Big Hits in Manufacturing Exports and Development*

William Easterly[†]
New York University

Ariell Reshef
University of Virginia

October 2009

PRELIMINARY. COMMENTS WELCOME

Abstract

Economic development is strongly correlated with success at exporting manufactures. What is the nature of this success? We systematically document remarkably high degrees of concentration in manufacturing exports for a sample of 151 countries over a range of 2,850 manufactured products. Manufacturing exports are dominated by a few "big hits", which account for most of export value. The per capita value of the top 3 product-destination export flows has a remarkably high correlation with income per capita (0.81 in logs). Overall export success is associated with higher degrees of concentration, after controlling for the number of export flows. This further highlights the importance of big hits. The distribution of exports closely follows a power law, especially in the upper tail. These findings do not support a "picking winners" policy for export development; the power law characterization implies that the chance of picking a winner diminishes exponentially with the degree of success. Moreover, we find that on average demand shocks are almost as important to the variation in trade revenues as technological dispersion and trade barriers combined, which further lowers the benefits from trying to pick winners.

*We wish to thank Peter Debare, Wayne-Roy Gale, Steven Stern, Jorg Stoye and Michael Waugh for useful comments and discussions. We thank Shushanik Hakobyan for excellent research assistance.

[†]NBER.

How do countries succeed at economic development? Many descriptions of success stories have stressed the important role of manufacturing exports as a vehicle for success. Indeed, manufacturing exports per capita have a striking correlation with per capita income across countries with a correlation of 0.88 (in log values), see **Panel A of Figure 1**. Causality could go either way in this association, or both variables may reflect other factors. The figure does support a descriptive statement that success at manufacturing exports and success at development are closely interrelated. This warrants a close examination of the characteristics of success in manufacturing exports.

The manufacturing exports include 2,985 possible products at the 6-digit level (HS1992). We also explore the patterns of exports by importing country – there are 217 possible destinations in our dataset. Hence, in theory, there are 647,745 possible product-destination combinations. **Panel B of Figure 1** depicts the relationship between the value of the top three product-destination export flows per capita and income. The correlation with income per capita is remarkably high: 0.81 (in log values). And it is remarkably similar to Panel A, which illustrates the main point we stress in this paper. A huge amount of trade value is concentrated in very few product-destination export flows – "big hits" – and they are related to development in the very same way to income as overall exports. Studying export success means studying the big hits.

In this paper we systematically document that manufacturing export success is characterized by a remarkable degree of specialization for virtually all countries. Manufacturing exports in each country are dominated by a few big hits, which account for most of export value and where the "hit" includes finding both the right product and the right market. Moreover, we show that higher export volumes are associated with higher degrees of concentration, after controlling for the number of destinations a country penetrates (the latter reflects absolute advantage and size of country). This highlights the importance of big hits. In addition, we estimate that most of the variation, and hence concentration, in export is driven by technological dispersion of the exporting country, rather than demand shocks from the importing destinations. However, given the size of the economy, developing countries are more exposed to demand shocks than rich ones.

Hausmann and Rodrik (2006), in a seminal paper which helped inspire this one, had previously pointed out the phenomenon of hyper-specialization, although only for a few countries and products, and not including the destination component, in contrast to the comprehensive scope of our work. We also make a very significant addition to the Hausmann and Rodrik findings, in that we characterize the probability of "big hits" as a function of the size of the hit – by a power law.

We specify a “hit” as a product-by-destination export flow. We chose this categorization because some export products are shipped to several destinations, while the typical export product is shipped to few destinations (with a mode of one). A few examples of big hits and their relationship to concentration illustrate the nature of a "big hit". Egypt gets 23 percent of its total manufacturing exports from exporting one product – “Ceramic bathroom kitchen sanitary items not porcelain” – to one destination, Italy, capturing 94 percent of the Italian import market for that product. Fiji gets 14 percent of its manufacturing exports from exporting “Womens, girls suits, of cotton, not knit” to the U.S., where it captures 42 percent of U.S. imports of that product. The Philippines get 10 percent of their manufacturing exports from sending “Electronic integrated circuits/microassemblies, nes” to the U.S. (80 percent of U.S. imports of that product). Nigeria earns 10 percent of its manufacturing exports from shipping “Floating docks, special function vessels nes” to Norway, making up 84 percent of Norwegian imports of that product.

Examining big hits that are exported almost exclusively to one destination for what one would think would be fairly similar countries reveals a surprising diversity of products and destinations. Why does Colombia export paint pigment to the U.S., but Costa Rica exports data processing equipment, and Peru exports T-shirts? Why does Guatemala export candles to the U.S., but El Salvador exports toilet and kitchen linens? Why does Honduras export soap to El Salvador, while Nicaragua exports bathroom porcelain to Costa Rica? Why does Cote d’Ivoire export perfume to Ghana, while Ghana exports plastic tables and kitchen ware to Togo? Why does Uganda export electro-diagnostic apparatus to India, while Malawi exports small motorcycle engines to Japan?

The high specialization across products and destinations shows up in high concentration ratios. The top 1 percent of nonzero product-destination pairs account for an average of 52 percent of manufacturing export value for 151 countries on which we have data.[1]

[1] At this point we do not analyze specialization (concentration) along the time dimension. One attempt to do so is Imbs and Wacziarg (2003). However, they address specialization in total production, not exports, and, hence, do not analyze the destination dimension, which we believe captures additional product

The difference between successful and unsuccessful exporters is found not just in the degree of specialization, but also in the scale of the "big hits." For example, a significant part of South Korea's greater success than Tanzania as a manufacturing exporter is exemplified by South Korea earning \$13 billion from its top 3 manufacturing exports, while Tanzania earned only \$4 million from its top 3.

The probability of finding a big hit ex ante decreases exponentially with the magnitude of the hit. We show that the upper part of the distribution of export value across products (defined both by destination and by six-digit industry classifications) is close to following a power law.[2] On average across our sample, the value of the top ranked product-destination export category is 13.5 times larger than that of the 10th ranked product-destination export category (the corresponding median ratio is lower, only 5 times, because of the skewness of this statistic in our sample). The value of top ranked product-destination export category is on average 1,064 times (median 48 times) larger than the 100th ranked product-destination export category. In this paper we will estimate just how much the entire distribution of export values within each country is explained by a power law, and will place it in context of a trade model with demand and productivity shocks.[3]

Realizing that export success is driven by a few big hits changes our understanding of "success" and poses challenges for economic policy. Power laws may arise because many conditions have to be satisfied for a "big hit," and hence the probability of success is given by multiplying the probability of each condition being satisfied times each other (if probabilities are independent). Source country s's success at exporting product p to destination country d depends on industry-specific and country-specific productivity factors in country s, the transport and relational connections between s and d for product p, and the strength of destination country d's demand for product p from country s. All of these components

differentiation.

[2] Pareto distributions follow a so-called "power law", in which the probability of observing a particular value decreases exponentially with the size of that value. The distributions of word frequencies (Zipf's law), sizes of cities, citations of scientific papers, web hits, copies of books sold, earthquakes, forest fires, solar flares, moon craters and personal wealth all appear to follow power laws; see Newman (2005). See also Table 1 in Andriani and McKelvey (2005) for more examples. Describing concentrated distributions in economics has a long tradition, starting with Pareto (1896). Sutton (1997) provides a survey of the literature on the size distribution of firms starting with the observation of proportional growth by Gibrat (1931) (Gibrat's law).

[3] Luttmer (2007) constructs a general equilibrium model with firm entry and exit that yields a power law in firm size. He combines a preference and a technology shock multiplicatively to obtain a variable he refers to as the firm's total factor productivity.

are subject to shocks in country-industry technology, firms, country policy, input sectors, shipping costs and technologies, trading relationships, brand reputation, tastes, competitors, importing countries, etc.

The policy discussion about making such success more likely tends to be sharply polarized. Hausmann and Rodrik argue that a firm in country s that first succeeds at exporting product p (they do not examine the destination dimension) is making a discovery that such a product export is profitable, which then has an externality to other firms who can imitate success. They argue therefore that such a discovery process should receive a public subsidy, which may imply a conscious government industrial policy.

Our analysis raises a new issue. In addition to the possible knowledge externality to a successful export, there is also a knowledge problem about the discovery itself. Who is more likely to discover the successful product-destination category – the public or private sector? We show that success (in both the product and destination dimensions) closely follows a power law. Hence, ex ante picking a winning export category (or discoverer) would be very hard indeed. The main argument for private entrepreneurship against the government "picking winners" relies on the view of private entrepreneurship as a decentralized search process, which is characterized by many independent trials by agents who have many different kinds of specific knowledge about sectors, markets, and technologies. Thus, *a priori*, it seems that private entrepreneurship is more likely to find a "big hit" than a process relying on centralized knowledge of the state. However plausible these arguments may be, in the end it is an empirical question which approaches work. We hope to stimulate this debate in this paper, but do not believe that we can resolve it definitively.

A complementary point to ours is made by Besedes and Prusa (2008). They find that most new trade relationships fail within 2 years and that the hazard rate of such failure is higher for developing countries.[4] Nevertheless, developing countries have the highest increase in trade relationships: there seems to be a lot of attempts in discovery as it is. These two last facts together imply that there are more attempts in discovery in developing countries than in developed ones.[5] However, entry (the extensive margin) does not account for much growth in trade. All this, together with our stress on the importance and difficulty

[4] Their sample is 1975-2003 and relates to bilateral 4-digit SITC relationships.

[5] This might not be surprising, given that developed countries have already "discovered" a larger share of the pool of possible markets and products.

of discovering big hits (at a higher level of disaggregation), this implies that Hausmann and Rodrik's point might be misplaced.

Although addressing the Hausmann-Rodrik argument is our main goal, our work is related to a few other recent papers. The observation that trade is concentrated has not been lost on economists. Bernard, Jensen, Redding, and Schott (2007) document concentration across U.S. exporting firms, while Eaton, Eslava, Kugler, and Tybout (2007) find that Colombian exports are dominated by a small number of very large (and stable) exporters. Arkolakis and Muendler (2009) make a similar point for Brazilian and Chilean exporting firms and find that the distribution is approximately Pareto.

In contrast to these and other contributions, we document concentration and Pareto-like distributions for many more countries (151); we do so at the product-destination level; and we try to assess how much of this concentration is driven by technological dispersion versus demand. Eaton, Kortum, and Kramarz (2008) also relate trade patterns to productivity and demand shocks. But while they dissect trading patterns only for French firms, regardless of which products each firm exports (there could be more than one product per firm), we analyze trade at the product level for many countries.[6]

In the next section we document concentration and distributions of exports for 151 countries in the product-destination dimension and perform preliminary analysis. In section 2 we estimate the contribution of technology versus demand to the distribution and concentration of exports. Section 3 concludes.

1 Empirical facts

Our main data source is the UN Comtrade database. The U.N. classifies exported commodities and manufactured products by source and destination at the six-digit level (roughly 5,000 categories). We use the 1992 Harmonized System classification (HS1992) for the year 2000, to maximize the available bilateral trade pairs. Using a less disaggregated classification might have lead to better coverage of countries (say, 4-digit SITC), but would miss the extreme concentration within finely defined products.[7]

[6] The distribution of exports across products is similar to what they find for French firms.

[7] An analysis of the distribution of product-destination export flows at the 4-digit SITC level reveals similar patterns, but lower levels of concentration, as one might expect.

We restrict our sample to manufactured categories, i.e. we drop from the sample all agriculture and commodities exports. Our focus on manufactured products stems from our interest on exports that are not dependent on country-specific natural endowments, and could potentially be produced everywhere in the world. We basically exclude products that rely directly on natural resources. Natural resources create strong comparative advantage for extractables and agricultural products. Therefore, *a priori*, focusing on manufacturing also reduces the degree of concentration, especially for developing countries.

Some importers in the original dataset did not correspond to well-defined destinations, so we dropped those destinations from the analysis.[8] Eventually, our sample contains 151 exporters, 2984 export categories, which may be shipped to at most 217 destinations (importers).

1.1 Concentration of exports

Our first observation is that exports are highly concentrated. That is, for each country a few successful products and destination markets account for a disproportionately large share of export value. We initially examine manufactured products, while ignoring the destination market dimension (we will incorporate the destinations shortly). **Table 1** shows that the median export share of the top 1%, 10% and 20% within nonzero export products for a country is 47%, 86% and 94%, respectively.[9] In fact, for the median country, the top 3 products account for 28% of exports, and the top 10 products account for a staggering 49%. The median share for the bottom 50% of exported products is a mere 0.8 %. This implies a high degree of concentration indeed.[10]

One issue that complicates the interpretation of the concentration ratios is that countries also differ a lot in how many export products they export at all (i.e. product exports with nonzero entries for each country) – from a minimum of 10 to a maximum of 2950, with a median of 1035. We will examine the role of number of products in the next section.

Another surprising fact is just how few destination markets each product penetrates. **Figure 2** shows the average across all 151 exporters of the share of export value accounted

[8] For example, "Antarctica", "Areas, nes", "Special Categories", etc.

[9] Our basis for comparisons are always nonzero export flows for each country separately. In calculating percentages we never compare to potential export products that are exported by all countries (2984 in total).

[10] **Table A1** in the appendix reports these shares for all 151 countries in our sample.

for by products that have the number of destinations shown on the X-axis. The largest shares go to products that are exported to only one destination, the next largest share goes to products that are exported to only two destinations, and then it falls off to a long tail. This observation led to our decision to treat the product-destination pair as the unit of analysis for the bulk of our analysis.

We now incorporate the destination dimension. In all the analysis that follows, we stick to this unit of account: the product-destination export flow. We will refer to this simply as "export flow". The same observation about concentration at the product level holds for product-destination trade flows, i.e. when each observation is an export of a particular product to a particular destination. **Table 2** shows that for the median exporter the top nonzero 1% of product-destination pairs account for 52% of total export value! The top 10% account for 89% and the bottom 50% for only 0.8%.[11]

The number of nonzero entries in the product-destination matrix varies enormously across countries, and is always far below the potential number. We define the number of potential export flows for some source country s as $P_s = I_s \times 215$, where I_s is the number of exported products for source country s and 215 is the maximum number of destinations. So for source country s reaching potential means serving all destinations with the entire set of exported products, I_s. This is an appropriate concept of potential for comparing across countries that export very different sets of products. After all, we cannot expect all countries to export every product.

The median number of nonzero product-destination entries per exporter is 3,055, with a minimum of 10 and a maximum of 195,417. The median number of nonzero entries is roughly 1.5 percent (Zimbabwe) of the potential number, with a minimum of 0.5 percent (Greenland) and a maximum of 31 percent (Germany). Baldwin and Harrigan (2007) and have previously made the observation that most potential product-destination flows are absent and find that incidence of zeros is negatively correlated with distance and positively correlated with importer size (i.e. it follows "gravity"). Here we show that this is another important dimension of variation in the degree of success of exports. In the next section we systematically relate this to concentration and the prevalence of big hits.

