental variables to identify the causal effects of trade costs on geographical export diversification. I treat Doing Business export costs as the only potentially endogenous variable, since world tariffs and distance should be exogenous to changes in a single country's level of geographical export diversification. I also re-run the baseline regression using dependent and independent variables expressed as ratios with respect to a constant comparator country (the USA). This first difference formulation eliminates the sector fixed effects, and therefore avoids any problems that might result from endogeneity of those fixed effects. I also estimate the first difference model instrumenting for export costs.

Results for these additional robustness checks are presented in Table 4. Column 1 re-estimates the baseline model using five year lags of aggregate and per capita GDP, to ensure there are no endogeneity problems with respect to those variables. Results are almost identical to the baseline in all respects, and indeed are even slightly stronger in terms of statistical significance. Next, columns 2 and 3 present instrumental variables estimates using the IV Poisson model developed by Mullahy (1997), and a standard two-step GMM estimator. I use Doing Business data on import times and the number of official import documents as instruments for export costs. The intuition is that a significant component of export costs is driven by the efficiency of customs procedures and administrations, and thus should also be reflected in corresponding import data. However, there is no reason to believe that import procedures might be endogenous to the number of export markets entered. Similarly, import procedures would not be expected to independently impact the number of export markets entered.

Results in columns 2 and 3 are strongly supportive of the hypotheses advanced above. With the exception of tariffs in the IV Poisson estimates and GDP per capita in both estimates (statistically insignificant), all coefficients have the expected signs and magnitudes, and are at least 10% significant. Interestingly, all elasticities are noticeably larger in absolute value than in the baseline specification. Most importantly, export costs and distance both have an elasticity of -0.5 in the IV Poisson specification, compared with elasticities of -0.2 and -0.3 respectively in the baseline model. Taking account of potential endogeneity of export costs can be seen to strengthen the baseline results.

As can be seen from Table 4, results are very similar from the IV Poisson and GMM specifications. Although these are count data, it is therefore appropriate to rely on the results of diagnostic tests based on the GMM results. Column 3 shows that endogeneity is indeed a serious issue for export costs, with the null of exogeneity rejected at the 1% level. In terms of the appropriateness of import times and documents as instruments for export costs, the first stage results in column 4 show the instruments to be strongly correlated with the regressors of interest: the first stage F-test (12.72) rejects the null hypothesis that the coefficients on the two excluded instruments are jointly equal to zero at the 1% level. There is thus little concern that the instruments in this case are weak. Moreover, the Hansen J-test fails to reject the null that the overidentifying restriction is valid. Together, these results strongly support the relevance, exogeneity, and excludability of the instrumental variables, and provide a sound basis for concluding that trade costs have a causal impact on geographical export diversification.

To deal with the potential endogeneity of the sector fixed effects, I re-estimate the baseline model expressing all variables in terms of ratios with respect to a common comparator country (the USA). Column 5 of Table 4 presents results for the simple first difference specification, columns 6-7 present instrumental variables results using Poisson and GMM, and column 8 presents the first-stage instrumental variable regression. In all but one case, estimated coefficients have the expected signs, magnitudes that are comparable to the baseline and fixed effects IV specifications, and are statistically significant at the 5% level. First difference results, with and without IV estimation, thus also tend to reinforce the evidence already presented in support of the hypotheses of this paper. The exception to that rule is the tariff variable. In the simple first difference specification, it is positive and statistically significant. However, the coefficient loses statistical significance in the instrumental variables specifications. The reason for these unexpected results would appear to be that the tariffs faced by the USA are very closely correlated (0.82) with those faced by other countries. One reason for this is that multilateral tariff reductions through the GATT/WTO system have brought about a generalized fall in tariffs over time, with results generally made available to all trading partners via the most favored nation clause. As a result of this effect, it is difficult to extract any meaningful information from the ratio of the tariff rates faced by the exporting country and the USA.

4 Conclusions

This paper has provided some of the first evidence on the factors driving the geographical spread of developing country trade. In the preferred econometric specification estimated using a Poisson IV model, I find that 10% reductions in export costs and transport costs (distance) are associated with approximately 6% increases in the number of export markets served. A 10% reduction in foreign tariffs is associated with a 5% increase in geographical export diversification, but this effect is only statistically significant at the 20% level. Finally, the data also suggest that export costs have stronger effects on geographical export diversification in poorer countries, and that export costs and transport costs have stronger effects in sectors that are relatively more differentiated. Since differentiation is usually associated with increased manufacturing value added, these results are of particular interest from a development point of view.

