World Development, 38(9), 1217-1228.

Geographical Diversification of Developing Country Exports

Ben Shepherd, Principal.

March 25, 2010.

developing trade consultants
policy • research • capacity building

349 5th Avenue
New York, NY 10016
Ben@Developing-Trade.com

GEOGRAPHICAL DIVERSIFICATION OF DEVELOPING COUNTRY EXPORTS

BEN SHEPHERD

Principal, Developing Trade Consultants Ltd.

Email: Ben@Developing-Trade.com

March 25, 2010

SUMMARY

This paper shows that export costs, tariffs, and international transport costs are all robustly associated with geographical export diversification in a sample of 117 developing countries. Reducing each of them by one standard deviation could lead to increases in the number of export destinations of 12%, 3%, and 4% respectively. From a geographical diversification point of view, trade facilitation at home is an important complement to improving market access abroad. Customs procedures and document preparation in exporting countries have particularly strong effects. Trade costs in general have larger effects in manufacturing, and highly differentiated sectors.

Keywords**:** Trade and development; Market access; Trade facilitation; Export diversification; Economic geography.

AUTHOR'S ACKNOWLEDGEMENTS

This paper was drafted while the author was working in the World Bank's Development Research Group, as part of a DFID-funded project on "Trade Costs and Facilitation: The Development Dimension". It was revised during a postdoctoral fellowship at Princeton University's Niehaus Center for Globalization and Governance. The findings, interpretations, and conclusions expressed in this paper are those of the author only. Allen Dennis, Bernard Hoekman, Patrick Messerlin, Chad Bown , Matt Cole, Felix Eschenbach, Ana Margarida Fernandes, Joe Francois, and anonymous referees all provided helpful comments and guidance.

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 product extensive margin); an increase in the unit values of traded goods (the quality margin); and creation of trading relationships between new partners (the geographical extensive margin). Although there is a vast literature on the determinants of intensive margin trade growth (e.g., Anderson & Van Wincoop, 2003), and an emerging body of work on the product extensive margin (e.g., Hummels & Klenow, 2005; Broda & Weinstein, 2006) and the quality margin (Schott, 2004; Baldwin & Harrigan, 2007), there is almost no empirical work specifically on the geographical extensive margin. Yet recent findings suggest that growth at the geographical extensive margin—which is akin to the concept of geographical export diversification in the policy literature—can be an important mechanism through which developing countries can become more integrated in the world trading system. For example, Evenett & 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 & Newfarmer (2007) suggest that the proportion was around 18% for the period 1995-2004. Although Besedes & 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, geographical expansion dominates product-dimension expansion in poorer countries.

The current financial crisis and trade collapse provide an additional rationale for diversifying exports geographically. Although all major markets have been affected by the “Great Recession”, the depth of the resulting drop in demand, as well as the timing and rate of recovery, differ noticeably across markets. Imperfect correlation among major market demand shocks means that developing country exporters serving a wider range of overseas markets may be less affected by overseas recessions than those serving a small number of markets. The argument is, of course, stronger for more "standard" recessions, in which the correlation across major markets is usually substantially weaker than in the present case. Geographical diversification can act like a form of portfolio diversification for developing country exporters, helping to minimize risk for a given level of return (Brainard and Cooper, 1965). This leads to a more stable flow of export income, in addition to other gains such as learning about foreign market conditions and technologies through exporting.

This paper's focus is on the instruments available to developing country policymakers concerned with supporting geographical export diversification. It aims to fill the void that currently exists in relation to the determinants of trade growth at the geographical extensive margin by examining the impact of three sets of factors: market size and development level in the exporting country; international trade costs (distance, tariffs) facing the exporting country; and export costs (border formalities, customs, documentation, and inland transport) 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 geographical extensive margin positively, but the remaining factors have a negative impact. Moreover, improved trade facilitation—i.e., lower export costs at home—has the potential to increase geographical diversification more strongly than comparable changes in market access abroad or international transport costs.

These results are highly robust to estimation using disaggregated data (by exporting country 2-digit ISIC sector) and aggregate data (by exporting country, summing over sectors), inclusion of a wide range of additional control variables, and estimation via instrumental variables techniques. I also find evidence that the effect of export costs is stronger in manufacturing compared with primary industry, and within manufacturing is stronger for relatively differentiated goods. In policy terms, these results are particularly relevant to lower income countries engaged in industrialization, i.e., a shift towards increasingly differentiated manufactured goods, and away from primary industry.

What is the economic intuition behind these results? Recent advances in trade theory provide a powerful explanation for why countries export goods to some overseas markets but not others. According to recent models in which firms are heterogeneous in productivity—Melitz (2003) is the canonical version—only a relatively small proportion of firms in an economy export. The rest serve the domestic market only. The set of foreign markets entered by exporters is determined by the entry costs they face, which can vary across countries. 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 a given 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. Aggregating these effects makes it possible to analyze the process in terms of the set of foreign markets entered, rather than individual bilateral trading relationships. An expansion in the set of markets entered is the process of trade growth at the geographical extensive margin that is central to this paper. Theory suggests that the range of factors that can shift cost cutoffs and thus influence this process can include trade costs, market size, and technology. I find support for these predictions in the data.

This paper's results complement those of Evenett & Venables (2002), and Eaton et al. (2008), the two main previous contributions to deal explicitly with trade growth at the geographical extensive margin.[1] Evenett & Venables (2002) examine the export growth of 23 developing countries to 93 foreign markets over the period 1970-1997. Conducting logit regressions separately for each SITC 3-digit product and country-pair, 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 the authors argue could be consistent with learning effects. They also find some evidence that a common border and common language increase the probability of observing trade for a given country dyad.