[11] **Table A2** in the appendix shows these numbers for all countries.

1.2 Bilateral flows and potential flows

Our main focus is on the distribution of value across product-destination export flows. However, we want to first place the statistics above in context. To do so, we provide a brief descriptive analysis of export patterns and concentration ratios. We start by illustrating the very strong (log-linear) association between the number of nonzero product-destination export flows and the value of total manufacturing exports, as can be seen in **Figure 3**. One way to succeed at exporting is to export more products to more places. This is a result of absolute advantage, which allows penetrating more markets with more products. The slope of the relationship in **Figure 3** is about 1.5 (greater than unity), which implies that a larger number of nonzero flows is positively associated with a higher average flow for each nonzero product-destination.

Larger economies export more products to more destinations by virtue of sheer size and diversity, and richer countries might have a better chance to penetrate more markets due to better technology. This relationship between the number of product-destination export flows, size and income is well captured by the following regression, which we fit to data on 135 countries

$$\log(\text{no. of nonzero flows}) = \underset{(0.084)}{-12.73} + \underset{(0.043)}{0.64} \cdot \log(\text{Pop}) + \underset{(0.066)}{1.29} \cdot \log(\text{GDP per capita}) \ ,$$

where robust standard errors are reported in parentheses and $R^2 = 0.8$.[12] Poorer and smaller economies indeed penetrate less markets with less products. However, in terms of explaining export success, this is not the entire story, as **Figure 4** shows. Even after controlling for size (population) and income (GDP per capita) the association between export success and the number of nonzero product-destination export flows remains strong.

Table 3 shows that the number of nonzero export flows is more important for explaining export success than population and GDP per capita: the beta coefficient is 1.77 times larger than that of population (in absolute value) and 2.7 times larger than that of GDP per capita. Favorable productivity or demand shocks are necessary to overcome a threshold to realize a nonzero entry (for either product or destination). Therefore, countries that exhibit higher productivity levels also get to draw from a more favorable productivity distribution

[12] We could obtain GDP data for only 135 countries in our sample.

and penetrate more destinations with more products. We take this into our estimation procedure below.

In order to assess the degree of success in terms of the number of bilateral trade flows, it is useful to have a benchmark. Such a benchmark is provided by the framework of Armenter and Koren (2009). Fix the set of export industries, I. There are 215 potential destination for each industry. Let the share of industry i in exports (not bilateral) be x_i, $i = 1, 2, ...I$. Let the share in exports of destination d (over all industries) be y_d, $d = 1, 2, ...215$ (some y_d may be zero). Assume that all destinations import from a given exporter in the same proportion and that each product is shipped in the same proportion to all destinations. This defines a set of "bins", each of which is of size $x_i y_d$.

Given the total number of shipments n from source s, and if each shipment, or "ball", is randomly assigned to a bin, then the expected number of nonzero bins is given by

$$E\left(\text{non empty bins } | n\right) = \sum_{i=1}^{I} \sum_{d=1}^{215} \left[1 - (1 - x_i y_d)^n\right] \ .$$

The problem is that we do not have n in our data. Instead, we perform the following calibration. Armenter and Koren (2009) calculate that 32% of potential bins are expected to be filled for the U.S. Let $m_{U.S.}$ be the number of nonempty bins for the U.S. We calibrate the factor β to satisfy

$$\sum_{i=1}^{I} \sum_{d=1}^{215} \left[1 - (1 - x_i y_d)^{\beta \cdot m_{U.S.}}\right] = 0.32 \times P_{U.S.} \ ,$$

where $P_{U.S.}$ is the number of potential bins for the U.S. Essentially, β is the average number of shipments per bilateral flow. We use β to calculate the number of shipments per nonempty bin for all countries. The expected number of nonempty bins for each country is thus given by

$$E\left(\text{non empty bins } | m_s, \beta\right) = \sum_{i=1}^{I} \sum_{d=1}^{215} \left[1 - (1 - x_i y_d)^{\beta \cdot m_s}\right] \ . \tag{1}$$

This calculation assumes that the average number of shipments per bilateral flow for all countries is the same as for the U.S.

Figure 5 plots the predicted percent of potential non empty bilateral export flows – calculated according to (1) – against the actual percent. A handful countries export more than predicted according to this benchmark, most notably Hong Kong and Kuwait as outliers (keep in mind the log scale). But the vast majority falls below potential, as can be seen from the number of points above the 45 degree line.

A regression of the percent of overpredicted export flows, defined as the predicted percent minus the actual percent of potential export flows, yields

$$100 \times \frac{\text{Overpredicted flows}}{\text{Potential}} = -\underset{(2.9)}{24} + \underset{(0.014)}{1.2} \cdot \log(\text{Pop}) + \underset{(0.2)}{1.1} \cdot \log(\text{GDP per capita}) \ ,$$

where robust standard errors are reported in parentheses and $R^2 = 0.44$. This means that larger and richer countries actually fall farther below their (random assignment) predicted flows than smaller and less developed countries. Poorer countries seem to do a better job in reaching destinations with the (more limited) set of products that they manage to export relative to richer countries, on average. The last result might follow from the fact that potential, as it is defined above, varies positively with size and income. But even when regressing the log of the number overpredicted flows on the same covariates yields a similar result

$$\log(\text{Overpredicted flows}) = -\underset{(6.2)}{9.8} + \underset{(0.08)}{6.5} \cdot \log(\text{Pop}) + \underset{(0.09)}{0.97} \cdot \log(\text{GDP per capita}) \ ,$$

where robust standard errors are reported in parentheses and $R^2 = 0.65$.[13] So the pattern above is not driven by the definition of potential.

One caveat is that our calculation of potential flows assumes that the average number of shipments per actual flow, β, is the same for all countries. If β is smaller for smaller and poorer countries, then our results above might be misleading. Unfortunately, at this stage we are unable to pursue this point further and we leave the reader with this warning.

1.3 Concentration revisited

We now return to describe concentration of exports across products and destinations. **Table 4** shows the bivariate correlations between all the concentration statistics given above. We see that the "top x" and "top x percent" of nonzero flows are not measuring the same thing; they are sometimes actually negatively related to each other. The problem is that neither statistic is invariant to the number of nonzero product-destination flows, which varies a

[13] We loose a few observations because "overpredicted flows" is negative for a few countries

lot across countries, as we have seen. For mechanical reasons, a larger number of nonzero product-destinations drives down the share of the "Top 3" or "Top 10", but drives up the share of the "Top 1%" or "Top 10%" (exactly the same effect on the concentration ratios is true for total manufacturing export value). The statistics on ratios of the top product-destination to the 10th ranked or 100th ranked are closely related to the shares of the top 3 or top 10, and are related to the other variables in the same way.

It is not clear whether we can construct an ideal concentration ratio when the number of nonzero product-destinations varies so much across countries. Our main results below don't rely on concentration ratios; instead, we characterize the entire shape of the distribution of nonzero entries.

Finally, **Table 5** examines the partial correlations of overall success in exporting with the number of nonzero product-destinations export flows and concentration. The interesting result is that controlling for the number of nonzero product-destination export flows, export revenue per capita is always positively associated with all the different definitions of concentration (with both the top x and top x percent measures). It seems that the most successful exporters by value per capita also have the highest concentration ratios for top x products or top x percent of product-destination exports, conditional on the number of nonzero product-destination export flows they have. It is noteworthy that we obtain very similar results when the regressand is total export revenue (rather than per capita).[14] This strengthens our point about the importance of big hits, because it stresses the magnitude effect of big hit.

The effects of concentration and the number of nonzero product-destination export flows can be related to absolute and comparative advantage. Countries that export a large number of products to many destinations exhibit absolute advantage, or higher average productivity. For a given exporter facing all possible destinations with entry fixed costs, a higher average productivity will allow penetrating more destinations with more products. But given the number of destinations an exporting country penetrates, higher overall export value comes from productivity draws that are high relative to the rest; these are the big hits. Thus, high concentration – or big hits – reflects comparative advantage. The upshot of this is that big hits – i.e. extreme specialization, as reflected in concentration ratios – increases overall

[14]These results are availble by request.

export success, over and above absolute advantage.[15]

1.4 The distribution of exports: mixed lognormal-power law

A country's most successful products account for the bulk of its total export value and therefore the distribution of export values appears to be highly right-skewed. A candidate distribution to describe this distribution would be the Pareto distribution which, as detailed above, is used to explain a variety of highly skewed phenomena.

The Pareto distribution would imply a straight line on a log-log scale of export rank and export value. We plot these rank graphs for all countries but observe that we have a straight line only in the tails of the distributions as illustrated in **Figure 6** for a selection of countries.[16] Eaton, Kortum, and Kramarz (2008) document similar rank graphs for French firms. Here we show that the shape is remarkably similar for practically every country in our dataset. These graphs indicate that the whole distribution does not fit the Pareto. But this is not unusual in economic applications of the Pareto distribution; the same holds for income, firm size and city size.[17] In all cases, a log normal distribution explains well the bottom of the distribution, whereas the Pareto distribution fits well the upper tail.

We simulated a mixed Pareto-log normal random variable and a log-normal random variable, and plotted their respective rank graphs in **Figure 7**. The simulated mixed Pareto-log-normal random variable remarkably resembles the empirical distributions in **Figure 6**. A visual comparison of the two simulated random variables in **Figure 7** indicates that the empirical graphs are "too straight" to fit the log normal. In other words, the distribution of "success" across exports is so skewed that even the highly skewed log normal distribution can not be used to characterize it; it seems to require some combination of the log normal – which is necessary for the lower ranked product-destinations – and something even more skewed, possibly a Pareto distribution (power law) – for the top ranked product-destinations. The simulated mixed Pareto-log normal distribution seems to provide a good candidate.

[15] This feature is taken into account below, when we estimate the contribution of specialization due to technology versus demand shocks.

[16] U.S. (an established industrialized OECD economy), Ghana (a poor African country), Argentina (a middle-income South American country), South Korea (a newly industrialized country, new to the OECD), China (the fast-growing giant) and Estonia (a small open transition economy). The data is by product category by destination and is demeaned by destination to control for the effects of gravity and trade barriers.

[17] For example, see Eeckhout (2004).

To formally reject lognormality of the data we performed two different tests on log export values: the Kolmogorov-Smirnov test and a normality test based on D'Agostino, Belanger, and D'Agostino (1990). Normality is rejected in 85% with the former and in 93% of the cases using the latter test. We conclude that the data cannot be described by a log-normal alone.

In what follows we construct a simple demand-supply framework that yields a distribution of export values which is determined by log normal demand shocks and Pareto productivity dispersion. Our innovation is to derive the lognormal-Pareto mixture distribution for export values and determine the relative role the power law part plays.[18]

2 Technology, demand and bilateral trade barriers

In this section we raise the following question: how much of the variation in export revenue is driven by technological dispersion in the source country; how much is driven by demand shocks from destination countries; and how much is attributable to bilateral factors such as trade barriers. Essentially, we perform a variance decomposition into those three sources. Our interpretation of demand shocks is broad. Demand shocks may capture true taste shocks or "popularity", or finding a good match and successful marketing. Answering this question can advise policy on the types of tools that might – and those that might not – be relevant for promoting trade.

Suppose that demand shocks are the most important source of variation. This would imply that the stress on finding one's comparative advantage is misplaced, because other forces determine trade flows. An implication is that penetrating markets is more about marketing and finding a good match than high productivity in the narrow sense. On the other hand, if technological dispersion is more important, and if it follows a power law, then it would be very hard to predict big hits, because the probability of predicting diminishes exponentially with the size of the hit (this is the definition of a power law).

In order to address these issues we lay out a demand-supply framework which is similar to the backbone of many modern trade models. This framework will allow us to estimate a parameter that governs the distribution of technological dispersion and a parameter that

[18] Arkolakis (2008) develops a model with market penetration that takes into account marketing costs and matches the distribution of exports better than a simple Pareto or log normal can.

governs the importance of demand shocks, as well as incorporate trade barriers. We examine empirically which accounts for a larger share of the variation in the data, country by country. Our results indicate that, on average, demands shocks are almost as important as technological dispersion and trade barriers combined.

In order not to burden the reader with familiar structure we present only the necessary minimum of our framework and relegate the rest to the appendix.

2.1 Revenue and selection equations

Each destination country (market) d is represented by one consumer, whose preferences over products are represented by a CES aggregator. Products are indexed both by the product's "name" i and by source s.[19] Optimal price taking behavior gives rise to the familiar CES demand schedule

$$X_{sid} = \alpha_{sid} \left(\frac{P_{sid}}{P_d} \right)^{-\sigma} \frac{Y_d}{P_d} \ , \tag{2}$$

where α_{sid} is a demand shock, P_{sid} is the price of product i from source s in destination d, P_d and Y_d are the price level and nominal income in destination country d, respectively.[20] As usual, $\sigma > 1$ is assumed. Since we will not be able to identify σ in the estimation procedure, we are silent on whether σ varies across destinations. It is also assumed that α_{sid} is *ex ante* independent of X_{sid}.

In source country s, producer i may export to any destination country d, including domestic sales ($d = s$). Technology is linear in inputs, which for simplicity are assumed to be only labor.[21] For a particular destination d producer i chooses P_{sid} to maximize profits

$$\pi_{sid} = P_{sid} \cdot X_{sid} - C_{si} \cdot X_{sid} - K_{sd}$$

subject to the demand schedule (2). The producer's constant marginal cost, C_{si}, is given by

$$C_{si} = \frac{W_s}{z_{si}} \ ,$$

where W_s are wages in source country s and z_{si} is labor productivity. $K_{sd} > 0$ is a bilateral

[19] This follows the organization of the data in Comtrade and it implies product differentiation at the good-source level. So widgets from Kenya are differentiated from widgets from Costa Rica, even if they are both called "widgets" in the data. This is essentially an Armington assumption.

[20] See the appendix for a more complete description.

[21] One could also entertain a composite of inputs, the cost of which is the same across industries.

fixed setup cost for business in country s to penetrate destination market d. These capture beach head costs and shipping costs. In this we depart from the usual iceberg trade costs, but it can be shown that adding iceberg trade costs does not alter the estimation below. Since K_{sd} varies bilaterally, it captures any trade cost, whether due to distance, language, etc.[22] The implicit assumption here is that there is just one producer of product i in source country s, which can export to all destinations. This follows the limitations of the data, which aggregates over producers.