There are a number of ways in which future research could extend the results presented here. First, as additional data become available from the Doing Business project, it will become possible to extend the empirical analysis to a panel data framework, and thus to take better account of the dynamics of geographical diversification. Second, it will be important to pay attention as well to the lessons that can be learned from firm level data that track the entry of individual exporters into overseas markets. Existing evidence (Eaton et al., 2005; Lawless and Whelan, 2008) is patchy on the market entry ordering postulated here, and it would be interesting to investigate alternative mechanisms at the micro-level, and to then implement them in a fully specified theoretical model. Finally, future work could usefully address the welfare economics of geographical export diversification, to complement the instrument-based approach taken here. In policy terms, it will be important to accurately identify the full range of costs and benefits associated with diversification.

References

Amurgo-Pacheco, Alberto, and Martha D. Pierola, 2008, "Patterns of Export Diversification in Developing Countries: Intensive and Extensive Margins", Policy Research Working Paper No. 4473, The World Bank.

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

Baldwin, Richard E., and James Harrigan, 2007, "Zeros, Quality, and Space: Trade Theory and Trade Evidence", Working Paper No. 13214, NBER.

Besedes, Tibor and Thomas J. Prusa, 2007, "The Role of Extensive and Intensive Margins and Export Growth", Working Paper No. 13628, NBER.

Brenton, Paul, and Richard Newfarmer, 2007, "Watching More than the Discovery Channel: Export Cycles and Diversification in Development", Policy Research Working Paper No. 4302, The World Bank.

Broda, Christian, and David E. Weinstein, 2006, "Globalization and the Gains from Variety", *The Quarterly Journal of Economics*, 121(2), 541-585.

Brainard, William C., and Richard N. Cooper, 1965, "Uncertainty and Diversification in International Trade", Discussion Paper No. 197, Cowles Foundation.

Cadot, Olivier; Céline Carrère; and Vanessa Strauss-Kahn, 2007, "Export Diversification: What's Behind the Hump?", Discussion Paper No. 6590, CEPR.

Cameron, A. Colin, and Pravin K. Trivedi, 2001, "Essentials of Count Data Regression" in Badi H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Malden, Ma.: Blackwell.

Chaney, Thomas, 2008, "Distorted Gravity: The Intensive and Extensive Margins of International Trade", *American Economic Review*, 98(4), 1707-1721.

Djankov, Simeon, Caroline Freund, and Cong S. Pham, Forthcoming, "Trading on Time", *Review of Economics and Statistics*.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, 2005, "An Anatomy of International Trade: Evidence from French Firms", Mimeo, www.econ.umn.edu/kortum/papers/ekk1005.pdf.

Evenett, Simon J., and Anthony J. Venables, 2002, "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows", Mimeo, http://www.evenett.com/working/setvend.pdf.

Gaulier, Guillaume, Soledad Zignago, Dieudonné Sondjo, Adja A. Sissoko, and Rodrigo Paillacar, 2007, "BACI: A World Database of International Trade at the Product Level 1995-2004 Version", Mimeo, http://www.cepii.fr/anglaisgraph/bdd/baci/baciwp.pdf.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, 2008, "Estimating Trade Flows: Trading Partners and Trading Volumes", *Quarterly Journal of Economics*, 123(2), 441-487.

Hummels, David L., and Peter J. Klenow, 2005, "The Variety and Quality of a Nation's Exports", *American Economic Review*, 95(3), 704-723.

Imbs, Jean, and Romain Wacziarg, 2003, "The Stages of Diversification", *American Economic Review*, 93(1), 63-86.

Kaufmann, Daniel; Aart Kraay; and Massimo Mastruzzi, 2008, "Governance Matters VII: Aggregate and Individual Governance Indicators, 1996-2007", Policy Research Working Paper No. 4654, The World Bank.

Laborde, David, Houssein Boumelassa, and Maria C. Mitaritonna, Forthcoming, "A Consistent Picture of the Protection Across the World in 2004: MAcMap HS6 V2", Working Paper, CEPII.

Lawless, Martina, and Karl Whelan, 2008, "Where Do Firms Export, How Much, and Why?", Working Paper No. 200821, University College, Dublin.