There are two main differences between this paper and Evenett & Venables (2002). First, this paper includes a range of policy-related trade costs, in addition to distance as a proxy for international transport costs. As a result, it has potentially wider implications for trade and development policy. Second, the analytical approach of this paper focuses on the set of overseas markets served by a given exporting country, rather than the existence or not of a particular bilateral trading relationship at the product line level. Results from this single-equation, pseudo-panel estimation framework are thus much easier to interpret than the 4,000 sets of parameter estimates reported by Evenett and Venables (2002).

Eaton et al. (2008) 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. By the same token, larger foreign markets tend to attract more entry by French firms. In counterfactual simulations, they show that lowering traditional (variable cost) trade barriers increases exports primarily at the intensive margin, but that lowering (fixed cost) entry barriers produces a large effect at the extensive margin, as more French firms enter each foreign market. (Using similar data, Koenig, 2009 finds that distance—a proxy for trade costs—and foreign market size have significant effects at the extensive margin.)

This paper builds on these firm-level results in two ways. First, it uses a theoretical framework with similar foundations but aggregates it to the country-level, so that it is possible to use a global database to test the model's predictions; the analysis of Eaton et al. (2008) is at the firm-level, and is limited to a single country (France). Second, this paper introduces a range of trade cost factors that are of interest from a policy perspective. Since these factors vary substantially across countries but not within, a global framework is needed to examine theoretical predictions as to the impact of trade costs on geographical diversification.

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 policy implications as well as directions for future research.

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. Whereas the first generation of trade models with product differentiation relied on an analysis using a single, representative firm (e.g., Krugman, 1979), the new class of models following Melitz (2003) and Chaney (2008) allow for each firm in the economy to have a different level of productivity. These new models provide an explicit theoretical basis for the extensive margin of trade in the geographical and product dimensions.

I do not set out a full model here, 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) assumes a world of J countries. Although the analysis takes place using a representative sector, all results are easily generalized to a multi-sector framework by including additional sectors multiplicatively in the utility function (see Chaney, 2008 for an example). Identical consumers in each country have Dixit-Stiglitz preferences over a continuum of varieties with intra-sectoral 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 or technology. The interaction between c_j and a means that a can be interpreted as a within-country index of relative firm-level 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 shape parameter k and support $[a_L, a_H]$. In addition to being analytically tractable, the Pareto distribution also reflects the empirical regularity that a few firms in a sector are very large and productive, but the vast majority of them are small and relatively unproductive. As in Melitz (2003), firms self-select into export markets based on their productivity draw: 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.[2] 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:

$$M_j = \{a_{ij} \,|\, a_{ij} \geq a_L\} \tag{1}$$

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 geographical extensive margin--or geographical export diversification in the policy literature—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 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-sectoral 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 at first 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] 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, but that 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 an instrumental variables estimator to account for the possibility of reverse causation.

Export markets are counted at the ISIC 2-digit level, and the empirical work proceeds at this level of aggregation. Although overseas tariffs are the only independent variable to vary at the country-sector level, a disaggregated approach is appropriate because it makes it possible to test hypotheses based on inter-sectoral differences. I show, for example, that the association between trade costs and geographical export diversification is stronger for manufactured versus agricultural goods, and for relatively differentiated manufactures versus relatively homogeneous ones. In any case, I also show in robustness checks that the paper's results continue to hold if the data are aggregated to the country level so as to eliminate concerns over clustering of the errors in the sectoral specification.

(a) Data

Data and sources are set out in full in Table 1, and descriptive statistics are in Table 2. Two aspects of the data are 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).[6]

First, I define

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

where $I(a_{ij})$ 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 country j are observed. I operationalize m_j in terms of its empirical counterpart m_{es}, which varies by exporter (e) and sector (s). To do this, I use UN Comtrade data to produce counts of the number of export markets served by each country in each 2-digit ISIC sector.[7]

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 simple average distance of each exporter from the rest of the world. (Results are also robust to using a GDP-weighted average of distance.) The second dimension of trade costs captured here is effectively applied tariffs (i.e., including preferences) of overseas export markets. These are sourced from the TRAINS database for the year 2005, and aggregated to the ISIC 2-digit level using trade-weighted averages.

In addition to distance and 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. 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.

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. It covers elements of variable costs (transportation and handling charges), and fixed costs (standardized document preparation). 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 $1,278 per container (excluding OECD and EU countries). I scale the data by expressing them as a percentage of per capita income, to take account of the fact that the same dollar amount can represent a vastly different business constraint in rich and poor countries.

(b) Empirical model and baseline estimation results

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. Poisson is an ideal workhorse model, since it provides consistent estimates even if the data are not distributed as Poisson (see e.g., Santos Silva & Tenreyro, 2006). The most common alternative, the negative binomial model, does not have this property; however, it can give more efficient estimates if the data are in fact distributed as a negative binomial, and the results presented here are robust to use of the alternative estimator (available on request).

The mean and variance of the Poisson process are equal to μ_{es}, and its density conditional on a set of independent variables $\mathbf{X}_{es}$ is given by:

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

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

$$\mu_{es} = \delta_s \exp\begin{bmatrix} \beta_1 \frac{cost_e}{gdppc} + \beta_2 \ln(dist_e) + \beta_3 \ln(1 + t_{es}) + \ldots \\ \ldots + \beta_4 \ln(gdp_e) + \beta_5 \ln(gdppc_e) \end{bmatrix} \tag{8}$$

Export costs, distance, and foreign market 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. Since export costs are expressed relative to per capita income, and all other independent variables are in logarithms, the estimated parameters can all be interpreted as elasticities. The sector fixed effects δ_s control for unobservables that impact all exporters in a given sector in the same way. Important examples of such factors include sector-specific technology, and worldwide sectoral demand. 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 precision, all coefficients are statistically significant at the 1% level, except for tariffs (5%) and GDP per capita (15%). Since per capita income is a noisy measure of country technology, it is not too surprising that its coefficient should be insignificant even though it carries the correct sign.

