The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs*

Alfonso Irarrazabal†, Andreas Moxnes‡, and Luca David Opromolla§

July 2011

Abstract

International economics has overwhelmingly relied on Samuelson's (1954) assumption that trade costs are proportional to value. We develop a quantitative analytical framework that features both additive and multiplicative (iceberg) trade costs, building on a model of international trade with heterogeneous firms. We structurally estimate the magnitude of additive trade costs, for every product and destination available in our firm-level data of Norwegian exporters. Identification is aided by the theoretical finding that the elasticity of demand to producer price is dampened, in absolute value, when prices are low, and this mechanism is magnified when additive trade costs are high. This magnification mechanism becomes useful in the subsequent econometric analysis. Estimated additive trade costs are substantial. On average, additive costs are 33 percent, expressed relative to the median price. This leads us to reject the pure iceberg cost assumption. We assess the importance of these costs in shaping aggregate world trade flows. Interestingly, our micro estimates of additive trade costs explain most of the geographical variation in aggregate trade, suggesting that the role of multiplicative

*This is a substantially revised version of a paper previously circulated under the title "The Tip of the Iceberg: Modeling Trade Costs and Implications for Intra-Industry Reallocation". Acknowledgements: We would like to thank Costas Arkolakis, Gregory Corcos, Don Davis, Rob Johnson, Samuel Kortum, Ralph Ossa, Arvid Raknerud, Andrés Rodríguez-Clare, Alexandre Skiba, Karen Helene Ulltveit-Moe, and Kjetil Storesletten for their helpful suggestions, as well as seminar participants at Banco de Portugal, Dartmouth, DIME-ISGEP Workshop, IIES Stockholm, ITSG-Bocconi Workshop, LMDG Workshop, LSE, NBER Summer Institute ITI2010, Norwegian School of Management, University of Oslo, and Yale. We thank Statistics Norway for data preparation and clarifications. We thank the project "European Firms in a Global Economy: Internal Policies for External Competitiveness" (EFIGE) for financial support. Alfonso Irarrazabal thanks the hospitality of the Chicago Booth School of Business where part of this research was conducted. The analysis, opinions, and findings represent the views of the authors, they are not necessarily those of Banco de Portugal.

†Norges Bank, alfonso.irarrazabal@norges-bank.no.

‡Dartmouth College, Department of Economics, andreas.moxnes@dartmouth.edu.

§Banco de Portugal, Research Department and Research Unit on Complexity and Economics (UECE), ldopromolla@bportugal.pt.

(iceberg) costs must be limited. An implication of our work is that inferring trade costs from standard gravity models suffers from specification bias, since these models by assumption assume away the role of additive trade costs.

JEL Classification: F10 *Keywords*: Trade Costs, Heterogeneous Firms, Exports.

1 Introduction

The costs of international trade are the costs associated with the exchange of goods and services across borders. Trade costs impede international economic integration and may also explain a great number of empirical puzzles in international macroeconomics (Obstfeld and Rogoff 2000). Since Samuelson (1954), economists usually model and estimate variable trade costs as iceberg (i.e. multiplicative) costs, implying that pricier goods are costlier to trade. Trade costs change the relative price of domestic to foreign goods and therefore alter the worldwide allocation of production and consumption. Gains from trade typically occur because freer trade allows prices across markets to converge.

In this paper we take a different approach. We depart from Samuelson's framework, modeling variable trade costs as comprising both a multiplicative (iceberg) and an additive part.[1] Multiplicative costs are defined as a constant percentage of the producer price per unit traded, while additive costs are defined as a constant monetary cost per unit traded (conditional on a product type, e.g. shoes).[2] Even though more expensive varieties of a given product may be costlier to export, those costs are presumably not proportional to the product price. For example, a \$200 pair of shoes will typically face much lower multiplicative costs (i.e. cost relative to producer price) than a \$20 pair of shoes.[3] A number of trade policy instruments also act like additive trade costs. According to the World Trade Organization (WTO), 19 percent of U.S. non-agricultural imports are subject to additive tariffs.[4] Quotas

[1] Trade costs are broadly defined to include "...all costs incurred in getting a good to a final user other than the production cost of the good itself" (Anderson and van Wincoop, 2004). This includes transportation costs, policy barriers, information costs, contract enforement costs, costs associated with the use of different currencies, legal and regulatory costs, and local distribution costs.