Mayer, Thierry, and Soledad Zignago, 2006, "Notes on CEPII's Distance Measures", Working Paper, CEPII.

Melitz, Marc J., 2003, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, 71(6), 1695-1725.

Mullahy, John, 1997, "Instrumental Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior", *Review of Economics and Statistics*, 79(4), 586-593.

Royston, Patrick, and Nicholas J. Cox, 2005, "A Multivariable Scatterplot Smoother", *Stata Journal*, 5(3), 405-412.

Schott, Peter K., 2004, "Across-Product Versus Within-Product Specialization in International Trade", *Quarterly Journal of Economics*, 119(2), 646-677.

Appendix: Comparative Statics of the Helpman et al. (2008) Model

Under the assumptions set out in the main text, Helpman et al. (2008) show that their model's equilibrium can be described by the following relations (see their equations 4, 5, and 7):

$$a_{ij}^{1-\varepsilon} = \frac{c_j f_{ij}}{Y_i(1-\alpha)} \left(\frac{\alpha}{\tau_{ij} c_j}\right)^{1-\varepsilon} P_i^{1-\varepsilon} \equiv \frac{c_j f_{ij}}{Y_i(1-\alpha)} \left(\frac{\alpha}{\tau_{ij} c_j}\right)^{1-\varepsilon} \left[\tilde{P}_i + \left(\frac{c_j \tau_{ij}}{\alpha}\right)^{1-\varepsilon} N_j V_{ij}\right] \tag{9}$$

$$P_i^{1-\varepsilon} = \sum_{j=1}^{J} \left(\frac{c_j \tau_{ij}}{\alpha}\right)^{1-\varepsilon} N_j V_{ij} \equiv \tilde{P}_i + \left(\frac{c_j \tau_{ij}}{\alpha}\right)^{1-\varepsilon} N_j V_{ij} \tag{10}$$

$$V_{ij} = \begin{cases} \int_{a_L}^{a_{ij}} a\, dG(a) & a_{ij} \geq a_L \\ 0 & a_{ij} < a_L \end{cases} \tag{11}$$

$$G(a) = \frac{a^k - a_L^k}{a_H^k - a_L^k} \tag{12}$$

The first condition is the zero profit marginal cost cutoff for the country pair $\{i, j\}$. The second and third equations define a CES price index for each country, and the fourth is a truncated Pareto distribution with support $[a_L, a_H]$ from which marginal costs are drawn. The only endogenous variables are the marginal cost cutoff and the price index, and it is possible to use these equations to solve for them in terms of model parameters.

To generate the hypotheses tested in this paper, it is sufficient to focus on the marginal cost cutoff condition. Together, the comparative statics below suggest that geographical export diversification should be positively associated with the size and sophistication of the home market, but negatively associated with fixed and variable trade costs.

Variable Trade Costs

Taking the derivative of the export cutoff with respect to variable trade costs gives:

$$(1-\varepsilon)\,a_{ij}^{-\varepsilon}\frac{da_{ij}}{d\tau_{ij}}=\frac{c_j f_{ij}}{Y_i(1-\alpha)}\left[(\varepsilon-1)\left(\frac{\alpha}{c_j}\right)^{1-\varepsilon}\tau_{ij}^{\varepsilon-2}\tilde{P}_i+N_j\frac{dV_{ij}}{da_{ij}}\frac{da_{ij}}{d\tau_{ij}}\right] \tag{13}$$

$$\therefore\ \frac{da_{ij}}{d\tau_{ij}}=\frac{\frac{(\varepsilon-1)c_j f_{ij}}{Y_i(1-\alpha)}\left(\frac{\alpha}{c_j}\right)^{1-\varepsilon}\tau_{ij}^{\varepsilon-2}\tilde{P}_i}{(1-\varepsilon)\,a_{ij}^{-\varepsilon}-\frac{c_j f_{ij}}{Y_i(1-\alpha)}N_j\frac{dV_{ij}}{da_{ij}}}<0 \tag{14}$$

where the final inequality follows from the fact that the constraints placed on the model parameters ensure $\varepsilon>1$and $0<\alpha=1-\frac{1}{\varepsilon}<1$. To derive the sign of $\frac{dV_{ij}}{da_{ij}}$, I substitute the Pareto CDF into the expression for V_{ij} to get:

$$V_{ij}=\frac{k}{a_H^k-a_L^k}\int_{a_L}^{a_{ij}}a^{k-\varepsilon}da \tag{15}$$

and so by the fundamental theorem of calculus, $\frac{dV_{ij}}{da_{ij}}=\frac{ka_{ij}^{k-\varepsilon}}{a_H^k-a_L^k}>0$