Optimal pricing is a fixed markup over marginal cost. Thus, revenue for producer i in source country s selling in destination d is given by

$$R_{sid} = \alpha_{sid} \left(\frac{\sigma}{\sigma - 1} \frac{W_s}{z_{si}} \right)^{1-\sigma} P_d^{\sigma-1} Y_d \ .$$

Taking logs we get the following expression

$$r_{sid} = \beta_0^r - \beta_s^w + \beta_d^{py} + \ln \alpha_{sid} + (\sigma - 1) \ln z_{si} \ , \tag{3}$$

where $r_{sid} = \ln R_{sid}$, $\beta_0^r = (1-\sigma) \ln \frac{\sigma}{\sigma-1}$, $\beta_s^w = (\sigma - 1) \ln W_s$, $\beta_d^{py} = (\sigma - 1) \ln P_d + \ln Y_d$. This is the revenue equation.

Equation (3) describes observed revenue conditional on positive profits. Given real income in the destination country and costs in the source country, profits will be positive if

$$\pi_{sid} = R_{sid} - C_{si} \cdot X_{sid} - K_{sd} \geq 0 \ .$$

Using the previous results, we have

$$\alpha_{sid} \cdot z_{si}^{\sigma-1} \geq \sigma^{\sigma} (\sigma - 1)^{1-\sigma} \frac{K_{sd}}{Y_d} \left(\frac{W_s}{P_d} \right)^{\sigma-1} \ .$$

This expression means that the demand shock and productivity must overcome a threshold. The threshold is larger if costs (W_s) are higher in the source country or if the bilateral fixed setup cost (K_{sd}) is larger. The threshold is lower for larger markets (Y_d) that in which the price level (P_d) is higher. Taking logs and rearranging yields

$$\ln \alpha_{sid} + (\sigma - 1) \ln z_{si} \geq \beta_0^s + \beta_s^w - \beta_d^{py} + \beta_{sd}^k \ , \tag{4}$$

[22] One can interpret the fixed cost to include bribes at the border, making connections with potential buyers, adjusting the good to comply with local regulations, etc'.

where $\beta_0^s = \ln\left(\sigma^\sigma (\sigma-1)^{1-\sigma}\right)$, $\beta_{sd}^k = \ln K_{sd}$ and β_s^w and β_d^{py} were defined above. This is the selection equation.

2.2 Empirical specification

We would like to estimate the relative contribution of z_{si} versus α_{sid} to the variation of export revenues, while taking into account bilateral factors. To this end we will make some distributional assumptions that will enable us to write down a likelihood function for export revenue. We will then maximize it in order to retrieve the distribution parameters of the underlying productivity and demand shocks. Using this information, we will be able to decompose the variance.

We assume that α_{sid} is distributed log-normal such that $\ln \alpha_{sid}$ is distributed normal with zero mean and variance η^2.[23] We do not index η^2 by destination d, which reflects our assumption that in percent terms demand shocks should not be different across countries. We assume that z_{si} in source country s is distributed Pareto,

$$Z \sim F_s(z) = 1 - \left(\frac{c_s}{z}\right)^{a_s},$$

where $z > c_s > 0$ and $a_s > 0$.[24] It is assumed that α and z are independent.

Equations (3) and (4) can then be written as

$$\text{Revenue} \quad : \quad r_{sid} = \beta_0^r - \beta_s^w + \beta_d^{py} + \theta_{si} + \delta_{sid} \tag{5}$$

$$\text{Selection} \quad : \quad \theta_{si} + \delta_{sid} \geq \beta_0^s + \beta_s^w - \beta_d^{py} + \beta_{sd}^k\ , \tag{6}$$

where $\delta_{sid} = \ln \alpha_{sid}$ is distributed normal with zero mean and variance η^2; and $\theta_{si} = (\sigma-1)\ln z_{si}$ is distributed conditional exponential

$$F_s(\theta) = 1 - c_s^{a_s} e^{-\frac{a_s}{\sigma-1}\theta} = 1 - e^{-\lambda_s(\theta - m_s)}\ ,$$

[23] Eaton, Kortum, and Kramarz (2008) also include lognormal demand shocks in their analysis of French firms exporting behavior.

[24] Helpman, Melitz, and Yeaple (2004) also assume a Pareto distribution for productivity, but do not allow it to change by source country.

where

$$\begin{aligned} \theta &\geq (\sigma - 1)\ln(c_s) \equiv m_s \\ \lambda_s &\equiv \frac{a_s}{\sigma - 1} \ . \end{aligned}$$

Thus θ_{si} is distributed conditional exponential with mean $m_s + 1/\lambda_s$.[25]

2.3 Maximum likelihood estimation

We will estimate λ_s, m_s and η for each source country separately. Therefore, the source specific dimensions are absorbed in the constant and only destination specific parameters are separately identified. For an arbitrary source country, equations (5) and (6) can be collapsed into

$$\text{Revenue} \ : \ r_{id} = \beta_d + \theta_i + \delta_{id} \tag{7}$$

$$\text{Selection} \ : \ r_{id} \geq r_d^{\min} \ . \tag{8}$$

Note that standard Heckman correction is not appropriate due to different distributional assumptions. Likewise, estimating (7) by least squares is not feasible because the mean of θ_{si} is not zero in general, so the intercept is not separately identified.

Let I denote the set of industries for some source country s and let $D\,(i)$ denote the set of destinations for industry $i \in I$, so that each industry may have a different set of destinations which it serves with non zero export flows. First, note that a (super) consistent estimator of $r_d^{\min}$ is

$$t_d \equiv \widehat{r}_d^{\min} = \min_{i \in I} \{r_{id}\} \ .$$

We will use this directly in the likelihood function. The log likelihood function for an arbitrary source country is given by

$$\log \mathcal{L} = \sum_{i \in I} \log \int_m \left[\prod_{d \in D(i)} \frac{\frac{1}{\eta}\phi\left(\frac{r_{id} - \beta_d - \theta_i}{\eta}\right)}{1 - \Phi\left(\frac{t_d - \beta_d - \theta_i}{\eta}\right)} \right] \lambda e^{\lambda m} e^{-\lambda \theta_i} d\theta_i \ , \tag{9}$$

where Φ and ϕ are the standard normal CDF and pdf, respectively, and we already replaced

[25] Notice that $(\sigma - 1)\ln(c_s)$ can be positive or negative, but since $c_s > 0$ and $\sigma > 1$, $(\sigma - 1)\ln(c_s)$ is bounded away from $-\infty$. This is not a standard exponential random variable, in the sense that θ can be less than zero, but all the properties of the exponential distribution are preserved.

$r_d^{\min}$ with its estimator, t_d. See the appendix for the derivation of (9).

Maximizing this likelihood presents us with some challenges. In principle, we could try to estimate all β_d coefficients, but this is not feasible for two reasons. First, it would be computationally prohibitively taxing for most countries. But more important, the β_d coefficients are not separately identified from m. The best way to see this is to realize that m is just a mean shifter for θ_i; this will become apparent in (10) below. The same logic that applies to perfect colinearity with a full set of dummy variables in the presence of a constant intercept applies here as well.

Therefore, we replace β_d with $b_d = \overline{r}_d$, which is the average of trade flows to destination d, calculated from the data. This will bias the estimator of m, but no better solution is available at this point. However, Monte-Carlo simulations suggest that this only biased the estimator of m and hardly affects the estimators of λ and η. The reason for the relative insensitivity of λ and η is that they are identified from the shape of the distribution, not its location. This will become apparent in (10) below.

Another challenge is to evaluate the integral in (9). First, we use a change of variables in order to simplify the integral,

$$\begin{aligned} u &= \lambda\left(\theta - m\right) \\ &\Rightarrow\; du = \lambda d\theta \\ &\Rightarrow\; \theta = u/\lambda + m\ . \end{aligned}$$

Thus, (9) becomes

$$\log \mathcal{L} = \sum_{i=1}^{I} \log \int_0 \left[\prod_{d(i)=1}^{D(i)} \frac{\frac{1}{\eta}\phi\left(\frac{r_{id}-\beta_d-u/\lambda-m}{\eta}\right)}{1-\Phi\left(\frac{t_d-\beta_d-u/\lambda-m}{\eta}\right)} \right] e^{-u} du\ . \tag{10}$$

Given this form, we can apply Gaussian Quadrature to approximate the integral to an arbitrary level of precision. Operationally, (10) is approximated by

$$\log \mathcal{L} \approx \sum_{i=1}^{I} \log \sum_{j=1}^{N} \left[\prod_{d(i)=1}^{D(i)} \frac{\frac{1}{\eta}\phi\left(\frac{r_{id}-b_d-u_j/\lambda-m}{\eta}\right)}{1-\Phi\left(\frac{t_d-b_d-u_j/\lambda-m}{\eta}\right)} \right] W_j\ , \tag{11}$$

where $\{u_j, W_j\}_{j=1}^{N}$ are obtained from Stroud and Secrest (1966). We use $N = 40$. For each of the 151 source countries in our data we maximize (11) with respect to λ_s, m_s and η. Note that although η is assumed not to vary by country, in practice we obtain different estimates of η, for each source country. However, it turns out that the estimates of η across countries are very similar; they are centered around 2 with a standard deviation of 0.25.

In order to ensure that our code works, we performed Monte Carlo simulations and backed out the original parameters successfully. The initial values for the numerical optimizer were chosen as empirical moments from the data. For each source country the initial value for λ was chosen as the reciprocal of the average over all $r_{id} - t_d$. The initial value for η was chosen as the standard deviation of $r_{id} - t_d$. The initial value for m was -1. Changing the initial values for the search did not affect the results.[26]

2.4 Estimation results and variance decomposition

The parameters for almost all countries are very precisely estimated (see appendix for detailed results by exporter). **Table 6** provides summary statistics. As mentioned above, the estimates of the standard deviation of demand shocks are tightly estimated around 2. Although some outliers exist, this parameter is remarkably stable across countries, with slightly larger variances for richer countries; a regression of η on log GDP per capita yields a small, but highly statistically significant coefficient of 0.06.

Estimates of λ do not vary systematically with size, income or number of export flows. **Figure 8** plots the estimated λ by country against log GDP per capita. Almost all estimates of λ fall within 0.5 and 1 with three noticeable outliers: Comoros, Gabon and Suriname.[27] The median estimate is 0.8. These estimates are admittedly very low, since it implies an infinite mean for actual export revenues (not in logs). However, it is interesting that the estimates are rather similar and do not vary systematically with income. The upshot is that the distribution of technology is remarkably similar across countries, assuming elasticities of demand are also similar (recall: $\lambda = a/(\sigma - 1)$). This can inform theoretical modeling.[28]

[26] As a robustness check, we used perturbed point estimates as initial values in a second round of optimization search. These perturbations were plus of minus 50% of the point estimate from the first round of estimation.

[27] The estimation failed to for Cook Isds, Greenland, Maldives, Turks and Caicos Isds. These countries have very few export flows, most of which are shipped to only one destination.

[28] Typical estimates of σ in similar settings are well above 2, in the range of 5-12 . This would place the estimate of the Pareto coefficient, a, above 2, which is reassuring, because it restricts the primitive distribution of productivity in the model to have finite first and second moments.

This does not imply that the *level* of the distributions of technology are the same in all countries. Higher m_s makes it more likely to penetrate any given destination market. The values of m_s vary considerably and are positively correlated (not in absolute value) to population. However, recall that these estimates are biased because we use b_d instead of β_d in the estimation. Unfortunately, we cannot make any conclusions based on the estimates of m.

Although the estimates themselves are somewhat interesting, we are more interested in their implication for the variation in export revenues. Given estimates of λ, m and η, and given the set of values for b_d for each source country, we decompose the variance of r_{id} given $r_{id} \geq t_d$ where t_d replaces $r_d^{\min}$.

$$\begin{aligned}
V\left(r_{id}|r_{id} \geq t_d\right) &= V\left(\beta_d + \theta_i + \delta_{id}|r_{id} \geq t_d\right) \\
&= V\left(\beta_d|r_{id} \geq t_d\right) + V\left(\theta_i|r_{id} \geq t_d\right) + V\left(\delta_{id}|r_{id} \geq t_d\right) \\
&= V\left(\beta_d\right) + V\left(\theta_i|\beta_d + \theta_i + \delta_{id} \geq t_d\right) + V\left(\delta_{id}|\beta_d + \theta_i + \delta_{id} \geq t_d\right) \\
&= V\left(b_d\right) + V\left(\theta_i|\theta_i + \delta_{id} \geq t_d - b_d\right) + V\left(\delta_{id}|\theta_i + \delta_{id} \geq t_d - b_d\right) \ .
\end{aligned}$$

We can replace $V\left(\beta_d|r_{id} \geq t_d\right)$ with $V\left(\beta_d\right)$ because β_d does not depend on $r_{id} \geq t_d$. We can replace $V\left(\beta_d\right)$ with $V\left(b_d\right)$ because β_d and b_d differ only by a constant that does not vary with d. All parameters were estimated using b_d instead of β_d. Therefore, we use

$$V\left(r_{id}|r_{id} \geq t_d\right) = V\left(b_d\right) + V\left(\theta_i|\theta_i + \delta_{id} \geq t_d - b_d\right) + V\left(\delta_{id}|\theta_i + \delta_{id} \geq t_d - b_d\right) \ .$$

We compute the second and third conditional variances by simulation. In doing so, we take into account the fact that the condition $\theta_i + \delta_{id} \geq t_d - b_d$ varies across destinations, with d. In order to address this issue, we decompose each conditional variance according to the variance decomposition (ANOVA) formula

$$V\left(X|\theta_i + \delta_{id} \geq t_d - b_d\right) = V_d\left[E\left(X|\theta_i + \delta_{id} \geq t_d - b_d, d\right)\right] + E_d\left[V\left(X|\theta_i + \delta_{id} \geq t_d - b_d, d\right)\right] \ ,$$

where X represents either θ or δ. See appendix for complete details.

We report summary statistics for the contribution to the variation in trade revenue due to technology (θ), demand shocks (δ) and bilateral trade barriers (b_d) in **Table 7**. On average, 31.3% of the variance is due to technological dispersion, 46.3% is due to demand shocks and 22.4% is due to bilateral trade barriers (by definition, these must sum to 100%). Demand accounts for as much as both technological dispersion and trade barriers, combined. However, there is a lot of dispersion around these average statistics.

In **Table 8** we report some correlates of the contribution to the variation in trade revenue. First, size and income seem to be associated with a higher contribution of demand, and lower contributions of technology and trade barriers. We find similar results when using size and overall export success (log export value per capita) instead income. This implies that for richer countries, finding the right market is the most important determinant of the variation in trade revenues. Consistent with this, the number of trade flows is also positively correlated with a higher contribution of demand. Both these facts point to absolute advantage being the main driver of exports for rich countries. Consistent with this view, overall export success, measured by the log of total export revenue per capita, is also associated with a higher role for demand shocks. However, the role of the number of bilateral export flows seems to be more important.