Thus far, the data tend to support the core contentions of this paper. Concretely, 10% reductions in international transport costs, and importer tariffs are associated with increases of 2%, and 5% respectively in the number of export destinations. A 10% increase in the size of the domestic market is associated with increased geographical diversification of 3%. The elasticity of export destinations with respect to Doing Business export costs is smaller in absolute value than for distance or tariffs: a reduction of 10 percentage points in the ratio of export costs to per capita income is associated with a nearly 1.5% increase in the number of foreign markets served.

How important are trade costs for geographical export diversification in a quantitative sense? To examine this question, consider one standard deviation decreases in each of the three trade cost factors independently, i.e. changing one variable but keeping all others constant. Evaluated at the sample mean, one standard deviation falls in transport costs (distance) and overseas tariffs increase the number of export markets by 4% and 3.5% respectively. A similar reduction in export costs relative to per capita GDP increases geographical diversification by more than 12%. These results line up well with the trade facilitation literature, in which measures that reduce non-tariff trade costs are usually found to have bigger trade impacts than tariff cuts (see e.g., Hertel & Keeney, 2006).

Another way of looking at the impact of lower export costs is in terms of their absolute US dollar level. Keeping per capita income constant, reducing the US dollar cost of exporting in Tajikistan (the highest cost market, $4,300) to the level of the median country (St. Lucia, $1,053) would be associated with an increase of nearly 40% in the number of foreign markets entered.

The Doing Business data make it possible to zoom in on particular sources of export costs, to see where the strongest diversification effects come from. Four categories can be separately identified: customs;

documentation; inland transport; and port costs. Column 2 of Table 3 shows that each of these types of costs indeed has a different effect on diversification. The largest impact comes from customs costs, with an estimated coefficient of -0.7. Documentation and inland transport costs have estimated coefficients of -0.2 and -0.1 respectively. All three are 1% statistically significant. The only surprising result is for port costs, which have an unexpected positive and 1% significant coefficient. One possible explanation for this result is that the US dollar cost of port services is not adjusted for the quality of service provided; i.e., higher costs might in part indicate better port logistics, which would tend to promote exports. Customs and documentation costs, by contrast, do not have so much of a quality component: higher costs are much more likely to indicate only a higher burden on exporting firms.

From a policy point of view, it is important to test whether the associations set out above are in fact causal in nature. Endogeneity is potentially a serious issue in these data. For example, countries entering more export markets, and exporting more, have an incentive to reduce the costs facing their exporters.[8] The effect of this dynamic would be to bias the coefficient on export costs towards zero, thereby making it harder to reject the null hypothesis that export costs have no impact on the number of export destinations. It is therefore unlikely that endogeneity would in this case result in drawing an inference that is overly favorable to the paper's contentions.

Nonetheless, I investigate this issue further using an instrumental variables strategy. Although an IV Poisson estimator is available (Mullahy, 1997), it proved to have poor stability and convergence properties with these data, most likely due to the presence of fixed effects. As a second best, I convert the dependent variable to $\ln(0.001 + destinations)$ and use standard linear regression techniques instead of Poisson. To show that this approach only induces minimal bias, column 3 presents results from a simple OLS regression analogous to the baseline Poisson model, but with the modified dependent variable. The OLS coefficient on export costs is -0.112*** compared with -0.139*** under the baseline--a difference that is not statistically significant.

Columns 4-5 present instrumental variables results using a standard two-step GMM estimator, with aggregate and per capita income lagged by five years as an additional precaution against endogeneity.[9] The instruments for export costs are Doing Business market entry costs in the exporting country--the cost of starting a business as a percentage of per capita income--and the number of documents required to complete an import transaction in the exporting country. Both instruments should be positively correlated with export costs, since the first proxies the general level of business costs, which should be reflected in the costs facing exporters, and the second proxies the general level of trade-related bureaucracy that is not directly related to exports. The first stage regression results in column 5 show that this is indeed the case, and that the instruments are strong enough to pass the first instrument validity test: both have positive and 1% significant coefficients, and a joint F(2,30) test of 1718.324***.

In line with the contention in the previous paragraph, the second-stage GMM results (Table 3, column 5) disclose a larger absolute value coefficient on export costs than under the baseline. Endogeneity indeed appears to bias the export costs coefficient towards zero. The difference in coefficients is non-negligible (baseline = -0.139***, IV = -0.361***), and a Durbin-Wu-Hausman endogeneity test confirms that it is indeed a significant issue in these data (chi2-1 = 4.621**).

The second condition for instrument validity is excludability from the main (second-stage) regression. This condition is violated if the instruments affect the number of export destinations other than through their relationship to export costs, or if they are not genuinely exogenous to the model. Hansen's J test does not reject the null hypothesis, thereby confirming that the instruments satisfy these conditions (chi2-1 = 0.034, prob. = 0.855).

Together, these results support three conclusions. First, entry costs and the number of import documents are strong instruments for the level of export costs. Second, they are valid instruments since they are both strong and excludable from the main regression. Third, endogeneity leads to noticeable bias in the uncorrected estimates, but correcting for it makes the main results stronger—the coefficient on export costs is larger in absolute value in the IV specifications than in the baseline model.

(c) Results with additional control variables

Apart from the possible impact of endogeneity, another important aspect of identification is the exclusion of other 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, Table 4 includes a variety of alternative specifications with additional controls for exporter characteristics.

In column 1, I include Doing Business export time as an additional measure of indirect export costs. Djankov et al. (2009) find that export time is negatively correlated with intensive margin trade, and the results presented here suggest that the same holds true of the geographical extensive margin. Export costs and export time both carry negative and statistically significant (1%) coefficients, suggesting that both factors can have important impacts on geographical diversification of exports.