Fixed Trade Costs

The derivative of the export cutoff with respect to fixed trade costs is:

$$(1-\varepsilon)\,a_{ij}^{-\varepsilon}\frac{da_{ij}}{df_{ij}}=\frac{c_j}{Y_i(1-\alpha)}\left[\left(\frac{\alpha}{\tau_{ij}c_j}\right)^{1-\varepsilon}\tilde{P}_i+N_jV_{ij}+f_{ij}N_j\frac{dV_{ij}}{da_{ij}}\frac{da_{ij}}{df_{ij}}\right] \tag{16}$$

$$\therefore\ \frac{da_{ij}}{df_{ij}}=\frac{\frac{c_j}{Y_i(1-\alpha)}\left(\frac{\alpha}{\tau_{ij}c_j}\right)^{1-\varepsilon}\tilde{P}_i+N_jV_{ij}}{(1-\varepsilon)\,a_{ij}^{-\varepsilon}-\frac{c_j f_{ij}}{Y_i(1-\alpha)}N_j\frac{dV_{ij}}{da_{ij}}}<0 \tag{17}$$

where the sign of the derivative again follows from the models constraints on the elasticity of substitution, and the fact that $\frac{dV_{ij}}{da_{ij}}>0$.

Home Market Technology

Next, take the derivative of the export cutoff condition with respect to c_j, an inverse measure of home country technology:

$$(1-\varepsilon)\,a_{ij}^{-\varepsilon}\frac{da_{ij}}{dc_j}=\frac{f_{ij}}{Y_i(1-\alpha)}\left(\varepsilon\left(\frac{\alpha}{\tau_{ij}c_j}\right)^{1-\varepsilon}\widetilde{P}_i+N_jV_{ij}\right)+\frac{c_j f_{ij}}{Y_i(1-\alpha)}N_j\frac{dV_{ij}}{da_{ij}}\frac{da_{ij}}{dc_j} \tag{18}$$

$$\therefore\ \frac{da_{ij}}{dc_j}=\frac{\frac{f_{ij}}{Y_i(1-\alpha)}\left(\varepsilon\left(\frac{\alpha}{\tau_{ij}c_j}\right)^{1-\varepsilon}\widetilde{P}_i+N_jV_{ij}\right)}{(1-\varepsilon)\,a_{ij}^{-\varepsilon}-\frac{c_j f_{ij}}{Y_i(1-\alpha)}N_j\frac{dV_{ij}}{da_{ij}}}<0 \tag{19}$$

where the final inequality follows from the same considerations as above. Since $c_j$is an inverse measure of exporting country technology, the negative sign on the derivative indicates that geographical export diversification should be positively associated with the level of technology.

Home Market Size

The expression used thus far for the export cutoff does not include the home market's GDP Y_j. To see the role of that factor, first note that exports from $j$to $i$can be expressed as follows (Helpman et al., 2008, equation 6):

$$M_{ij} = \left(\frac{c_j \tau_{ij}}{\alpha P_i}\right)^{1-\varepsilon} Y_i N_j V_{ij} \tag{20}$$

Summing over all destinations, including the home market, and imposing equality between income and expenditure gives:

$$Y_j \equiv \sum_{i=1}^{J} M_{ij} = \left(\frac{c_j}{\alpha}\right) N_j \sum_{h=1}^{J} \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} Y_h V_{hj} \tag{21}$$

which can be rearranged and solved for Y_i:

$$Y_i = \frac{1}{V_{ij}} \left(\frac{P_i}{\tau_{ij}}\right)^{1-\varepsilon} \left[\frac{Y_j \alpha}{N_j c_j} - \left(\frac{\tau_{jj}}{P_j}\right)^{1-\varepsilon} Y_j V_{jj} - \sum_{h \neq i,j}^{J} \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} Y_h V_{hj}\right] \tag{22}$$