Looking at the same table differently, we can conclude that relative to the average country, poorer countries rely more on technological dispersion, rather than demand shocks. Trade barriers are more important for them than for rich countries.

3 Conclusion

In this paper we document the high degree of specialization in exports in a sample of 151 economies. Specialization is remarkably high in exporting manufactures. The distribution is remarkably skewed. We find that very few "big hits" account for a disproportionate share of export volumes and can also explain high degrees of specialization. We also find that higher concentration (i.e., big hits) is positively correlated with export success, after controlling for the number of products that are exported and destinations that are reached. Larger countries export more products to more destinations and so do richer countries, where the latter is driven by absolute advantage. Controlling for the number of product-destination export flows, overall export volumes are positively correlated with higher concentration, which are explained by big hits. This is driven by comparative advantage.

We analyze the determinants of these big hits. We find that demand shocks explains almost half of the variation in export trade flows, with about one fourth due either to technological dispersion and trade barriers. Developing countries export less products to fewer destinations, which helps explaining this. Exporting to more destinations exposes a country to more demand shocks that are uncorrelated with technological dispersion. Therefore, as a country penetrates more markets with more products, demand shocks from those markets and for those products account for a larger percent of variation – and hence concentration – in exports. We also find that in poorer countries the relative contribution of technology and barriers to the variation in exports is higher and the relative contribution of demand shocks is smaller. Developing countries rely less on to demand shocks and more on comparative advantage.

Overall, countries with a higher share of variance due to demand shocks are more successful at achieving high manufacturing export values per capita, but the number of non-zero export flows is more important. This is consistent with the positive association between concentration ratios and overall manufacturing export success, controlling for number of non-zero export flows.

Our analysis leads us to some important conclusions that are relevant for policies that aim to promote trade. We find that a power law plays an important role in the distribution of export value across possible product-destination pairs. This makes the fierce debate about the relative weights on the government and the market in "picking winners" even more relevant than previously realized in the literature. A power law means that successfully picking a winner becomes less likely exponentially with the degree of success that is predicted. Over and above this mechanism, the high relative exposure of developing countries to demand shocks, given their successful export flows, implies an even smaller role for picking winners.

The "picking winners" debate is about two things: probability of discovering a "winner" and externalities from identifying the winner to other firms. The traditional argument for relying on free markets to decide what to produce is that they make possible a decentralized search by myriads of entrepreneurs, and provide means for scaling up successful hits through reinvestment of profits and financing by capital markets. The probability of any one agent – such as a government policymaker – finding which product-destination combination will be the big hit is very small. In fact, the track record of governments in picking winners is not great, as Lee (1996) demonstrates for Korea.[29] Hence, an alternative implication – nearly the opposite of Hausmann-Rodrik conclusion – of the hyper-specialization phenomenon is that entrepreneurs and financiers should be as unhindered as possible from any government intervention.

However, if there are externalities from the discovery of a "big hit" to other firms who can also export the same good-destination pair, then there is a market failure leading to too little discovery effort by any one entrepreneur. This lead to the traditional argument for government intervention to subsidize "discovery", as Hausmann and Rodrik emphasized. Perhaps one could try to get the best of both worlds by designing a blanket government subsidy to all "discovery" efforts, while leaving the process of identifying the winners to private entrepreneurs. How to design such a policy in practice, and whether the traditional arguments fully apply to the stylized facts we have uncovered is far from definitive. Our main contribution is to show that finding winning hyper-specializations is even harder – and yet the rewards to finding these hyper-specializations are also even larger – than previously thought.

[29] We are not saying that industrial policy in Korea did not contribute to its subsequent success. We only point out that the "picking winners" part of that policy has not proven to be successful.

Appendix

A Demand structure

There are D destination countries (importers). Let preferences in destination country d be given by

$$U_d = \left(\int \alpha_{sid}^{1/\sigma} x_{sid}^{\frac{\sigma-1}{\sigma}} d\left(s,i,d\right) \right)^{\frac{\sigma}{\sigma-1}} ,$$

where x_{sid} denotes product i from source country s and α_{sid} are preference weights (demand shocks) associated with those products. As usual, $\sigma > 1$ is assumed. We assume that elasticities of substitution in demand, σ, are the same in all countries. We assume that α_{sid} are *ex ante* independent of x_{sid}.

Maximizing this utility function under the following budget constraint

$$\int p_{sid} x_{sid} d\left(s,i,d\right) \leq y_d$$

gives rise to the demand schedule

$$x_{sid} = \alpha_{sid} \left(\frac{p_{sid}}{p_d} \right)^{-\sigma} \frac{y_d}{p_d} ,$$

where y_d denotes nominal national income and p_d is the perfect price index in destination d

$$p_d = \left(\int \alpha_{sid} \cdot p_{sid}^{1-\sigma} d\left(s,i,d\right) \right)^{\frac{1}{1-\sigma}} .$$

B Deriving the likelihood function

For some source country (index is omitted), industry i and destination d we have two equations

$$\begin{aligned} \text{Revenue} &: \; r_{id} = \beta_d + \theta_i + \delta_{id} \\ \text{Selection} &: \; r_{id} \geq r_d^{\min} . \end{aligned}$$

Plug r_{id} from the revenue equation into the selection equation and rearrange to get

$$\begin{aligned} \text{Revenue} &: \; \theta_i + \delta_{id} = r_{id} - \beta_d \\ \text{Selection} &: \; \theta_i + \delta_{id} \geq r_d^{\min} - \beta_d . \end{aligned}$$

Distribution assumptions:

$$\begin{aligned} \text{Technology} &: \; \theta_i \overset{iid}{\sim} \exp\left(\lambda\right), \; \theta_i \geq (\sigma - 1) \ln c \equiv m \\ \text{Demand} &: \; \delta_{id} \overset{iid}{\sim} N\left(0, \eta^2\right) \\ \text{Orthogonality} &: \; \theta_i \perp \delta_{id} . \end{aligned}$$

The density function for θ is

$$f_{\lambda,m}(\theta) = e^{\lambda m}\lambda e^{-\lambda\theta} = \lambda e^{-\lambda(\theta-m)}$$

and the CDF is

$$F_{\lambda,m}(\theta) = 1 - e^{\lambda m}e^{-\lambda\theta} = 1 - e^{-\lambda(\theta-m)} \ .$$

A (super) consistent estimator of $r_d^{\min}$ is

$$t_d \equiv \widehat{r}_d^{\min} = \min_i \{r_{id}\} \ .$$

We use this directly in the likelihood function.

Deriving the likelihood function:

1. Probability to observe r_{id} for each destination and industry, conditional on θ_i

$$\begin{aligned} f\left(r_{id} = \beta_d + \theta_i + \delta_{id} | \theta_i,\ r_{id} \geq t_d\right) &= f\left(\delta_{id} = r_{id} - \beta_d - \theta_i | \theta_i,\ \delta_{id} \geq t_d - \beta_d - \theta_i\right) \\ &= f\left(\frac{\delta_{id}}{\eta} = \frac{r_{id} - \beta_d - \theta_i}{\eta}\middle|\ \theta_i,\ \frac{\delta_{id}}{\eta} \geq \frac{t_d - \beta_d - \theta_i}{\eta}\right) \\ &= \frac{\frac{1}{\eta}\phi\left(\frac{r_{id}-\beta_d-\theta_i}{\eta}\right)}{1 - \Phi\left(\frac{t_d-\beta_d-\theta_i}{\eta}\right)} \ . \end{aligned}$$

2. Probability to observe $\{r_{id}\}_{d\in D(i)}$ for each industry, conditional on θ_i

$$\begin{aligned} & f\left(r_{id} = \beta_d + \theta_i + \delta_{id} | \theta_i,\ r_{id} \geq t_d,\ d \in D(i)\right) \\ = \ & f\left(\delta_{id} = r_{id} - \beta_d - \theta_i | \theta_i,\ \delta_{id} \geq t_d - \beta_d - \theta_i,\ d \in D(i)\right) \\ = \ & \prod_{d\in D(i)} \frac{\frac{1}{\eta}\phi\left(\frac{r_{id}-\beta_d-\theta_i}{\eta}\right)}{1 - \Phi\left(\frac{t_d-\beta_d-\theta_i}{\eta}\right)} \ , \end{aligned}$$

   where $D(i)$ denotes the set of destinations for industry i. We take the product due to the *iid* assumption for δ_{id}. This depends both on i (industry) and θ_i.

3. Probability to observe $\{r_{id}\}_{d=1}^{D(i)}$ for each industry, unconditional of θ_i

$$\begin{aligned} & f\left(r_{id} = \beta_d + \theta_i + \delta_{id} | r_{id} \geq t_d,\ d \in D(i)\right) \\ = \ & f\left(\theta_i + \delta_{id} = r_{id} - \beta_d | \theta_i + \delta_{id} \geq t_d - \beta_d \ , d \in D(i)\right) \\ = \ & \int_m \left[\prod_{d\in D(i)} \frac{\frac{1}{\eta}\phi\left(\frac{r_{id}-\beta_d-\theta_i}{\eta}\right)}{1 - \Phi\left(\frac{t_d-\beta_d-\theta_i}{\eta}\right)}\right] f_{\lambda,m}(\theta_i)\, d\theta_i \ . \end{aligned}$$

   This step is similar to the convolution theorem since we integrated out θ_i. This depends only on i (industry).

4. Likelihood function $\mathcal{L}\left(\theta_i + \delta_{id} = r_{id} - \beta_d | \theta_i + \delta_{id} \geq t_d - \beta_d,\ d \in D(i),\ i \in I\right)$

$$\begin{aligned}
\mathcal{L} &= \prod_{i \in I} f\left(r_{id} = \beta_d + \theta_i + \delta_{id} | r_{id} \geq t_d,\ d \in D(i)\right) \\
&= \prod_{i \in I} f\left(\theta_i + \delta_{id} = r_{id} - \beta_d | \theta_i + \delta_{id} \geq t_d - \beta_d\ , d \in D(i)\right) \\
&= \prod_{i \in I} \int_m \left[\prod_{d \in D(i)} \frac{\frac{1}{\eta}\phi\left(\frac{r_{id}-\beta_d-\theta_i}{\eta}\right)}{1-\Phi\left(\frac{t_d-\beta_d-\theta_i}{\eta}\right)}\right] f_{\lambda,m}\left(\theta_i\right) d\theta_i\ ,
\end{aligned}$$

where I is the set of industries for the source country.

Taking logs, we obtain the expression in the text.

C Gaussian Quadrature

Stroud and Secrest (1966) show that for some functions $f(x)$ and $w(x)$ there exist N points (nodes) x_i and N weights W_i such that

$$\int_a^b f(x)\, w(x)\, dx \approx \sum_{j=1}^{N} f(x_j)\, W_j\ .$$

It is assumed that the weighting function $w(x)$ has well defined finite moments

$$c_k = \int_a^b w(x)\, x^k dx\ ,\ \ k = 0, 1, 2, ...$$

and $c_0 \geq 0$. In our case $w(x)$ is e^{-x} and $a = 0$ and $b = \infty$, so this condition is satisfied. The higher N is, the better the approximation.

D Simulating conditional variances

Here we describe the algorithm for simulating the conditional variances for each source country. We start with estimates of λ, m and η, and the set of b_d and cutoff values t_d for each destination country.

1. Draw a large number S (we use S =50,000) of uniform (u) and standard normal (z) random variables and store them. Both vectors are ($S \times 1$) and will be used for all countries and destinations.

2. Simulate exponential productivity values, θ, and normal demand shocks, δ, as follows

$$\begin{aligned}
\theta &= m - \ln(1-u)/\lambda \\
\delta &= \eta * z\ .
\end{aligned}$$

   The vectors θ and δ are ($S \times 1$).

3. For each destination d, generate a $(S \times 1)$ logical indicator vector

$$I\left(\theta + \delta \geq t_d - b_d\right) \ .$$

4. Simulate

$$E\left[X|\theta + \delta \geq t_d - b_d\right] = \frac{E\left[X \cdot I\left(\theta + \delta \geq t_d - b_d\right)\right]}{E\left[I\left(\theta + \delta \geq t_d - b_d\right)\right]} \approx \frac{\frac{1}{D}X \cdot I\left(\theta + \delta \geq t_d - b_d\right)}{\frac{1}{S}\iota' I\left(\theta + \delta \geq t_d - b_d\right)} \ ,$$

   where $X \cdot I\left(\theta + \delta \geq t_d - b_d\right)$ means that X is taken into account only when the condition $\theta + \delta \geq t_d - b_d$ is met; the number of such cases is $D = \iota' I\left(\theta + \delta \geq t_d - b_d\right)$. The vector ι is a $(S \times 1)$ vector of ones. X can be either θ, θ^2, δ or δ^2. We use these values to compute variances according to

$$V\left(X|\theta + \delta \geq t_d - b_d\right) = E\left(X^2|\theta + \delta \geq t_d - b_d\right) - \left[E\left(X|\theta + \delta \geq t_d - b_d\right)\right]^2 \ .$$

5. Repeat 4 for each destination d, and store the results.

6. Use the decomposition

$$V\left(X|\theta + \delta \geq t_d - b_d\right) = V_d\left[E\left(X|\theta + \delta \geq t_d - b_d, d\right)\right] + E_d\left[V\left(X|\theta + \delta \geq t_d - b_d, d\right)\right]$$

   to compute the conditional variance of θ and δ, where the values inside brackets are calculated above. The operators over d ($V_d\left[\cdot\right]$ and $E_d\left[\cdot\right]$) use sample analogues. We calculate $V_d\left[\cdot\right]$ and $E_d\left[\cdot\right]$ in two ways: once without weights and then using the number of observations per destination as weights.

Repeat for each source country.

References

Andriani, P., and B. McKelvey (2005): "Why Gaussian statistics are mostly wrong for strategic organization," *Strategic Organization*, 3(2), 219–228.

Arkolakis, K. (2008): "Market Penetration Costs and the New Consumers Margin in International Trade," Yale Working Paper.

Arkolakis, K., and M.-A. Muendler (2009): "The Extensive Margin of Exporting Goods: A Firm-Level Analysis," Yale Working Paper.

Armenter, R., and M. Koren (2009): "A Balls-and-Bins Model of Trade," Working Paper.

Baldwin, R., and J. Harrigan (2007): "Zeros, Quality and Space: Trade Theory and Trade Evidence," NBER Working Paper 13214.

Bernard, A., B. Jensen, S. Redding, and P. Schott (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105–130.

Besedes, T., and T. J. Prusa (2008): "The Role of Extensive and Intensive Margins and Export Growth," Working Paper.

D'Agostino, R. B., A. Belanger, and R. B. J. D'Agostino (1990): "A Suggestion for Using Powerful and Informative Tests of Normality," *The American Statistician*, 44(4), 316–321.