Column 2 includes the square of per capita income, to allow for the type of nonlinear relation between income and sectoral diversification in industrial production (not trade) found by Imbs and Wacziarg (2003). The coefficient on export costs remains close to the baseline in terms of sign and significance, and the same is true of the other baseline variables. In line with the results of Imbs and Wacziarg (2003), per capita income in levels carries a positive coefficient, but the squared term has a negative one; both are 1% significant. These results suggest that a higher income level is associated with greater diversification for relatively poor countries, but that the effect is reversed at high levels of income. The turning point is at around $16,000 or approximately the income level of Hungary, which is broadly comparable to the various results reported by Imbs and Wacziarg (2003). From that point onwards, specialization appears to dominate diversification.

In line with the gravity model literature, columns 3-5 include controls for geographical and historical factors that have been widely found to influence trade costs. Column 3 includes a dummy for landlocked countries. Column 4 includes dummies for countries where one of the official languages can be considered ''international'', i.e. Spanish, French, English, Portuguese, or Russian. Column 5 includes dummies for countries that were colonized by the major colonial powers, namely Spain, France, Great Britain, Portugal, and Russia. In all three columns, the coefficient on export costs remains negative, and highly statistically significant. All other variables of interest also carry the expected signs, and are statistically significant (1%).

Two additional factors that might also affect diversification are governance and factor intensities. Francois & Manchin (2007) find evidence that stronger institutions can be trade promoting at the intensive and extensive margins. In standard trade theories, factor abundance obviously exerts a strong influence a country's industrial structure, and the sectoral composition of its trade. Column 6 proxies institutional development using the government effectiveness indicator from Kaufmann et al. (2008). Following Romalis (2004), Column 7 includes data on factor abundance taken from Hall & Jones (1999) (capital intensity of production, and human capital) and the World Development Indicators (land to labor ratio). In both cases, results remain close to the baseline in terms of sign and significance: in particular, export costs continue to have a negative and 1% significant coefficient. The only exception is the per capita GDP coefficient in column 6, which is unexpectedly negative and significant. A likely explanation for this is the very strong correlation between per capita income and governance ($\rho = 0.85$).

Column 8 of Table 4 pushes the additional controls strategy to its limits. The regression includes all controls from the seven previous regressions, i.e. 22 variables in addition to the sector fixed effects. Results are remarkably robust to this approach. All coefficients of interest have the expected signs, and are at least 5% significant. The magnitude of the export costs coefficient is noticeably smaller than in the baseline specification, but is also less precisely estimated due to the large number of variables included in the model. Nonetheless, it remains negative and statistically significant.

The lack of a true panel dataset makes it impossible to control for country fixed effects, which would be the ideal way to ensure that omitted country variables are not influencing the results. A second best approach is to use a mixed effects Poisson model with fixed effects by sector, and random effects by country. This model controls for all unobserved heterogeneity at the sector level, as in the baseline, and in addition captures some degree of unobserved heterogeneity at the country level. Of course, the country-level specification is more restrictive than fixed effects, since the unobserved heterogeneity is assumed to follow a normal distribution. But this is the best that can be done given the current state of the data.

Column 9's results once again confirm the robustness of the baseline model. Almost all coefficients of interest retain the expected signs, similar magnitudes to the baseline, and are statistically significant at the 5% level or better. The exception is GDP per capita, which is correctly signed but, as in the baseline, statistically insignificant. Combining results from the mixed effects model with the previous columns of Table 4 suggests that it is unlikely that country-level variables external to the baseline model are driving the observed association between export costs and geographical diversification.

(d) Results with aggregate data

The baseline model is estimated at the level of ISIC 2-digit sectors partly in order to obtain meaningful results on tariffs, and partly to facilitate the examination of cross-sectoral heterogeneity (see below).

The tariff rate faced by the exporting country is the only variable that varies at the exporter-sector level. Even though all standard errors are corrected for clustering at the sector level, there is still the possibility that using a dataset that primarily varies at the country, not sectoral, level might lead to erroneous inference due to biased standard error estimates. To deal with this possibility, Table 5 presents results using aggregate data, i.e. one observation per country. The dependent variable is now the total number of export destinations served by a country, aggregating over all sectors. Due to the greatly decreased number of observations, the model is estimated using data for all countries, not just developing countries.

Results in Table 5 show that the paper's core results are highly robust to re-estimation at the aggregate level. Column 1 contains the baseline specification. Although the coefficient on export costs is smaller than in the disaggregated specification, it remains negative and 1% significant. Distance is also negative and 10% significant, while GDP is positive and 1% significant. The tariff coefficient carries the expected negative sign, but is statistically insignificant. Per capita income has an unexpected negative sign, but the coefficient is statistically insignificant.

The remaining columns of Table 5 reproduce the regressions discussed in the previous section, in which additional control variables are introduced in groups. All coefficients except for per capita income retain signs and magnitudes that are close to the baseline. Export costs and distance have negative and statistically significant coefficients in all but one specification, while GDP has a positive and 1% significant coefficient in all specifications. The only model in which export costs have a statistically insignificant coefficient is the very last one, in which all control variables are entered simultaneously. It is important to keep that result in perspective, however, since including 22 independent variables in a model with only 109 observations must inevitably lead to imprecise estimates due to correlation among the right hand side variables. The fact that the coefficient retains the expected sign and is not statistically significantly different from those in the previous estimates suggests that the baseline results are not seriously called into question by this final regression.