[2] We use the terminology additive costs throughout the paper. Per-unit or specific trade costs are also terms frequently used in the literature.

[3] According to UPS rates at the time of writing, a fee of \$125 is charged for shipping a one kilo package from Oslo to New York (UPS Standard). They charge an additional 1% of the declared value for full insurance. Supposing that each pair of shoes weighs 0.2 kg, the multiplicative shipping costs are in this case 126 ((25+0.01*20)/20) and 13.5 ((25+0.01*200)/200) percent for the \$20 and \$200 pair of shoes respectively.

[4] 2006 data from the WTO are presented in Table 6. We discuss the data in more detail in the appendix. Until the 1950's, two-thirds of dutiable U.S. imports were subject to additive tariffs. This proportion fell to less than 40 percent by the early 1970's (Irwin, 1998).

(through the imposition of a quota license price) also act like a additive tariff.[5] In the U.S. and the European Union, 9.5 and 15.1 percent of the Harmonized System (HS) six-digit subheadings in the schedule of agricultural concessions are covered by tariff quotas. Distribution costs are also partly additive costs (e.g. Corsetti and Dedola, 2005).

The presence of additive trade costs has important consequences when firms charge different prices. First, when trade costs are incurred additively, trade costs not only alter relative prices *across* markets but also relative prices *within* markets. For example, the \$200 pair of shoes becomes cheaper relative to the \$20 pair in the presence of a additive tariff. As a consequence, and as shown by Alchian and Allen (1964), additive costs alter relative consumption patterns both within and across markets. Second, falling prices in the manufacturing sector (e.g. due to productivity growth) increase effective trade costs, if not accompanied by falling prices in the transport sector (or falling nominal tariffs). This illustrates the simple point that it is real trade costs, and not nominal ones, that determine the extent of economic integration. Third, the elasticity of demand to producer (f.o.b., free on board) price is dampened, in absolute value, when prices are low, and this mechanism is magnified when additive trade costs are high. This magnification mechanism becomes useful in the subsequent econometric analysis.

The first contribution of this paper is therefore to present a model of international trade with heterogeneous firms, building on Melitz (2003), Chaney (2008) and Eaton et al. (2010), but that features both iceberg costs and additive variable trade costs, as well as fixed entry costs, and to explore the economic implications of such a model. The second contribution is to document new firm-level facts about the relationship between f.o.b. prices and the volume of export across markets, consistent with the presence of additive trade costs. The third contribution is to develop a quantitative framework, derived from a subset of the model, that allows us to estimate the magnitude of additive trade costs, for every product and destination in our firm-level data of Norwegian exporters. The methodology is reminiscent of a difference-in-differences approach, where trade costs are identified by comparing the

[5] Demidova et al. (2009) use a trade model with heterogeneous firms to analyze the behavior of Bangladeshi garments exporters selling their products to the EU and to the U.S. and facing quotas as well as other types of barriers. Khandelwal, Schott and Wei (2011) investigate the impact of quota removal on aggregate productivity in China.

difference in the elasticity of sales to f.o.b. price between low- and high price firms, for a particular product, across destinations. Our model and quantitative framework are robust to heterogeneity in demand shocks (quality) across producers within a narrowly defined sector.