Eaton, J., M. Eslava, M. Kugler, and J. Tybout (2007): "Export Dynamics in Colombia: Firm-Level Evidence," NBER Working Paper 13531.

Eaton, J., S. Kortum, and F. Kramarz (2008): "An Anatomy of International Trade: Evidence from French Firms," Working Paper.

Eeckhout, J. (2004): "Gibrat's Law for (all) Cities," *American Economic Review*, 94(5), 1429–1451.

Gibrat, R. (1931): *Les inetǵalitetś etćonomiques; applications: aux inetǵalitetś des richesses, a' la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., dŠune loi nouvelle, la loi de lŠeffet proportionnel*. Librairie du Recueil Sirey, Paris.

Hausmann, R., and D. Rodrik (2006): "Doomed to Choose: Industrial Policy as Predicament," Working Paper.

Helpman, E., M. Melitz, and S. Yeaple (2004): "Exports vs. FDI with Heterogeneous Firms," *American Economic Review*, 94(1), 300–316.

Imbs, J., and R. Wacziarg (2003): "Stages of Diversification," *American Economic Review*, 93(1), 63–86.

Lee, J.-W. (1996): "Government Interventions and Productivity Growth," *Journal of Economic Growth*, 1(2), 391–414.

Luttmer, E. G. J. (2007): "Selection, Growth and the Size Distribution of Firms," *Quarterly Journal of Economics*, August, 1103–1144.

Newman, M. E. J. (2005): "Power Laws, Pareto Distributions and Zipf's Law," *Contemporary Physics*, 46(5), 323–351.

Pareto, V. (1896): *Cours d'Economie Politique*. Droz, Geneva.

Stroud, A. H., and D. Secrest (1966): *Gaussian Quadrature Formulas*. Prentice-Hall, Englewood Cliffs, N.J.

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35, 40–59.

Table 1: Concentration Ratios for Export Products by Country, Summary Statistics

	Median	Mean	Minimum	Maximum
Percent of the following in total manufacturing export revenues:				
Top 3 products	28	34	5	96
Top 10 products	49	52	13	100
Top 1%	47	48	18	92
Top 10%	86	85	43	99
Top 20%	94	93	66	99
Bottom 50%	0.8	1.3	0.1	17.3
Other statistics:				
Ratio of Top product value to 10th ranked product value	7.2	20.3	1.8	626.6
Ratio of Top product value to 100th ranked product value	104.8	1004.1	10.8	84478.2
Share of Top product in world import market for that product	0.018	0.066	0	0.698
Number of products exported (# of nonzero entries)	1035	1302	10	2950

Notes: 151 observations (countries). The numbers are for export values by product, regardless of the number of export destinations. Source: U.N. Comtrade and authors calculations.

Table 2: Concentration Ratios for Product-Destination Bilateral Trade Flows, Summary Statistics

	Median	Mean	Minimum	Maximum
Percent in manufacturing export value of:				
Top 3 product-destinations	18	24	1	93
Top 10 product-destinations	34	38	3	100
Top 1% *	52	52	20	85
Top 10% *	89	87	53	99
Top 20% *	95	94	72	99.5
Bottom 50% *	0.8	1.4	0.1	14.5
Other statistics:				
ratio product-destination 1 value to product-destination 10	5	14	2	317
ratio product-destination 1 value to product-destination 100	48	1064	5	121154
Share of top product-destination in destination's imports of that product	0.18	0.32	0	1
Nonzero products-destinations	3055	19985	10	195417
Nonzero products-destinations/Potential	0.015	0.039	0.005	0.31
Total manufacturing export value (dollars)	516,000,000	26,544,261,836	87,105	598,300,000,000

Notes: 151 observations (countries). All statistics reflect product-destination export flows. * Percentages apply to total number of nonzero product-destination export flows for each country. Total product categories = 2985. Total possible destinations = 215. Potential is given by the product of the number of the number of exported products for the source country times 215, which is the maximum number of destinations. Source: U.N. Comtrade and authors calculations.

Table 3: Export Success and Destinations

	(1)	(2)	(3)	(5)
	Dependent Variable: Log of Total Export Value Per Capita			
Log(Number of Nonzero Export Flows)	0.961***	1.378***	1.011***	0.78
	(12.48)	(27.19)	(12.24)	
Log Population		-0.872***	-0.642***	-0.44
		(-17.57)	(-11.33)	
Log GDP per capita			0.631***	0.29
			(5.120)	
Observations	145	145	135	
R-squared	0.565	0.861	0.886	

Notes: Number of Nonzero Export Flows is the number of product-destination categories that a country exports. GDP is corrected for PPP. Column (4) reports beta coefficients for the specification in column (3). Source: U.N. Comtrade, World Bank World Development Indicators. Robust t statistics in parentheses. *** significant at 1%. A constant was included but is not reported.

Table 4: Correlations Between Export Success and Concentration

	lvalue	N	Top 3	Top 10	Top 1%	Top 10%	Top 20%	log(p1/p10)	log(p1/p100)
Log Total Export Value	1								
No. of Export Flows	0.71	1							
Top 3 Flows	-0.68	-0.45	1						
Top 10 Flows	-0.75	-0.54	0.96	1					
Top 1% Flows	0.53	0.27	0.12	0.02	1				
Top 10% Flows	0.65	0.27	-0.11	-0.15	0.8	1			
Top 20% Flows	0.67	0.29	-0.17	-0.21	0.72	0.98	1		
Log(prod1/prod10)	-0.51	-0.4	0.9	0.8	0.26	-0.02	-0.08	1	
Log(prod1/prod100)	-0.56	-0.48	0.93	0.94	0.17	0.09	0.03	0.82	1

Notes: 151 observations (countries). Lvalue is the log of total export value. No. of Export Flows is the number of nonzero product-destination categories a country exports. Top 3 Flows (Top 3) is the percent of export value accounted by the largest 3 product-destination export flows from a country; similarly for Top 10. Top 1% Flows (Top1%) is the percent of total export value accounted by the largest 1% of nonzero product-destination export flows from a country; similarly for 10% and 20%. Log(prod1/prod10) is the log of the ratio of the largest product-destination export flow to the 10th largest; similarly for 100. Source: U.N. Comtrade and authors calculations.

Table 5: Export Success and Concentration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Log of Total Export Value Per Capita							
Log(Number of Nonzero Export Flows)	1.242***	1.483***	0.840***	0.769***	0.754***	1.038***	1.252***	0.343**
	(11.51)	(13.13)	(9.473)	(9.163)	(8.946)	(10.80)	(12.44)	(2.412)
Share of Top 3 Flows	3.760***							
	(2.914)							
Share of Top 10 Flows		5.297***						
		(4.955)						
Share of Top 1% Flows *			4.984***					
			(4.337)					
Share of Top 10% Flows *				11.727***				
				(5.510)				
Share of Top 20% Flows *					20.524***			
					(5.134)			
Log(prod1/prod10)						0.281		
						(1.358)		
Log(prod1/prod100)							0.553***	
							(4.449)	
Log Top 3 Flows Value								0.574***
								(5.490)
Observations	145	145	145	145	145	145	140	145
R-squared	0.593	0.621	0.617	0.643	0.646	0.571	0.568	0.637

Notes: 151 observations (countries). Top 3 Flows (Top3) is the percent of export value accounted by the largest 3 bilateral product-destination export flows from a country; similarly for Top 10. Top 1% Flows is the percent of total export value accounted by the largest 1% nonzero product-destination export flows from a country; similarly for 10% and 20%. * Percentages apply to total number of nonzero product-destination export flows for each country. Log(prod1/prod10) is the log of the ratio of the largest bilateral product-destination export flow to the 10th largest; similarly for 100. Source: U.N. Comtrade and authors calculations. Robust *t* statistics in parentheses. *** significant at 1%. A constant was included but is not reported.

Table 6: Summary Statistics for Estimates

	λ	m	η
Minimum	0.53	-4.15	1.07
Median	0.80	-2.25	1.93
Mean	0.85	-2.31	1.94
Maximum	2.24	-1.32	2.77
Standard Deviation	0.21	0.46	0.26

Notes: Statistics calculated for 147 observations (countries). The estimation failed to for Cook Isds, Greenland, Maldives, Turks and Caicos Isds. λ is the exponential exponent parameter. m is the exponential cutoff parameter. η is the standard deviation of demand shocks.

Table 7: Summary Statistics for Variance Decomposition

	Percent of Variance due to:		
	Technology	Demand	Barriers
Minimum	2.8	18.5	2.6
Median	30.1	47.1	21.0
Mean	31.3	46.3	22.4
Maximum	63.9	71.6	72.2
Standard Deviation	10.2	10.5	9.2

Notes: Statistics calculated for 147 observations (countries). The relative contributions of technology, demand and trade barrieirs were calcualted using the parameter estimates reported in **Table 6**. Variances were calculated using weights for destinations. See text for details.

Table 8: Correlates of Contributions to Variance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Percent Variance due to:	Technological Dispersion					Demand Shocks					Bilateral Trade Barriers				
Log Population	-0.964**				-1.058**	1.903***				1.815***	-0.940**				-0.757**
	(0.467)				(0.431)	(0.407)				(0.366)	(0.377)				(0.369)
Log GDP Per Capita, PPP	-1.028					4.187***					-3.159***				
	(0.701)					(0.610)					(0.566)				
Log export flows		-0.943**		-0.703			2.878***		2.571***			-1.935***		-1.868***	
		(0.391)		(0.625)			(0.333)		(0.523)			(0.321)		(0.510)	
Log Export Value Per Capita			-0.645**	-0.239	-0.713**			1.790***	0.308	1.908***			-1.146***	-0.0688	-1.195***
			(0.305)	(0.472)	(0.301)			(0.275)	(0.395)	(0.256)			(0.259)	(0.385)	(0.257)
Observations	134	147	143	143	143	134	147	143	143	143	134	147	143	143	143
R-squared	0.042	0.039	0.031	0.039	0.071	0.321	0.340	0.231	0.344	0.346	0.208	0.200	0.122	0.199	0.147

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. A constant was used in all specifications, but is not reported.

Table A1: Concentration Ratios for Export Goods by Country (Part 1 of 2)

Exporter	Top 3	Top 10	Top 1%	Top 10%	Top 20%	Bottom 50%	N
Albania	50	67	62	90	95	0.68	667
Algeria	28	56	53	95	99	0.12	821
Andorra	19	46	43	88	95	0.7	824
Anguilla	36	72	36	86	95	0.73	219
Antigua and Barbuda	36	52	52	87	94	0.85	965
Argentina	18	35	49	87	95	0.49	2578
Armenia	42	60	57	86	94	0.91	714
Australia	16	34	48	81	91	1.4	2840
Austria	8	18	31	76	89	1.33	2765
Azerbaijan	40	62	60	93	97	0.36	828
Bahamas	31	50	52	90	97	0.21	1086
Bahrain	53	80	77	98	99	0.1	851
Bangladesh	27	56	41	89	97	0.28	490
Barbados	29	53	58	93	98	0.23	1218
Belarus	21	36	50	86	94	0.66	2240
Belgium	15	22	34	76	88	1.73	2902
Belize	74	86	78	94	98	0.27	322
Benin	26	54	20	73	86	2.75	174
Bolivia	57	71	71	93	97	0.28	969
Botswana	26	45	58	93	97	0.34	1930
Brazil	20	34	47	84	93	0.65	2690
Bulgaria	7	19	34	83	94	0.61	2495
Burkina Faso	24	48	35	83	94	0.75	486
Burundi	90	99	68	90	95	1.03	25
Cambodia	41	65	55	94	98	0.11	507
Canada	27	42	56	86	94	0.66	2856
Cape Verde	50	72	64	93	97	0.36	575
Central African Rep.	29	60	20	66	83	3.19	128
Chile	36	49	60	91	97	0.32	2127
China	7	16	30	75	87	1.96	2928
Colombia	16	30	43	85	95	0.43	2235
Comoros	85	94	69	91	95	1.16	52
Cook Isds	80	99	43	72	80	6.41	14
Costa Rica	57	70	77	97	99	0.1	1706
Cote d'Ivoire	20	38	46	91	97	0.33	1321
Croatia	22	35	48	88	95	0.46	2302
Cuba	43	64	60	91	97	0.24	774
Cyprus	30	45	50	89	95	0.62	1471
Czech Rep.	11	22	35	76	89	1.4	2894
Denmark	9	19	33	77	90	1.09	2733
Dominica	68	92	68	97	99	0.16	264
Ecuador	24	42	39	89	96	0.5	893
Egypt	38	57	59	94	98	0.24	1075
El Salvador	14	30	39	88	96	0.47	1530
Estonia	40	49	58	88	95	0.61	2337
Ethiopia	81	93	73	88	94	0.86	52
Fiji	44	63	63	94	98	0.27	976
Finland	30	45	56	89	96	0.31	2757
France	11	24	40	75	87	2.26	2867
French Polynesia	45	75	65	92	96	0.57	544
Gabon	24	43	36	80	91	1.47	602
Gambia	70	87	64	89	94	1.12	127
Georgia	37	59	57	91	96	0.43	878
Germany	13	24	34	70	84	2.83	2890
Ghana	41	60	57	90	96	0.49	707
Greece	14	29	44	85	93	0.75	2445
Greenland	53	81	53	90	95	1.09	236
Grenada	86	93	86	97	99	0.1	285
Guatemala	19	35	48	90	96	0.37	1960
Guinea	95	98	92	99	99	0.08	145
Guyana	38	66	61	94	98	0.2	707
Honduras	51	69	69	95	98	0.12	962
Hong Kong	11	22	38	83	93	0.81	2813
Hungary	22	40	51	85	93	0.78	2236
Iceland	31	61	61	95	98	0.22	959
India	9	22	38	79	90	1.55	2855
Indonesia	11	24	38	83	94	0.58	2645
Iran	44	54	60	89	96	0.33	1535
Ireland	28	60	75	96	99	0.12	2467
Israel	26	42	54	91	97	0.26	1860
Italy	5	13	27	68	82	2.94	2915
Jamaica	52	76	74	95	98	0.2	839
Japan	16	28	43	83	93	0.74	2900
Jordan	17	32	40	81	90	1.71	1803
Kazakhstan	21	42	51	88	95	0.57	1513
Kenya	18	35	46	90	96	0.47	1652
Kuwait	66	83	83	97	99	0.17	906
Kyrgyzstan	25	46	48	87	95	0.76	1032
Latvia	16	30	42	84	94	0.65	2097

Table A1: Concentration Ratios for Export Products by Country (Part 2 of 2)