(e) Results with alternative samples

Having established that the model provides highly robust results on the determinants of geographical export diversification, it is useful to examine the extent to which these factors operate differently in particular country and sectoral settings. The first three columns of Table 6 estimate the model separately for different country samples. Column 1 uses data for all countries, developing and developed. Column 2 includes lower middle and low income countries only. Column 3 is limited to low income countries. The estimated coefficient on tariffs is noticeably larger in absolute value in columns 2 and 3, which indicates that tariffs are more of a constraint on geographical diversification in poorer countries. The distance coefficient follows the opposite pattern, however; in the low income sample, it even turns positive and statistically significant. The reason for this behavior is surely the role of trade preferences: South-South trade is characterized by substantially higher tariffs than those available under preference regimes such as the European Union's ''Everything but Arms'' program. As a result, poorer countries in many cases have a stronger incentive to trade with Northern economies, which tend to be more distant geographically, than with their Southern neighbors.

The coefficient on Doing Business export costs becomes noticeably weaker as the estimation sample becomes poorer. However, it is important to keep in mind the role of deflating these US dollar costs by per capita income. Evaluating the sensitivity of export destinations with respect to export costs in dollars at the sample average income level shows that the effect is about 17% stronger in low income countries compared with the full sample.

Since the Melitz (2003) model is based on product differentiation, it is most likely a better fit with manufactured goods sectors than it is with primary products. The degree of differentiation is relatively high in the first group, but much lower in the second. Columns 4 and 5 of Table 6 confirm that the model indeed has much greater explanatory power in respect of manufactured goods (R2 of 0.6 versus 0.2 for agriculture). The absolute value coefficient on export costs is much larger in the manufactures regression than in the primary products regression, or in the pooled results. It is statistically significant in the manufactures regression, but not in the primary products one. This contrast is important because it suggests that export costs in developing countries particularly inhibit their geographical diversification in manufactured goods exports.

Using estimates of the intra-sectoral elasticity of substitution due to Broda & Weinstein (2006), it is also possible to examine the role of product differentiation more directly.[10] Since their measures of substitutability are estimated at the HS 2-digit level, the remaining regressions in this section use data aggregated according to that scheme but limited to manufactured goods only (HS Chapter 28 and higher). I interact the three trade costs variables with dummies indicating differentiated goods sectors, so as to investigate whether the impact of trade costs is different according to the degree of substitutability of goods. Estimation uses the full country sample, in order to have maximum within-sample variation. Column 6 uses a loose definition of differentiated goods, i.e. those with an intra-sectoral elasticity of substitution less than the sample 75th percentile (approximately 12). In all three cases, the differentiated goods interaction term is negative, which suggests that the impact of trade costs is larger in more differentiated sectors. The effect is 1% significant in relation to Doing Business export costs, and is 15% significant in the case of tariffs.

Column 7 runs the same model with a stricter definition of differentiated goods, i.e. those with an intra-sectoral elasticity of substitution less than the sample median (approximately 9). Again, all three interaction effects have negative coefficients. The distance interaction term is 1% significant, and the tariffs interaction is 10% significant. For Doing Business export costs, the estimated coefficient is marginally significant at the 10% level (prob. = 0.106).

Results from the loose and strict definitions of differentiated goods consistently suggest that the negative impact of trade costs on the geographical extensive margin is stronger in the case of relatively differentiated manufactures. In other words, the elasticity of substitution provides a dampening effect. This finding is exactly in line with the predictions of the heterogeneous firms model due to Chaney (2008). As that author points out, these kinds of findings with respect to the extensive margin suggest that the trade costs factors used in the regression most likely have variable and fixed components, which was the proposition advanced above in relation to Doing Business export costs.

4. CONCLUSIONS

This paper has provided some of the first evidence on the factors driving the geographical spread of developing country trade. In the baseline econometric specification, I find that reducing international transport costs (distance), tariffs, and Doing Business export costs by one standard deviation leads to increases of 4%, 3.5%, and 12% respectively in the number of export markets served.

The data strongly suggest that export costs have heterogeneous effects on geographical diversification across countries and sectors. The link tends to be stronger in poorer countries, in manufactures versus agriculture, and in relatively differentiated manufactures versus relatively homogeneous ones. Since industrialization and movement into differentiated manufactured goods exports are closely associated with the development process, the results presented here are of particular interest from a policy point of view.

The results also have an important implication for trade policy. On the one hand, they show that market access abroad can of course help developing countries diversify their exports geographically. But they also highlight another policy option that is available, namely trade facilitation—understood as policies designed to reduce export costs at home. Trade facilitation can have a significant impact on diversification—perhaps even stronger than improved market access—and it can be pursued by developing countries unilaterally or regionally, as well as multilaterally. Although included in the WTO's Doha Round, progress on trade facilitation does not necessarily have to await progress in the broader talks. Since trade facilitation reforms are usually non-discriminatory, they have the added benefit of minimizing trade diversion and maintaining consistency with the basic rules of the multilateral system.

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. Use of a genuine panel framework is also necessary to fully account for unobserved exporter heterogeneity using country fixed effects. Second, it will be important to pay attention 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., 2008; 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. In policy terms, it will be important to accurately identify the full range of costs and benefits associated with diversification and specialization.

---

[1]The sample selection gravity model developed by Helpman et al. (2008) allows for extensive margin trade expansion in the product and market dimensions, in addition to an intensive margin. Their empirical work controls for both effects. However, their regressions aggregate the two extensive margins, in the sense that identification of

different effects at the two margins is not possible. Reworking their empirics to make separate identification possible could be a direction for future research, but is outside the scope of the present paper. A similar analysis applies to the Tobit model used by Amurgo-Pacheco & Pierola (2008), which distinguishes between the extensive and intensive margin, but does not empirically identify the two extensive margins referred to here (except in their descriptive statistics section). Besedes & Prusa (2007) focus on the duration of trading relationships, not on geographical diversification as such.

[2]The recent international trade literature provides extensive evidence in support of this setup. Firm productivity levels are well approximated by the Pareto distribution, and this is partly reflected in the fact that only a very small percentage of firms--those in the upper tail of the productivity distribution--become exporters. There is also extensive evidence suggesting that firms self-select into export status based on productivity. See Bernard et al. (2007) for a review of recent research using data from developed and developing countries.