Substituting this expression into the export cutoff and canceling terms gives:

$$a_{ij}^{1-\varepsilon} = \frac{c_j f_{ij}}{(1-\alpha)} \left(\frac{\alpha}{c_j}\right)^{1-\varepsilon} V_{ij} \left[\frac{Y_j \alpha}{N_j c_j} - \left(\frac{\tau_{jj}}{P_j}\right)^{1-\varepsilon} Y_j V_{jj} - \sum_{h \neq i,j}^{J} \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} Y_h V_{hj}\right]^{-1} \tag{23}$$

I can now take the derivative with respect to Y_j (ignoring indirect effects) and rearrange:

$$\therefore \frac{da_{ij}}{dY_j} = \frac{-\frac{c_j f_{ij}}{(1-\alpha)} \left(\frac{\alpha}{c_j}\right)^{1-\varepsilon} V_{ij} \left[\frac{Y_j \alpha}{N_j c_j} - \left(\frac{\tau_{jj}}{P_j}\right)^{1-\varepsilon} Y_j V_{jj} - \sum_{h \neq i,j}^{J} \left(\frac{\tau_{hj}}{P_h}\right)^{1-\varepsilon} Y_h V_{hj}\right]^{-2} \left(\frac{\alpha}{N_j c_j} - \left(\frac{\tau_{jj}}{P_j}\right)^{1-\varepsilon} V_{jj}\right)}{(1-\varepsilon)\, a_{ij}^{-\varepsilon}} \tag{24}$$

The denominator of this expression is clearly negative, based on the parameter constraints discussed above. However, the sign of the numerator is ambiguous. The sign of the derivative will be positive provided that $\frac{\alpha}{N_j c_j} > \left(\frac{\tau_{jj}}{P_j}\right)^{1-\varepsilon} V_{jj}$. To demonstrate that this condition will usually hold, I rearrange the expression, set $\tau_{jj} = 1$, and substitute the price index to show that the condition amounts to:

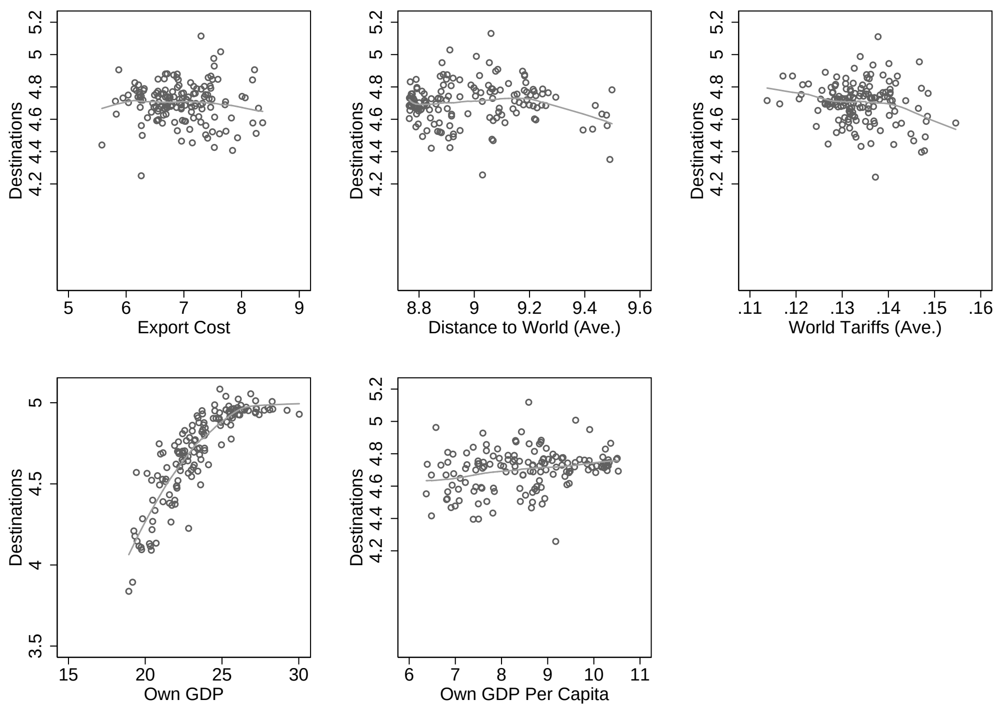
$$\frac{\alpha}{c_j} > \frac{N_j V_{jj}}{P_j^{1-\varepsilon}} \equiv \frac{N_j V_{jj}}{\left(\frac{c_j}{\alpha}\right)^{1-\varepsilon} N_j V_{jj} + \sum_{h \neq j}^{J} \left(\frac{c_h \tau_{jh}}{\alpha}\right)^{1-\varepsilon} N_h V_{jh}} \tag{25}$$