Exporter	Top 3	Top 10	Top 1%	Top 10%	Top 20%	Bottom 50%	N
Lebanon	14	28	37	80	90	1.39	1681
Lesotho	54	85	46	87	96	0.16	103
Lithuania	13	28	44	87	94	0.63	2416
Luxembourg	17	36	53	94	98	0.19	2194
Macao	20	44	53	96	99	0.08	1306
Madagascar	44	71	69	96	99	0.08	875
Malaysia	32	50	69	93	97	0.33	2703
Maldives	72	94	32	77	91	1.66	39
Mali	17	42	22	77	90	1.12	353
Malta	74	82	84	98	99	0.06	1249
Mauritius	54	76	81	98	99	0.1	1546
Mexico	16	31	50	88	95	0.38	2877
Mongolia	45	73	60	92	97	0.13	406
Montserrat	47	73	40	80	91	1.77	131
Morocco	22	44	55	93	98	0.1	1632
Mozambique	20	41	33	82	93	0.68	635
Namibia	59	70	76	93	97	0.35	1993
Nepal	50	75	50	88	95	0.35	228
Netherlands	14	30	44	78	89	1.5	2827
New Caledonia	22	40	38	83	92	1.44	845
New Zealand	17	29	44	83	93	0.91	2503
Nicaragua	29	52	43	88	95	0.67	699
Niger	57	73	73	94	98	0.13	909
Nigeria	53	79	46	89	95	0.59	160
Norway	9	22	39	85	94	0.6	2568
Oman	32	56	54	90	96	0.45	820
Panama	28	60	33	87	95	0.72	355
Papua New Guinea	48	75	62	95	98	0.22	437
Paraguay	30	54	35	79	91	1.3	323
Peru	38	54	64	93	98	0.23	1907
Philippines	55	73	79	96	99	0.1	1800
Poland	12	27	37	75	88	2.16	2249
Portugal	15	32	48	87	95	0.46	2592
Qatar	65	82	77	96	98	0.27	646
Rep. of Korea	26	44	57	88	95	0.61	2809
Rep. of Moldova	41	54	56	91	97	0.29	1158
Romania	11	24	38	86	95	0.53	2175
Russian Federation	12	25	41	83	93	0.7	2785
Saint Kitts and Nevis	73	90	77	97	99	0.2	337
Saint Lucia	58	84	70	96	98	0.27	468
Saint Vincent and the Grenadines	50	69	58	90	96	0.67	449
Sao Tome and Principe	64	91	38	71	83	2.22	32
Saudi Arabia	32	55	69	95	98	0.27	2100
Senegal	26	44	40	86	94	0.57	772
Serbia and Montenegro	10	21	31	79	91	1.22	1890
Singapore	31	53	66	91	96	0.57	2897
Slovakia	27	37	48	86	95	0.42	2641
Slovenia	16	26	41	82	93	0.5	2574
South Africa	23	33	46	82	91	1.3	2881
Spain	19	33	45	78	88	1.73	2920
Sudan	78	86	78	94	98	0.03	278
Suriname	26	48	33	82	93	0.75	426
Swaziland	54	73	84	97	99	0.11	1871
Sweden	19	33	43	80	91	0.82	2853
Switzerland	12	22	34	78	91	0.8	2945
TFYR of Macedonia	17	33	43	90	97	0.28	1601
Tanzania	27	59	39	90	96	0.43	458
Thailand	22	36	49	87	95	0.39	2702
Togo	49	75	49	88	95	0.79	261
Trinidad and Tobago	61	73	78	96	99	0.16	1724
Tunisia	20	40	51	89	96	0.25	1682
Turkey	14	28	44	85	94	0.62	2742
Turkmenistan	53	81	53	95	98	0.1	260
Turks and Caicos Isds	31	53	31	78	90	1.05	275
USA	14	25	40	75	86	2.63	2950
Uganda	29	49	33	78	90	1.5	372
Ukraine	12	24	36	82	93	0.65	2309
United Kingdom	10	26	42	76	87	2.37	2900
Uruguay	18	35	38	86	95	0.44	1118
Venezuela	16	36	51	91	97	0.36	1876
Zambia	53	72	70	95	98	0.12	864
Zimbabwe	20	37	46	86	95	0.61	1851
Minimum	5	13	20	66	80	0.03	-
Mean	34	52	52	87	94	1	-
Median	28	49	49	88	95	0.57	-
Maximum	95	99	92	99	99	6.41	-

Notes: Top 3 is the share of the largest 3 export categories. Top 10 is the share of the largest 10 export categories. Top #% is the share of the # percent largest export categories. Bottom 50% is the share of the 50% smallest export categories. N is the total number of export categories.

Table A2: Concentration Ratios for Export Product-Destinations by Country and Destination (Part 1 of 2)

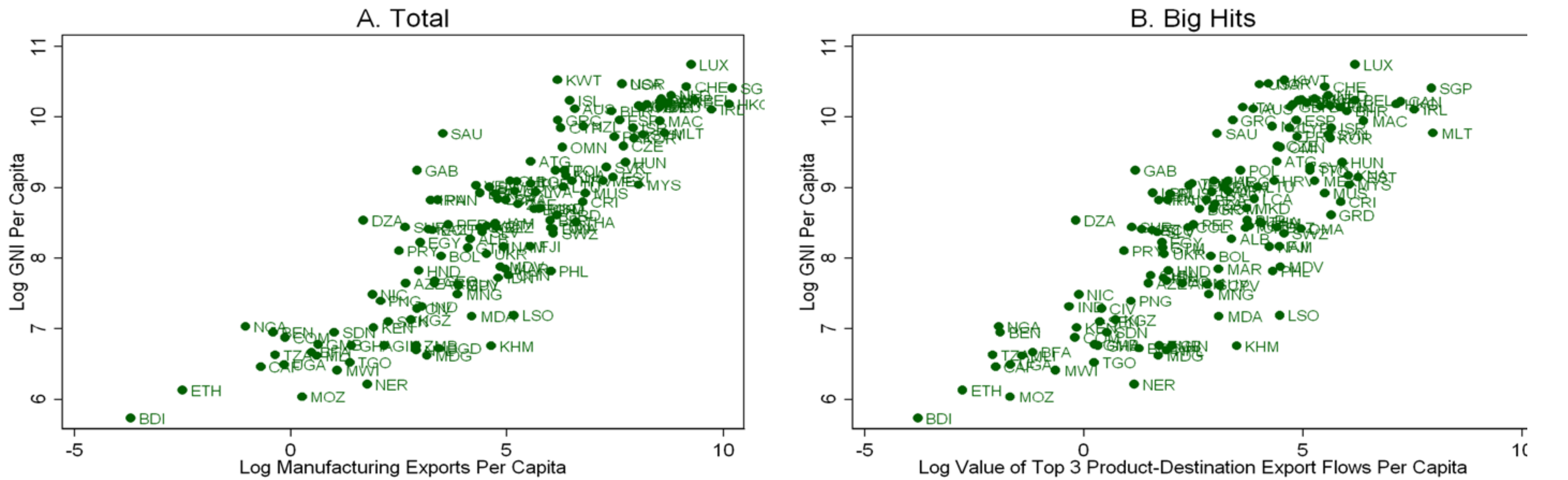
Exporter	Top 3	Top 10	Top 1%	Top 10%	Top 20%	Bottom 50%
Albania	46	61	60	89	95	0.01
Algeria	15	39	37	90	97	0.01
Andorra	12	32	30	80	90	0.02
Anguilla	31	64	23	74	88	0.03
Antigua and Barbuda	32	45	49	81	89	0.03
Argentina	14	24	59	91	96	0.01
Armenia	34	51	51	83	91	0.02
Australia	7	15	59	89	95	0.01
Austria	4	8	56	90	96	0.00
Azerbaijan	30	48	51	87	94	0.01
Bahamas	23	41	49	89	96	0.00
Bahrain	24	44	48	87	95	0.01
Bangladesh	12	25	45	89	95	0.01
Barbados	13	28	47	85	92	0.02
Belarus	10	22	56	91	97	0.00
Belgium	4	9	57	92	97	0.00
Belize	72	85	72	92	96	0.01
Benin	22	41	22	62	78	0.05
Bolivia	56	66	70	91	96	0.01
Botswana	25	38	50	88	94	0.01
Brazil	11	20	63	91	96	0.00
Bulgaria	5	11	45	86	94	0.01
Burkina Faso	19	33	30	73	86	0.03
Burundi	91	100	68	68	81	0.06
Cambodia	32	52	61	95	98	0.00
Canada	27	40	82	98	99	0.00
Cape Verde	46	70	53	88	95	0.01
Central African Rep.	27	49	20	58	76	0.06
Chile	12	24	60	90	96	0.01
China	3	7	62	92	97	0.00
Colombia	5	11	65	94	98	0.00
Comoros	12	27	54	92	97	0.00
Cook Isds	14	20	49	87	94	0.01
Costa Rica	93	100	47	82	93	0.04
Cote d'Ivoire	81	99	43	72	81	0.10
Croatia	41	56	77	95	98	0.00
Cuba	8	16	35	81	91	0.01
Cyprus	12	21	55	89	96	0.01
Czech Rep.	20	36	38	81	91	0.02
Denmark	21	29	49	85	92	0.02
Dominica	4	9	57	91	96	0.00
Ecuador	3	7	46	86	94	0.01
Egypt	33	61	38	87	94	0.01
El Salvador	18	31	47	86	94	0.01
Estonia	30	41	61	90	95	0.01
Ethiopia	6	16	34	81	92	0.01
Fiji	30	41	61	90	96	0.01
Finland	77	90	38	87	94	0.02
France	34	58	69	94	97	0.01
French Polynesia	7	15	60	91	96	0.00
Gabon	3	7	61	91	96	0.00
Gambia	23	60	51	90	95	0.01
Georgia	18	34	34	74	86	0.04
Germany	68	82	63	85	90	0.03
Ghana	24	39	45	87	94	0.01
Greece	4	9	54	90	96	0.00
Greenland	34	51	54	85	92	0.02
Grenada	6	14	50	86	94	0.01
Guatemala	53	81	40	86	93	0.02
Guinea	60	90	60	96	98	0.00
Guyana	10	19	40	85	93	0.01
Honduras	86	97	76	98	99	0.00
Hong Kong	35	58	51	88	94	0.01
Hungary	36	58	60	90	95	0.01
Iceland	16	26	68	94	98	0.00
India	21	36	49	85	94	0.01
Indonesia	3	8	50	86	93	0.01
Iran	5	12	56	90	96	0.01
Ireland	22	35	59	90	95	0.01
Israel	11	22	74	96	99	0.00
Italy	10	21	59	91	96	0.01
Jamaica	1	3	51	87	95	0.01
Japan	51	68	77	93	96	0.01
Jordan	9	14	64	93	98	0.00
Kazakhstan	16	32	50	87	94	0.01
Kenya	13	23	41	83	92	0.02
Kuwait	21	41	71	95	98	0.00
Kyrgyzstan	13	28	34	81	92	0.02
Latvia	6	15	41	84	93	0.01

Table A2: Concentration Ratios for Export Product-Destinations by Country and Destination (Part 1 of 2)

Exporter	Top 3	Top 10	Top 1%	Top 10%	Top 20%	Bottom 50%
Lebanon	6	14	42	79	88	2.59
Lesotho	50	83	42	85	95	0.42
Lithuania	10	18	50	87	94	0.98
Luxembourg	5	13	52	92	97	0.45
Macao	24	46	56	92	97	0.50
Madagascar	18	41	50	88	95	0.60
Malaysia	14	25	74	95	98	0.28
Maldives	69	91	32	83	94	0.84
Mali	13	34	21	68	83	3.43
Malta	51	68	81	97	99	0.24
Mauritius	27	51	71	95	98	0.34
Mexico	14	28	85	99	100	0.06
Mongolia	37	63	50	87	95	0.63
Montserrat	41	66	25	66	81	5.26
Morocco	15	26	57	93	98	0.28
Mozambique	14	31	25	75	87	2.72
Namibia	51	64	73	92	96	0.92
Nepal	36	58	63	94	98	0.40
Netherlands	4	9	60	92	97	0.39
New Caledonia	22	38	38	77	87	3.76
New Zealand	9	16	58	91	96	0.73
Nicaragua	14	33	31	81	90	2.19
Niger	54	70	70	89	94	1.28
Nigeria	41	72	32	87	94	1.05
Norway	3	9	51	88	95	0.66
Oman	17	34	52	87	94	1.00
Panama	23	39	35	78	89	2.09
Papua New Guinea	37	63	42	90	95	1.08
Paraguay	20	38	34	76	88	2.23
Peru	32	42	62	89	95	0.84
Philippines	18	38	81	97	99	0.17
Poland	6	14	42	77	87	3.64
Portugal	7	15	61	92	97	0.41
Qatar	18	40	56	93	97	0.46
Rep. of Korea	10	20	68	94	97	0.30
Rep. of Moldova	33	45	53	88	94	0.94
Romania	6	13	53	90	96	0.50
Russian Federation	6	14	58	92	97	0.43
Saint Kitts and Nevis	73	86	73	93	97	0.91
Saint Lucia	41	68	51	91	95	1.31
Saint Vincent and the Grenadines	46	62	53	82	90	2.34
Sao Tome and Principe	66	93	39	53	72	14.48
Saudi Arabia	62	82	71	96	98	0.30
Senegal	15	28	35	77	88	2.25
Serbia and Montenegro	6	13	35	78	90	1.73
Singapore	10	21	76	95	98	0.21
Slovakia	12	21	57	90	96	0.47
Slovenia	8	14	49	87	94	0.74
South Africa	10	19	62	90	96	0.45
Spain	6	13	62	91	96	0.56
Sudan	62	80	62	91	95	0.78
Suriname	21	40	25	68	83	4.19
Swaziland	22	48	61	95	98	0.37
Sweden	3	7	53	89	96	0.44
Switzerland	3	6	54	90	96	0.39
TFYR of Macedonia	13	23	41	86	94	0.94
Tanzania	8	15	66	94	98	0.32
Thailand	32	51	39	80	90	2.56
Togo	38	57	72	92	96	0.81
Trinidad and Tobago	10	21	50	91	97	0.43
Tunisia	5	11	62	91	96	0.60
Turkey	47	73	47	88	94	0.82
Turkmenistan	31	53	26	68	82	4.27
Turks and Caicos Isds	3	7	66	93	97	0.33
USA	22	36	29	69	82	4.57
Uganda	7	16	50	88	95	0.63
Ukraine	3	8	62	91	96	0.45
United Kingdom	18	44	27	78	90	1.89
Uruguay	15	31	45	84	93	1.06
Venezuela	15	27	52	89	95	0.83
Zambia	31	63	65	93	97	0.62
Zimbabwe	11	24	46	84	92	1.65
Minimum	1	3	20	53	72	0.06
Mean	24	38	52	87	94	1.37
Median	18	34	52	89	95	0.83
Maximum	93	100	85	99	100	14.48