[3]Although this mechanism is intuitively appealing, Eaton et al. (2008) find that it is not a sufficient explanation for the pattern of exporting behavior of French exporting firms. Lawless & Whelan (2008) report similar limitations 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 assumed here. However, an expansion of the canonical heterogeneous firms model in this direction is outside the scope of this paper.

[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]Although the export cost data discussed below are now available for a number of years, there are two obstacles to conducting a panel data analysis. First, trade and tariff data are only available for many developing countries with a

significant lag, making 2007 or 2008 the latest practicable year for analysis. Second, over that time period (2005-2008), the Doing Business indicators exhibit almost no systematic temporal variation. A true panel data analysis will need to await the availability of further data.

[7]Baseline results are very similar with an HS 2-digit aggregation scheme. However, the lack of homogeneity in HS chapters results in generally insignificant tariff coefficients. The ISIC classification is considerably more homogeneous, and thus generates more consistent and meaningful results.

[8]By contrast, the other trade cost factors—distance and foreign tariffs—are not under the control of the exporting country, and thus cannot give rise to endogeneity concerns.

[9]See Davidson & MacKinnon (2004), Chapter 9 on GMM estimation. This estimator is preferred to two-stage least squares because it provides more efficient estimates when there are more instruments than potentially endogenous variables, as is the case here.

[10]The regressions presented here use the intra-sectoral elasticities of substitution for the USA, since they provide improved data coverage compared with the country specific estimates in Broda & Weinstein (2006).

REFERENCES

Amurgo-Pacheco, A., & Pierola, M. (2008). Patterns of export diversification in developing countries: Intensive and extensive margins. Policy research working paper no. 4473, The World Bank.

Anderson, J., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *The American economic review*, *93*, (1), 170-192.

Baldwin, R., & Harrigan, J. (2007). Zeros, quality, and space: Trade theory and trade evidence. Working paper No. 13214, NBER.

Bernard, A., Jensen, J., Redding, S., & Schott, P. (2007). Firms in international trade. *Journal of economic perspectives, 21,* (3), 105-130.

Besedes, T., & Prusa, T. (2007). The role of extensive and intensive margins and export growth. Working paper no. 13628, NBER.

Brenton, P., & Newfarmer, R. (2007). Watching more than the discovery channel: Export cycles and diversification in development. Policy research working paper no. 4302, The World Bank.

Broda, C., & Weinstein, D. (2006). Globalization and the gains from variety. *The quarterly journal of economics*, *121*, (2), 541-585.

Brainard, W., & Cooper, R. (1965). Uncertainty and diversification in international trade. Discussion paper No. 197, Cowles Foundation.

Cadot, O., Carrère, C., & Strauss-Kahn, V. (2007). Export diversification: What's behind the hump?". Discussion paper No. 6590, CEPR.

Cameron, A., & Trivedi, P. (2001). Essentials of count data regression. In B. Baltagi (Ed.), *A companion to theoretical econometrics.* Malden, Ma.: Blackwell.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *The American economic review*, *98*, (4), 1707-1721.

Davidson, R., & MacKinnon, J. (2004). *Econometric theory and methods*. New York: Oxford University Press.

Djankov, S., Freund, C., & Pham, C. (Forthcoming). Trading on time. *Review of Economics and Statistics*.

Eaton, J., Kortum, S., & Kramarz, F. (2008). An anatomy of international trade: Evidence from French firms. Working paper, http://www.econ.nyu.edu/user/eatonj/EKK.pdf.

Evenett, S, & Venables, A. (2002). Export growth in developing countries: Market entry and bilateral trade flows. Working paper, http://www.evenett.com/working/setvend.pdf.

Francois, J., & Manchin, M. (2007). Institutions, infrastructure, and trade. Policy research working paper no. 4152, World Bank.

Hall, R., & Jones, C. (1999). Why do some countries produce so much more output per worker than others? *The quarterly journal of economics, 114*, (1), 83-116.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics*, *123*, (2), 441-487.

Hertel, T., & Keeney, R. (2005). What's at stake? The relative importance of import barriers, export subsidies, and domestic support. In K. Anderson, & W. Martin (Eds.), *Agricultural trade reform and the Doha Development Agenda* (pp. 37-62). Washington, D.C.: The World Bank.

Hummels, D., & Klenow, P. (2005). The variety and quality of a nation's exports. *The American economic review*, *95*, (3), 704-723.

Imbs, J., Wacziarg, R. (2003). The stages of diversification. *The American economic review*, *93*, (1), 63-86.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2008). Governance matters vii: Aggregate and individual governance indicators, 1996-2007. Policy research working paper no. 4654, The World Bank.

Koenig, P. (2009). Agglomeration and the export decision of French firms. *Journal of urban economics, 66,* 186-195.

Krugman, P. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of international economics, 9*, 469-479.

Lawless, M, & Whelan, K. (2008). Where do firms export, how much, and why? Working paper no. 200821, University College, Dublin.

Mayer, T., & Zignago, S. (2006). Notes on CEPII's distance measures. Working paper, http://www.cepii.fr/distance/noticedist_en.pdf.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*, (6), 1695-1725.

Mullahy, J. (1997). Instrumental variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of economics and statistics*, *79*, (4), 586-593.

Romalis, J. (2004). Factor proportions and the structure of commodity trade. *The American economic review, 94,* (1), 67-97.

Santos Silva, J., & Tenreyro, S. (2006). The log of gravity. *Review of economics and statistics*, *88*, (4), 641-658.

Schott, P. (2004). Across-product versus within-product specialization in international trade. *The quarterly journal of economics*, 119(2), 646-677.