$$\therefore 1 > \frac{\frac{c_j}{\alpha} N_j V_{jj}}{\left(\frac{c_j}{\alpha}\right)^{1-\varepsilon} N_j V_{jj} + \sum_{h \neq j}^{J} \left(\frac{c_h \tau_{jh}}{\alpha}\right)^{1-\varepsilon} N_h V_{jh}} \tag{26}$$

All summation terms in the denominator are positive, so summing over large J and large N_j should result in a denominator that is significantly larger than the numerator, thereby ensuring that the condition holds, and the derivative is positively signed.

Tables and Figures

Figure 1: Non-parametric regression results.



1. Results obtained using the multivariate Lowess smoother. All variables are in logarithms.
2. The dependent variable is the number of export markets entered, aggregated across all sectors. The independent variables are Doing Business export costs, average distance to the world, world tariffs, exporting country GDP, and exporting country GDP per capita.
3. Estimation is based on the full sample, i.e. all countries for which data are available.

Figure 2: Estimated elasticities from per capita GDP interaction model.

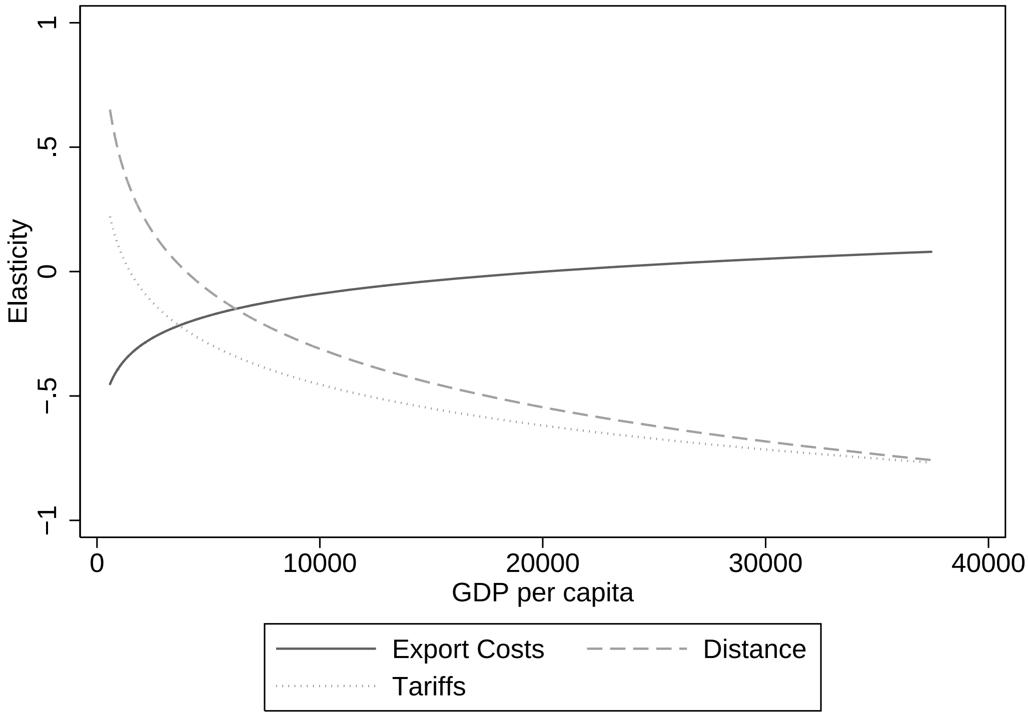


Table 1: Data and sources.