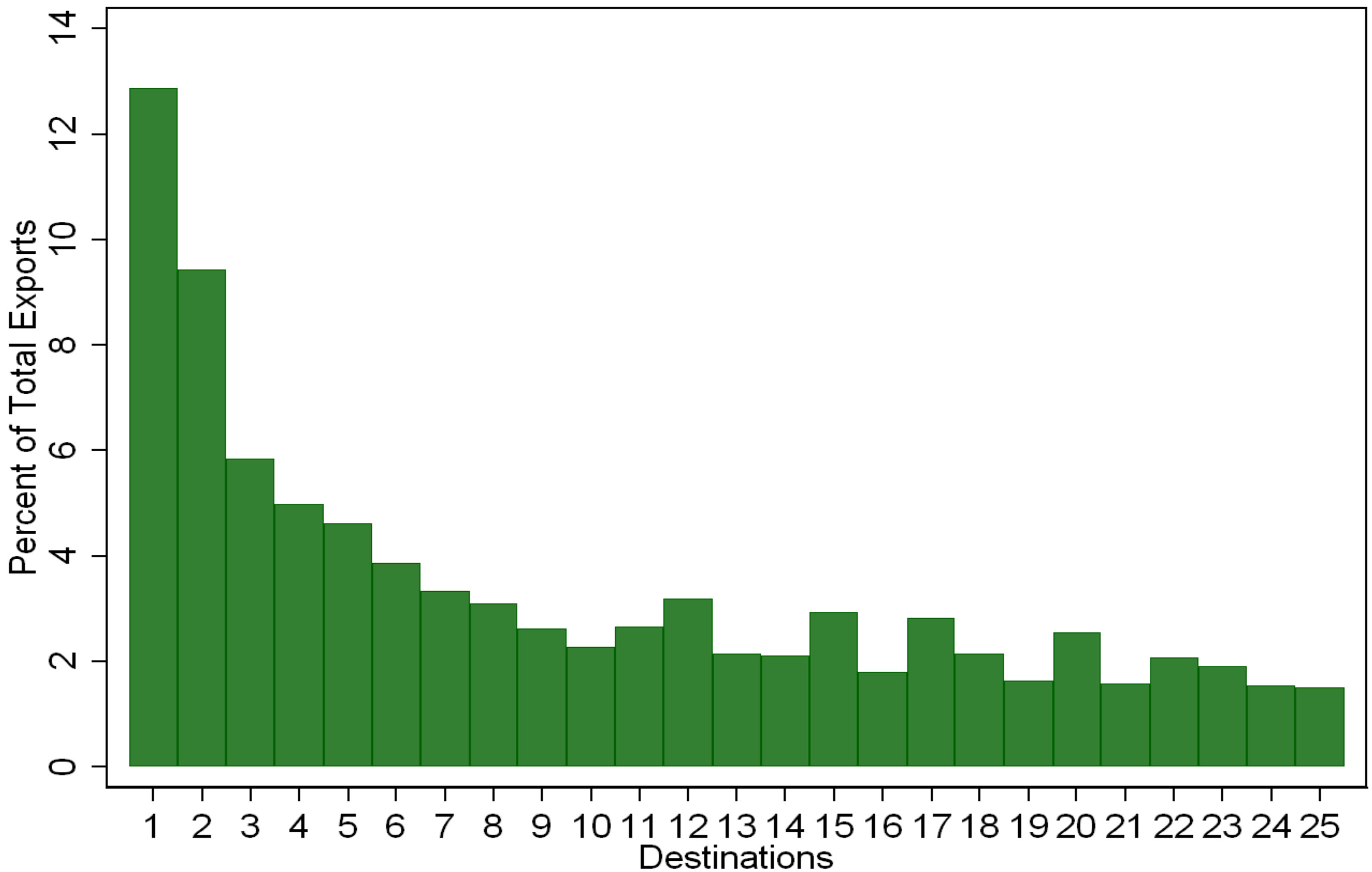
Notes: Top 3 is the share of the largest 3 export flows by product-destination. Top 10 is the share of the largest 10 export flows by product-destination. Top #% is the share of the # percent largest export flows by product-destination. Bottom 50% is the share of the 50% smallest export flows by product-destination.

Figure 1: Manufacturing Exports and Development



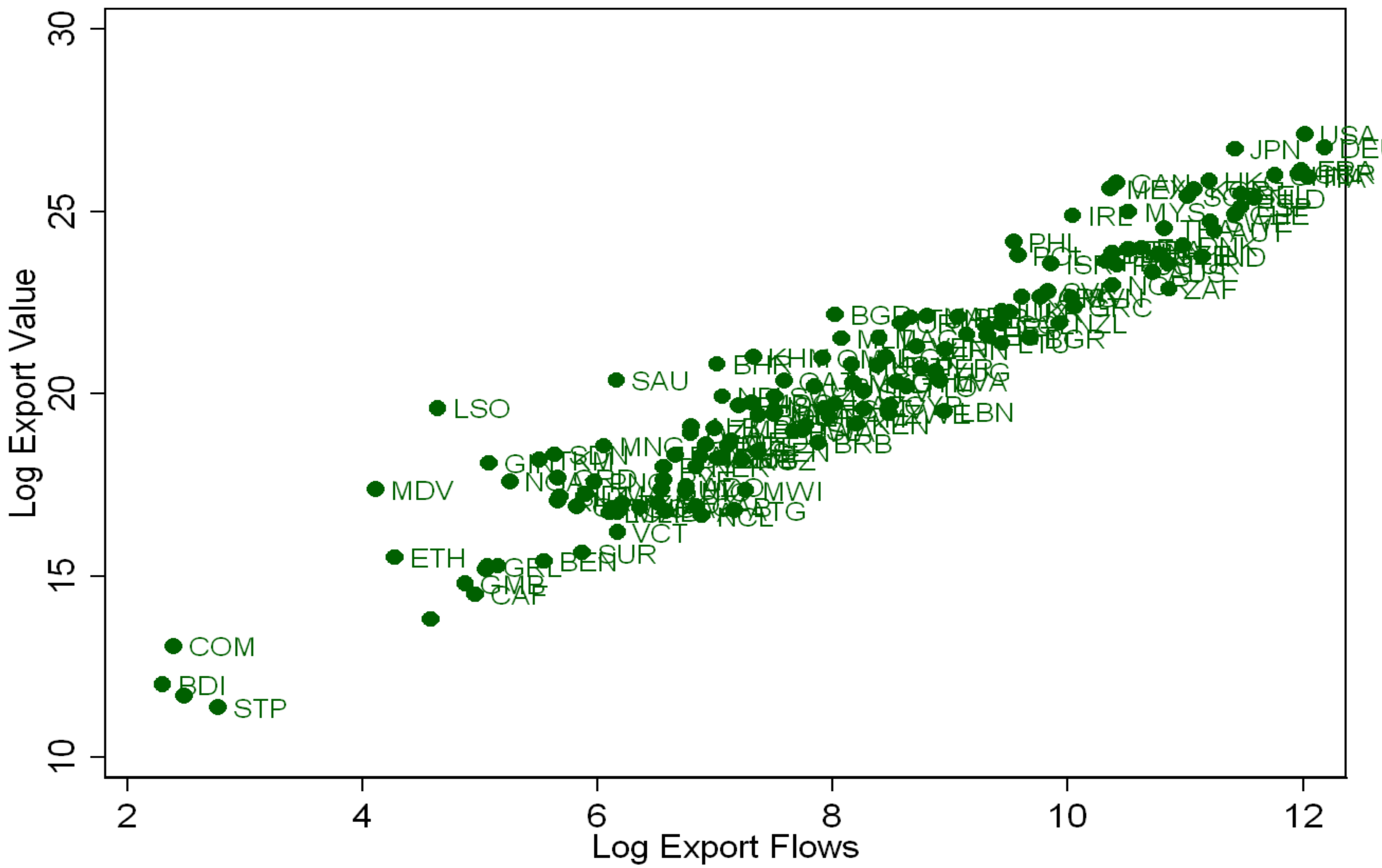
Notes: All data for 2000. GNI is corrected for PPP. Source: The World Bank, World Development Indicators, U.N., Comtrade.

Figure 2: Distribution of Export Revenue Across Destinations



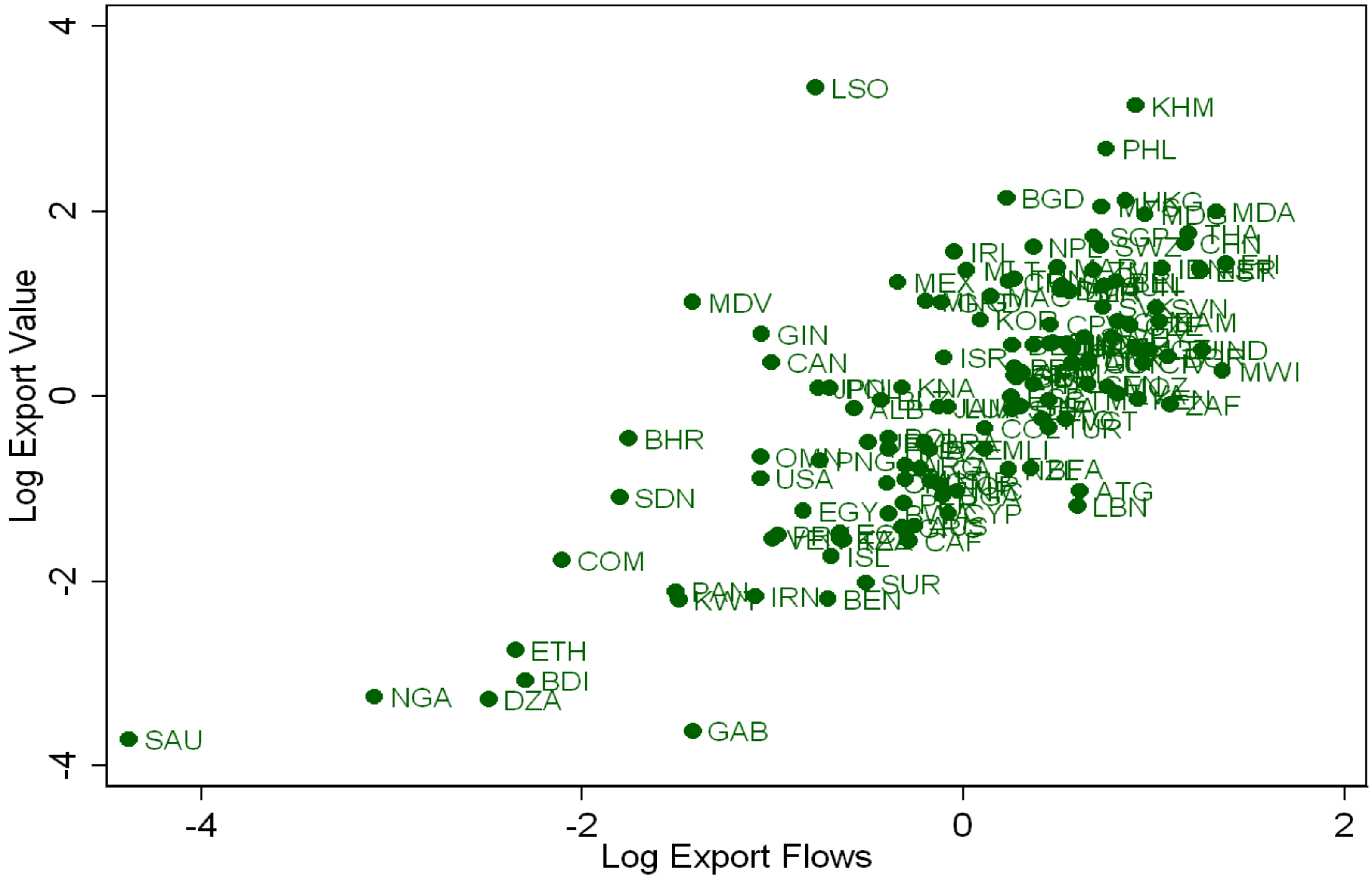
Notes: The figure displays the average percent of export revenue accounted for by products shipped to X destinations. For each country, export values by product were assigned to bins according to the number of destinations to which that product was exported. Each bin was assigned the percent of total export value that it accounted for. The figure depicts the average percent over all 151 exporters in the sample. The rest of the distribution continues to diminish slowly after 25 destinations.

Figure 3: Export Success and Product-Destination Flows



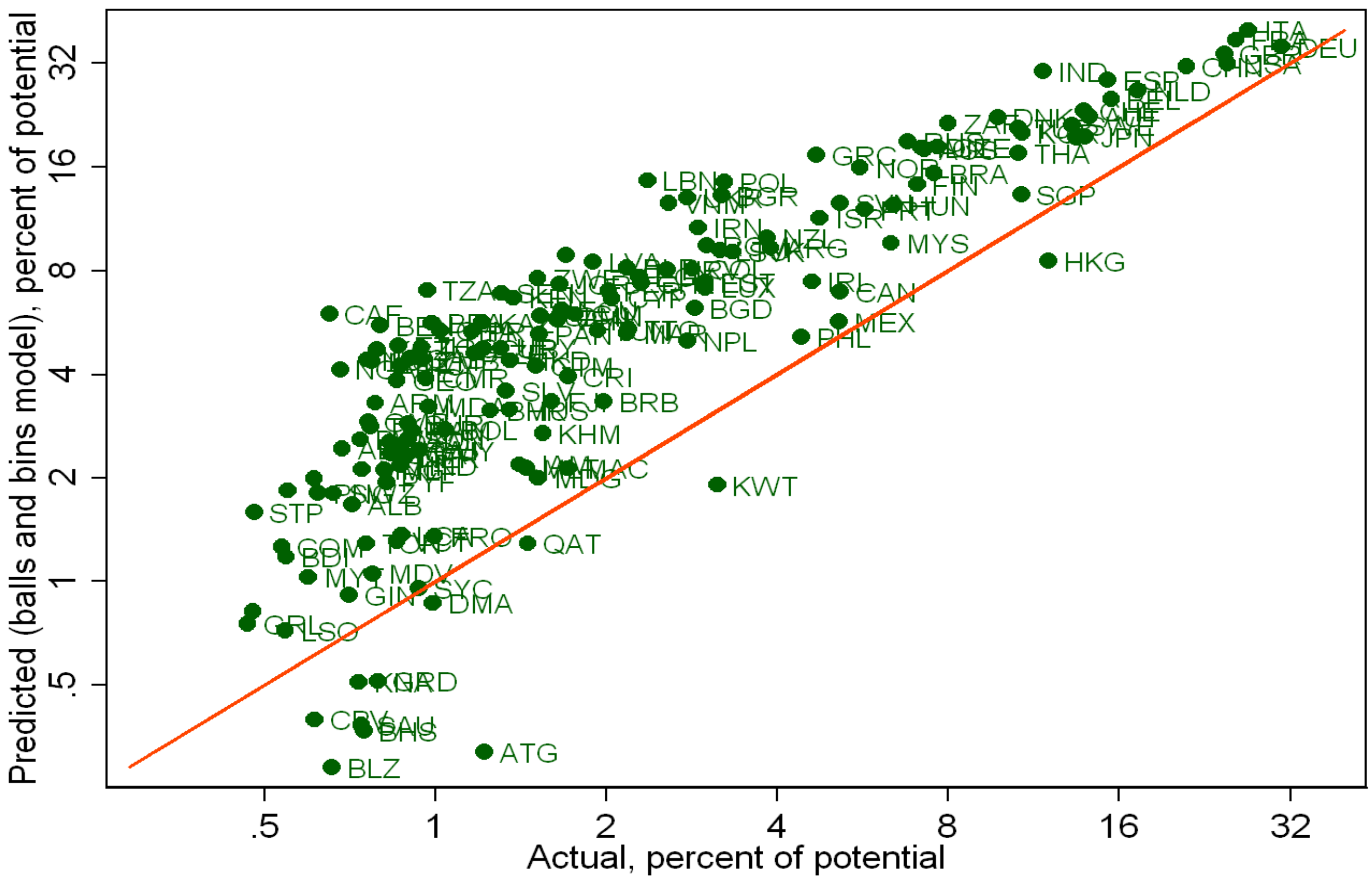
Notes: Each observation is a country. Export Flows is the number of nonzero product-destination categories that a country exports. Export Value is the log of total export value that a country exports. Source: U.N., Comtrade.