APPENDIX: COMPARATIVE STATICS

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):

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

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

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

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

where in addition to the variables defined in the main text: P_i is a standard CES price index, aggregating over the set of varieties V_{ij}, as defined by the second and third equations above; $G(a)$ is the CDF of the Pareto productivity distribution, a defined by the fourth equation; and α is related to the intra-sectoral elasticity of substitution ε by $\varepsilon = \frac{1}{1-\alpha}$. With these definitions, the first condition is the zero profit marginal cost cutoff for the country pair $\{i, j\}$. 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.

(a) Variable Trade Costs

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

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

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

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:

$$V_{ij} = \frac{k}{a_H^k - a_L^k}\int_{a_L}^{a_{ij}} a^{k-\varepsilon}\,da \quad (15)$$

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

(b) Fixed Trade Costs

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

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

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

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

(c) Home Market Technology

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

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

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

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.

(d) 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.

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 ISIC 2-digit sector.	2005	WITS-Comtrade
Differentiated	Dummy variable equal to unity for differentiated manufactured goods, defined as: 1) goods with an elasticity of substitution less than the manufacturing median; or 2) goods with an elasticity of substitution less than the manufacturing 25th percentile.	1990-2001	Broda & Weinstein (2006)
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.	2005	Doing Business
Export Time	Time required to comply with all official procedures required to export goods.	2005	Doing Business
Factor Intensities	Physical capital to output ratio, human capital per worker (Hall & Jones, 1999), and land to labor ratio (WDI).	1999/2005	Hall & Jones (1999); World Development Indicators
GDP	Gross domestic product, current USD.	2000, 2005	World Development Indicators
GDPPC	GDP per capita, current USD.	2000, 2005	World Development Indicators
Governance	Government effectiveness indicator, rescaled to min. = 1.	2005	World Governance Indicators
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.	2005	Doing Business
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
Tariffs	Trade-weighted average applied ad valorem tariff in the rest of the world, by ISIC 2-digit sector.	2005	WITS-Trains

Table 2: Descriptive statistics, main variables only.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.	Correlation with Destinations
Destinations	7006	39.62	44.55	0.00	158.00	1.00
Between	*31*		19.67	0.14	62.06	
Within	*226*		40.13	-22.44	161.85	
Export Cost	4681	0.52	0.83	0.01	6.20	-0.36
Ln(Distance)	6293	9.01	0.20	8.76	9.50	-0.08
Ln(GDP)	5549	23.24	2.42	17.69	30.03	0.73
Ln(GDPPC)	4743	8.52	1.12	6.37	10.53	0.55
Ln(1+Tariff)	6217	0.04	0.06	0.00	0.92	0.01
Between	*31*		0.03	0.00	0.15	
Within	*200.548 (ave.)*		0.05	-0.11	0.80	

a. The variable Destinations includes 764 observations equal to zero, i.e. approximately 10% of the sample.

Table 3: Estimation results—baseline model and instrumental variables.

	(1)	(2)	(3)	(4)	(5)
	Poisson	Poisson	OLS	GMM	
	Baseline	*Breakdown*	*Baseline*	*2nd Stage*	*1st Stage*
Export cost/GDPPC	-0.139***		-0.112***	-0.560***	
	(0.020)		(0.028)	(0.216)	
Log(distance)	-0.209***	-0.238***	-0.367**	-0.573***	-0.205***
	(0.058)	(0.057)	(0.159)	(0.152)	(0.019)
Log(1+tariff)	-0.519**	-0.540**	-0.510	-0.424	0.217
	(0.244)	(0.244)	(0.562)	(0.486)	(0.290)
Log(GDP)	0.292***	0.295***	0.399***	0.378***	-0.062***
	(0.009)	(0.009)	(0.021)	(0.026)	(0.001)
Log(GDPPC)	0.033	0.035	0.064	-0.177	-0.408***
	(0.022)	(0.022)	(0.039)	(0.111)	(0.004)
Customs cost/GDPPC		-0.733***			
		(0.055)			
Documents cost/GDPPC		-0.224***			
		(0.033)			
Transport cost/GDPPC		-0.141***			
		(0.022)			
Port cost/GDPPC		0.297***			
		(0.036)			
Entry cost/GDPPC					0.096***
					(0.002)
Import documents					0.022***
					(0.001)
Observations	3310	3369	3310	3310	3310
Groups	31	31	31	31	31
R2	0.456	0.451	0.379	0.338	0.459
Instrument F-Test (2,30)					1734.60***
Hansen's J (chi2-1)				0.034	
Fixed effects	Sector	Sector	Sector	Sector	Sector

b. The dependent variable in columns 1, 2, and 4 is Destinations. In column 3 it is ln(0.001+Destinations). In column 5 it is Export Cost/GDPPC. GDP and GDPPC are lagged by 5 periods in columns 4-5.

c. Robust standard errors corrected for clustering by ISIC 2-digit sector are in parentheses. Statistical significance is indicated by * (10%), ** (5%), and *** (1%).

d. R2 in columns 1-2 is calculated as the squared correlation between Destinations and the fitted values from each regression.

Table 4: Robustness checks using additional control variables.