Variable	Definition	Year	Source
Colonization	Dummy variables equal to unity for countries colonized by Britain, France, Spain, Portugal, and Russia, else zero.	NA	CEPII
Destinations	Count of the number of countries to which the exporting country has strictly positive export flows, by HS 2-digit sector.	2005	MAcMap
Distance	Average of the great circle distances between the main cities of the exporting country and all other countries.	NA	CEPII
Export Cost	Official fees levied on a 20 foot container leaving the exporting country, including document preparation, customs clearance, technical control, terminal handling charges, and inland transit.	2006	Doing Business
GDP	Gross domestic product, current USD.	2000, 2005	WDI
GDPPC	GDP per capita, current USD.	2000, 2005	WDI
Governance	Government effectiveness indicator, rescaled to min = 1.	2005	WGI
Import Documents	Official documents required to import a 20 foot container, including bank documents, customs declaration and clearance documents, port filing documents, and import licenses.	2006	Doing Business
Import Time	Time taken to complete all official procedures for importing a 20 foot container.	2006	Doing Business
Industry %	Industry value added, % GDP.	2005	WDI
Internal Distance	Internal distance of the exporting country = $\frac{2}{3}\sqrt{\frac{area}{\pi}}$.	NA	CEPII
Landlocked	Dummy variable equal to unity for landlocked countries, else zero.	NA	CEPII
Language	Dummy variables equal to unity for countries with English, French, Spanish, Portuguese, or Russian as an official language, else zero.	NA	CEPII
Sigma	Elasticity of substitution for the USA, simple average by HS 2-digit sector.	1990-2001	Broda and Weinstein (2006)
Tariffs	Average applied ad valorem tariff in the rest of the world, by HS-2 sector. Aggregated by reference groups (Laborde et al., 2007).	2005	MAcMap

Table 2: Descriptive statistics.

Variable	Obs	Mean	Std. Dev.	Min.	Max.	$\rho\,(destinations)$
Col. ESP	20064	0.11	0.32	0	1	0.03
Col. FRA	20064	0.17	0.37	0	1	-0.21
Col. GBR	20064	0.34	0.48	0	1	-0.14
Col. PRT	20064	0.03	0.18	0	1	-0.08
Col. RUS	20064	0.07	0.26	0	1	-0.09
Destinations	20064	30.28	36.66	0	142	NA
Distance	19584	8309.39	1798.87	6401.68	13342.11	-0.08
Export Cost	15936	1192.71	754.65	265	4300	-0.29
GDP	17280	1.99E+11	9.35E+11	4.83E+07	1.10E+13	0.39
GDPPC	14688	8895.74	9379.04	584.22	37436.53	0.6
Governance	18528	3.15	1.02	1	5.38	0.55
Import Documents	16224	10.38	4	2	19	-0.31
Import Time	16224	35.88	25.46	3	139	-0.37
Industry %	15456	30.1	13.26	7.09	94.21	0.1
Internal Distance	19776	197.22	233.64	1	1554.24	0.31
Landlocked	19776	0.17	0.38	0	1	-0.21
Lang. ESP	20064	0.12	0.32	0	1	-0.02
Lang. FRA	20064	0.18	0.39	0	1	-0.16
Lang. GBR	20064	0.33	0.47	0	1	-0.14
Lang. PRT	20064	0.04	0.19	0	1	-0.05
Lang. RUS	20064	0.01	0.12	0	1	0.01
Tariffs	20064	0.14	0.15	0.03	12.81	-0.01

Table 3: Baseline estimates and robustness checks.

	Base	Base + Controls	Base + Controls	Base + Controls	Base + Controls	Base + Controls	All Countries	Base + GDPPC	Base + Sigma
Export Cost	-0.151**	-0.102*	-0.058	-0.113	-0.077	-1.363***	-0.138**	-1.267**	-0.176**
	[0.069]	[0.062]	[0.070]	[0.070]	[0.067]	[0.002]	[0.057]	[0.528]	[0.072]
Distance	-0.329*	-0.286	-0.017	-0.119	-0.311*	-0.733***	-0.388**	2.808	-0.358*
	[0.179]	[0.186]	[0.245]	[0.283]	[0.182]	[0.001]	[0.172]	[1.887]	[0.196]
Tariffs	-0.279	-0.209	-0.331	-0.251	-0.326	0.959***	-0.079	1.735***	1.212
	[0.318]	[0.271]	[0.339]	[0.299]	[0.322]	[0.002]	[0.226]	[0.553]	[0.844]
GDP	0.353***	0.369***	0.362***	0.356***	0.446***	0.419***	0.296***	0.297***	0.353***
	[0.019]	[0.018]	[0.020]	[0.020]	[0.036]	[0.002]	[0.022]	[0.019]	[0.019]
GDPPC	0.142***	0.012	0.176***	0.157***	0.051	-0.282	0.098**	2.286	0.143***
	[0.046]	[0.056]	[0.044]	[0.047]	[0.058]	[0.230]	[0.038]	[1.857]	[0.046]
Exp.*GDPPC								0.128**	
								[0.059]	
Dist.*GDPPC								-0.339	
								[0.207]	
Tar.*GDPPC								-0.238***	
								[0.070]	
Exp.*Sigma									0.002**
									[0.001]
Dist.*Sigma									0.002
									[0.004]
Tar.*Sigma									-0.084*
									[0.049]
Obs.	11808	11424	11808	11808	11808	11808	14496	14496	11808
Countries	123	119	123	123	123	123	151	151	123
Additional Controls		Industry % Governance	Language Dummies	Colonization Dummies	Internal Dist. Landlocked	Ctry. Fixed Effects			