Figure 4: Export Success and Product-Destination Flows, Conditional on Size and Income



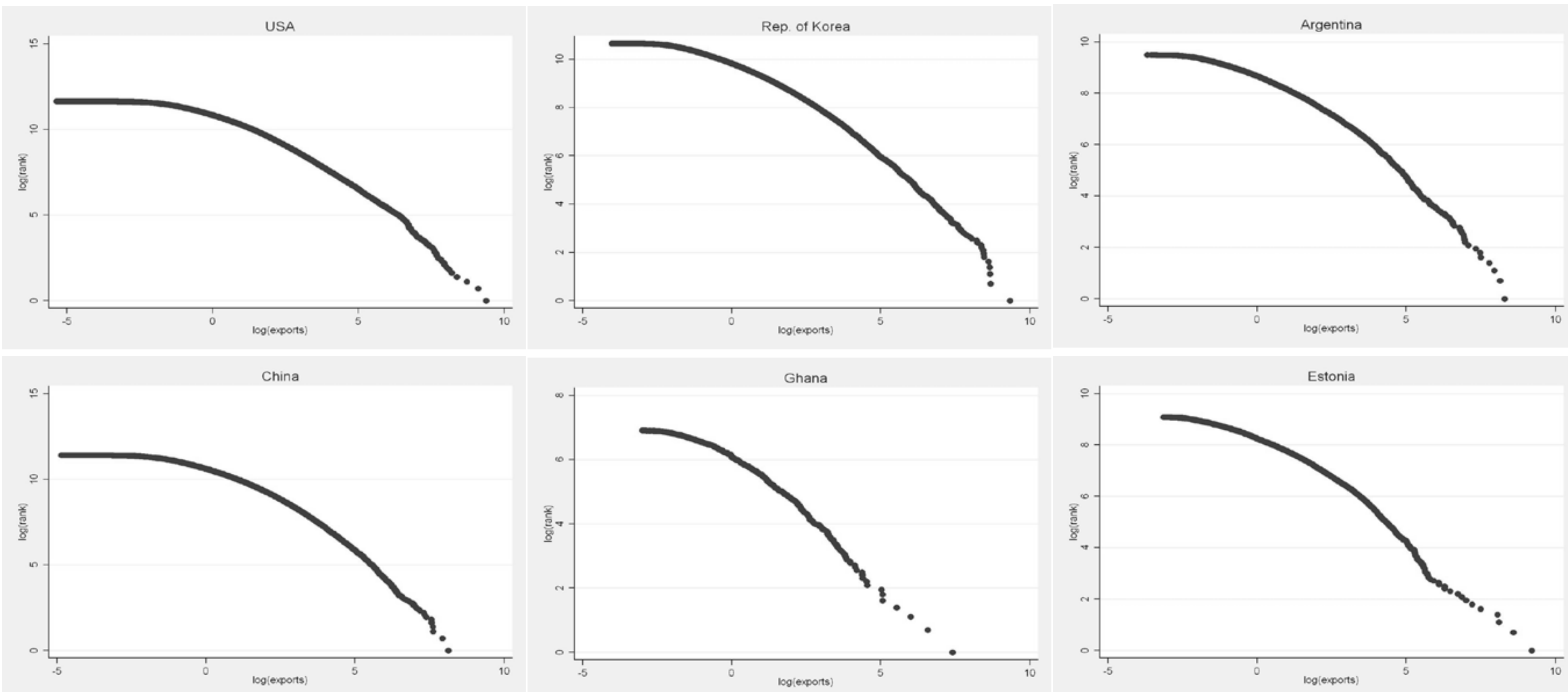
Notes: Each observation is a country. Export Flows is the number of nonzero product-destination categories that a country exports. Export Value is the log of total export value that a country realizes. The data are residuals from regressions on log GDP and log GDP per capita, both corrected for PPP. Source: U.N., Comtrade, World Bank, World Development Indicators.

Figure 5: Predicted versus Actual Nonzero Export Bilateral Flows



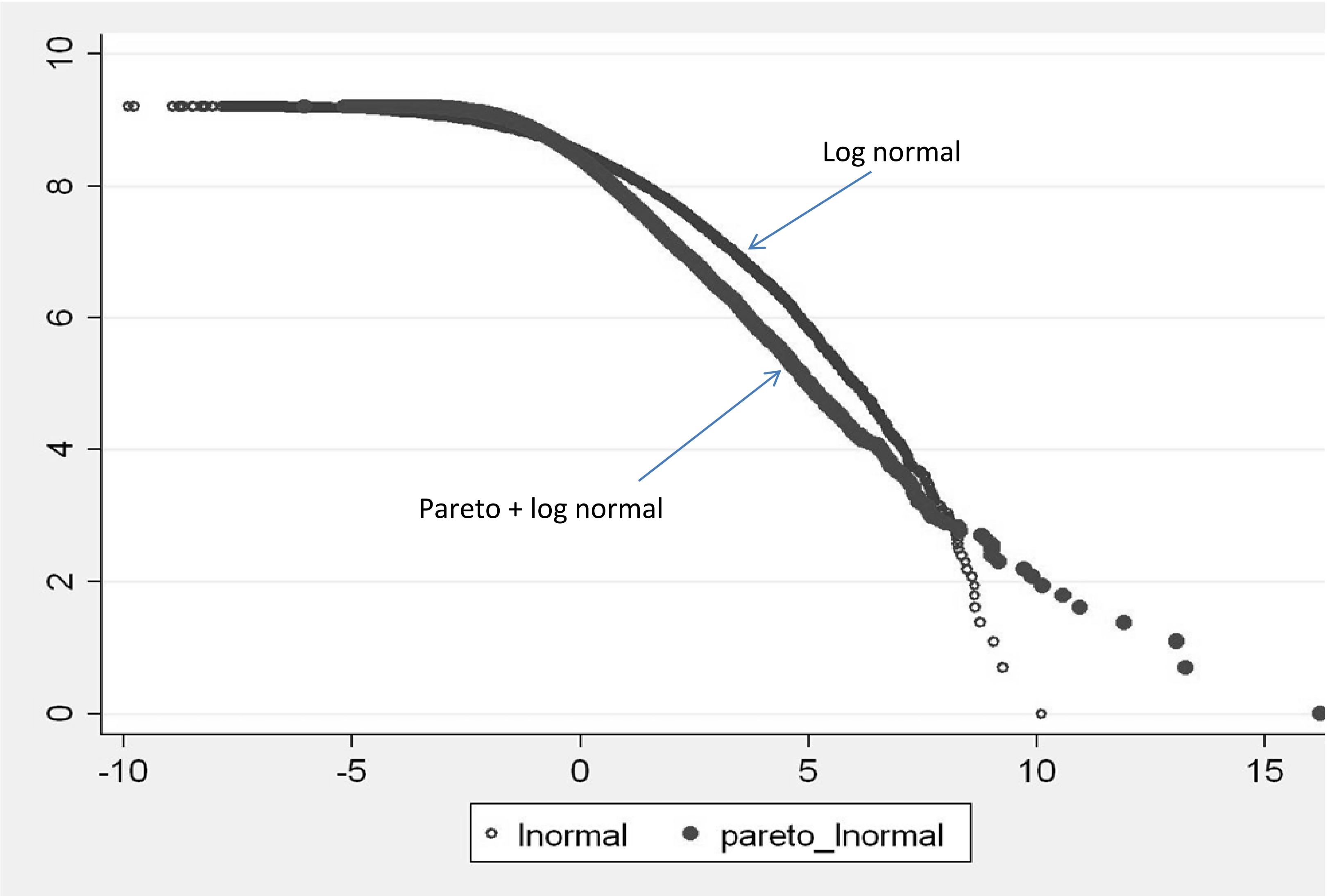
Notes: Log scale. Actual percent of nonzero bilateral export flows is the ratio of actual flows to potential flows times 100. Potential flows for each country are given by the number of export industries times 215; 215 is the maximum number of destinations in the data. Predicted nonzero flows are calculated from the Balls and Bins (random assignment) model; the number of shipments for each country is the number of actual bilateral flows times a factor that is equal to the average number of U.S. shipments per nonzero bilateral flow. This factor makes the U.S. predicted percent equal to 32%, as in Armenter and Koren (2009); see additional details in the text. The predicted percent is the predicted number of nonzero flows divided by potential times 100. The straight line is at 45 degrees. Sample: 157 countries. Montserrat was dropped from the figure as an outlier: its actual nonzero flows were 0.64% out of potential, while the predicted was 0.013%.

Figure 6: Log Export Rank and Log Export Value



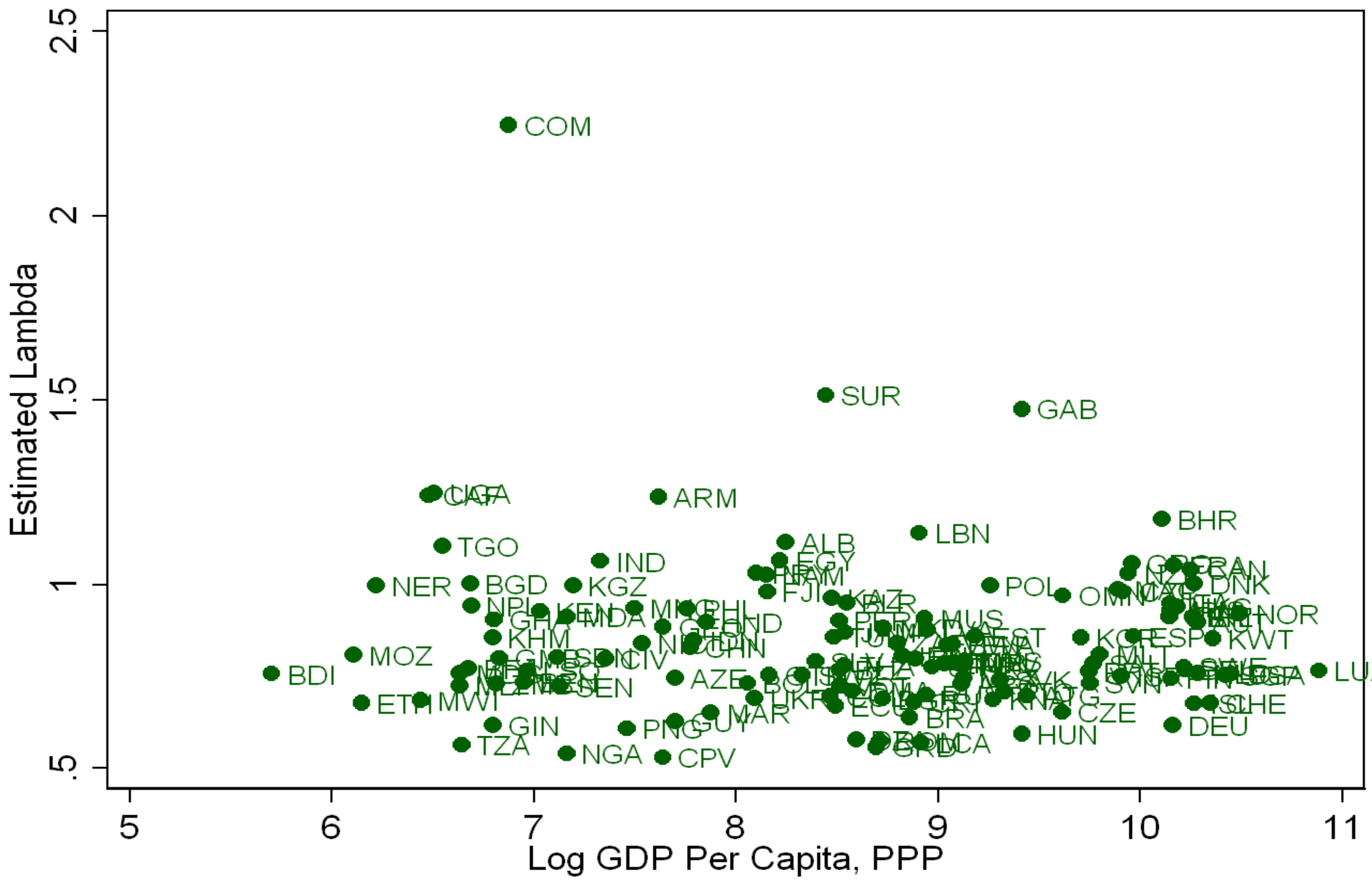
Notes: log(exports) is the log of bilateral product-destination export value. Log(rank) is the log of the rank of the product-destination export value. Source: U.N. Comtrade.

Figure 7: Simulated Rank Graphs for Log Normal and mixed Pareto-Log Normal



Notes: The simulation for the log normal uses the empirical standard deviation of export values averaged over all 151 countries. The the mixed Pareto-log normal is the sum of a Pareto and log normal random variables. The simulation uses the average estimated coefficients and standard deviations for all 151 countries from the estimation results below.

Figure 8: Estimates of Exponential exponent, Lambda



Notes: Lambda are estimated exponential exponent parameters. GDP per capita corrected for purchasing power parities were obtained from the World Bank World Development Indicators for 2000.