	(1) Poisson	(2) Poisson	(3) Poisson	(4) Poisson	(5) Poisson	(6) Poisson	(7) Poisson	(8) Poisson	(9) Poisson
Export cost / GDPPC	-0.215***	-0.154***	-0.192***	-0.206***	-0.215***	-0.254***	-0.180***	-0.031**	-0.100**
	(0.024)	(0.020)	(0.023)	(0.022)	(0.023)	(0.025)	(0.020)	(0.014)	(0.043)
Log(distance)	-0.308***	-0.297***	-0.297***	-0.211***	-0.158***	-0.306***	-0.120***	-0.166***	-0.316**
	(0.036)	(0.039)	(0.036)	(0.034)	(0.033)	(0.038)	(0.034)	(0.035)	(0.135)
Log(1+tariff)	-0.755***	-0.620***	-0.686***	-0.852***	-1.097***	-0.880***	-0.823***	-1.284***	-1.067***
	(0.221)	(0.203)	(0.206)	(0.231)	(0.251)	(0.247)	(0.282)	(0.373)	(0.072)
Log(GDP)	0.242***	0.243***	0.237***	0.239***	0.245***	0.241***	0.211***	0.226***	0.297***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.013)
Log(GDPPC)	-0.005	0.563***	0.043***	0.035**	0.016	-0.134***	0.062***	1.572***	0.058
	(0.016)	(0.147)	(0.015)	(0.014)	(0.014)	(0.015)	(0.016)	(0.148)	(0.036)
Observations	4325	4325	4325	4325	4325	4301	3150	3150	4325
Groups	31	31	31	31	31	31	31	31	31
R2	0.523	0.520	0.522	0.522	0.524	0.519	0.513	0.526	0.917
Fixed Effects	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector
Random Effects									Country
Additional Controls	Export Time	$GDPPC^2$	Landlocked	Language	Colonies	Governance	Factor Intensities	All	

a. The dependent variable in all cases is Destinations.

b. Robust standard errors corrected for clustering by ISIC 2-digit sector are in parentheses. Statistical significance is indicated by * (10%), ** (5%), and *** (1%).

c. R2 is calculated as the squared correlation between Destinations and the fitted values from each regression.

Table 5: Robustness checks using aggregate data.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Export cost / GDPPC	-0.054***	-0.054***	-0.037*	-0.040**	-0.049***	-0.052***	-0.060***	-0.056***	-0.011
	(0.018)	(0.017)	(0.020)	(0.019)	(0.016)	(0.018)	(0.018)	(0.019)	(0.025)
Log(distance)	-0.113*	-0.120*	-0.122*	-0.125*	-0.152*	-0.169*	-0.102	-0.039	-0.119
	(0.066)	(0.066)	(0.067)	(0.066)	(0.084)	(0.089)	(0.064)	(0.062)	(0.083)
Log(1+tariff)	-0.004	-0.061	0.113	-0.018	-0.025	0.068	-0.179	-0.499	-0.523
	(0.505)	(0.510)	(0.515)	(0.522)	(0.453)	(0.446)	(0.498)	(0.552)	(0.462)
Log(GDP)	0.094***	0.095***	0.096***	0.093***	0.094***	0.092***	0.093***	0.073***	0.070***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	(0.009)
Log(GDPPC)	-0.017	-0.026	0.194	-0.014	-0.017	-0.018	-0.051**	-0.007	0.418*
	(0.015)	(0.020)	(0.207)	(0.015)	(0.015)	(0.017)	(0.023)	(0.022)	(0.227)
Observations	151	151	151	151	151	151	150	109	109
R2	0.749	0.750	0.751	0.754	0.758	0.761	0.755	0.745	0.810
Additional Controls		Export Time	$GDPPC^2$	Landlocked	Language	Colonies	Governance	Factor Intensities	All

a. The dependent variable in all cases is Destinations, aggregated over all ISIC 2-digit sectors; i.e., it is the total number of foreign markets to which an exporting country ships at least one product line.

b. Robust standard errors are in parentheses. Statistical significance is indicated by * (10%), ** (5%), and *** (1%).

c. R2 is calculated as the squared correlation between Destinations and the fitted values from each regression.

Table 6: Robustness checks using alternative samples.

	(1) Poisson	(2) Poisson	(3) Poisson	(4) Poisson	(5) Poisson	(6) Poisson	(7) Poisson
Export cost/GDPPC	-0.213***	-0.071***	-0.037**	-0.160***	-0.030	-0.157***	-0.202***
	(0.023)	(0.018)	(0.015)	(0.016)	(0.022)	(0.055)	(0.043)
Export Cost/GDPPC *						-0.142*	-0.115
Differentiated						(0.080)	(0.071)
Log(distance)	-0.283***	-0.120*	0.581**	-0.240***	0.010	-0.395***	-0.348***
	(0.037)	(0.066)	(0.234)	(0.052)	(0.249)	(0.037)	(0.025)
Log(distance) *						-0.003	-0.098**
Differentiated						(0.045)	(0.042)
Log(1+tariff)	-0.682***	-1.559***	-0.858***	-0.212	-1.946***	0.363	0.202
	(0.207)	(0.344)	(0.329)	(0.166)	(0.561)	(0.357)	(0.247)
Log(1+tariff) *						-0.631	-0.658*
Differentiated						(0.422)	(0.357)
Log(GDP)	0.240***	0.318***	0.355***	0.286***	0.342***	0.284***	0.284***
	(0.008)	(0.009)	(0.013)	(0.009)	(0.034)	(0.009)	(0.009)
Log(GDPPC)	0.036**	0.086***	0.039	0.052**	-0.099**	0.052***	0.053***
	(0.014)	(0.026)	(0.033)	(0.021)	(0.047)	(0.012)	(0.011)
Observations	4325	2215	964	2547	707	9640	9640
Groups	31	31	31	31	31	69	69
R2	0.519	0.476	0.291	0.624	0.209	0.569	0.440
Country Sample	All	Lower Middle + Low Income	Low Income	No High Income	No High Income	All	All
Sector Sample	All	All	All	Manufacturing	Primary Industry	Manufacturing	Manufacturing

a. The dependent variable in all cases is Destinations.

b. Robust standard errors corrected for clustering by ISIC 2-digit sector (columns 1-5) or HS 2-digit sector (columns 6-7) are in parentheses. Statistical significance is indicated by * (10%), ** (5%), and *** (1%).

c. R2 is calculated as the squared correlation between Destinations and the fitted values from each regression.

d. Differentiated products are those with a substitution elasticity below the manufacturing 75th percentile (column 6) or median (column 7).