1. The dependent variable in all cases is *destinations* (in levels). All independent variables except *sigma* are in logarithms. In column 6, all independent variables except tariffs are replaced with dummy variables equal to unity if a country is above the sample median for that series, else zero.

2. Columns 1-6 and 9 estimate using data for developing countries only, i.e. all countries except those in the World Bank's high income group. The estimation sample in columns 7 and 8 consists of all countries.

3. All models are estimated by Poisson, and include fixed effects by HS 2-digit sector. Robust standard errors adjusted for clustering by country are in square brackets under the coefficient estimates. Statistical significance is indicated using * (10%), ** (5%), and *** (1%).

Table 4: IV and first-difference estimates.

	Base + Lags	IV Poisson	IV GMM	Stage One	First Differences	First Diff. + IV Poisson	First Diff. + IV GMM	Stage One
Export Cost	-0.172***	-0.564**	-0.869***		-0.173***	-0.547**	-0.860***	
	[0.066]	[0.283]	[0.315]		[0.067]	[0.250]	[0.300]	
Distance	-0.470**	-0.611**	-0.758**	-0.469**	-0.449**	-0.651***	-0.719**	-0.470**
	[0.191]	[0.252]	[0.322]	[0.227]	[0.193]	[0.251]	[0.311]	[0.226]
Tariffs	-0.206	-0.453	-0.690**	-0.436***	0.306***	0.096	-0.06	-0.242**
	[0.288]	[0.336]	[0.351]	[0.161]	[0.107]	[0.192]	[0.183]	[0.096]
GDP	0.376***	0.442***	0.409***	-0.075***	0.382***	0.391***	0.388***	-0.075***
	[0.018]	[0.032]	[0.028]	[0.024]	[0.018]	[0.028]	[0.028]	[0.024]
GDPPC	0.051	0.048	0.028	0.067	0.045	0.045	0.026	0.067
	[0.035]	[0.053]	[0.055]	[0.055]	[0.035]	[0.046]	[0.053]	[0.055]
Import Time				0.566***				0.567***
				[0.121]				[0.121]
Import Docs.				-0.332**				-0.332**
				[0.156]				[0.155]
Obs.	12576	12576	11731	11731	12576	12576	11731	11731
Countries	131	131	131	131	131	131	131	131
R^2			0.62	0.26			0.43	0.26
Instr. F			11.31***				11.41***	
Exog.			7.466***				8.274***	
Hansen J			0.009				0.013	

1. Dependent variable for the Poisson regressions is *destinations* and for the GMM regressions it is $ln(destinations)$. All independent variables are in logarithms.

2. All models are estimated using data for developing countries only, i.e. all countries except those in the World Bank's high income group.

3. Models in Columns 1-4 include fixed effects by HS 2-digit sector. Models in Columns 5-8 are in terms of ratios relative to the USA.

4. Robust standard errors adjusted for clustering by country are in square brackets under the coefficient estimates. Standard errors for the IV Poisson models are estimated by bootstrapping (200 replications). Statistical significance is indicated using * (10%), ** (5%), and *** (1%).

5. Instrument F tests the null hypothesis that the coefficients on the excluded instruments are jointly equal to zero in the first stage regression. Exogeneity tests the null hypothesis that export costs are exogenous to the number of markets entered. Hansen's J tests the joint null hypothesis that the instruments are uncorrelated with the main regression error term, and that they are correctly excluded from the main regression.