Several strong results emerge from the empirical analysis. First of all, additive costs are pervasive. The weighted mean of additive trade costs, expressed relative to the median price, is 33 percent. Our estimates are strongly positively correlated with observable proxies of trade costs, such as distance and product weight per value.[6] The pure iceberg model is therefore rejected. Second, we show that our micro estimates of additive trade costs can explain a substantial share of the geographical variation in world aggregate trade flows. Specifically, additive trade costs alone can explain between 40 to 70 percent of the elasticity of aggregate trade to distance. This suggests that the role played by multiplicative (iceberg) trade costs must be substantially more limited than previously thought. An implication of our work is that inferring trade costs from standard gravity models suffers from specification bias, since these models by assumption assume away the role of additive trade costs.

1.1 Previous Literature

More flexible modeling of trade costs is not new in international economics. Alchian and Allen (1964) pointed out that additive costs imply that the relative price of two varieties of some good will depend on the level of trade costs, and that relative demand for the high quality good increases with trade costs ("shipping the good apples out"). More recently, Hummels and Skiba (2004) found strong empirical support for the Alchian-Allen hypothesis. Specifically, the elasticity of freight rates with respect to price was estimated to be well below the unitary elasticity implied by the iceberg assumption. Also, their estimates implied that doubling freight costs increases average free on board (f.o.b.) export prices by $80-141$ percent, consistent with high quality goods being sold in markets with high freight costs. However, the authors could not identify the magnitude of additive costs, as we do here. Furthermore, our methodology identifies all kinds of trade costs, whereas their paper is concerned with shipping costs exclusively. Lugovskyy and Skiba (2009) introduce

[6] Hummels and Skiba (2004) find that distance has a positive and significant impact on freight costs.

a generalized iceberg transportation cost into a representative firm model with endogenous quality choice, showing that in equilibrium the export share and the quality of exports decrease in the exporter country size.

Our work also relates to a recent paper by Berman, Martin, and Mayer (2011). They also introduce a model with heterogeneous firms and additive costs, but in their model the additive component is interpreted as local distribution costs that are independent of firm productivity. Their research question is very different, however, as their paper analyzes pricing to market and the reaction of exporters to exchange rate changes. They show that, in response to currency depreciation, high productivity firms optimally raise their markup rather than the volume, while low productivity firms choose the opposite strategy.

Our work also connects to the papers that quantify trade costs. Anderson and van Wincoop (2004) provides an overview of the literature, and recent contributions are Anderson and van Wincoop (2003), Eaton and Kortum (2002), Head and Ries (2001), Hummels (2007), and Jacks, Meissner, and Novy (2008). This strand of the literature either compiles direct measures of trade costs from various data sources, or infers a theory-consistent index of trade costs by fitting models to cross-country trade data.[7] Our approach of using the within-market relationship between f.o.b. prices and exports is conceptually different and provides an complimentary approach to inferring trade barriers from data. This is possible thanks to the recent availability of detailed firm-level data. Furthermore, whereas the traditional approach can only identify iceberg trade costs relative to some benchmark, usually domestic trade costs, our method identifies the absolute level of (additive) trade costs.

The rest of the paper is organized as follows. Section 2 presents the model and summarize its implications. Since the subsequent empirical framework is formulated conditional on a set of general equilbrium variables, we present only the features of the model that is relevant to the empirical work, and choose to close the model later in the paper. In Section 3 we describe the data and present some empirical patterns that are suggestive of the presence of additive trade costs. Section 4 lays out the econometric strategy and presents the baseline estimates as well as robustness checks. In Section 5 we complete the theory and describe

[7] Helpman, Melitz and Rubinstein (2008) develop a gravity model that controls both for firm heterogeneity and fixed costs of exporting and make predictions about the response of trade to changes in trade costs.

the full equilibrium. In Section 6 we calibrate the model and evaluate the importance of additive trade costs in shaping world trade flows. Finally, Section 7 concludes.

2 The Model

In this section, we present a stylized model of heterogeneous firms and international trade that features both iceberg and additive trade costs. We keep the model as parsimonious as possible with the purpose of showing that this simple modification has important consequences when firms are heterogeneous. In Section 2.4 we summarize a number of important implications of the model, among them that variation in f.o.b. prices translates into less variation in exports when additive trade costs are high. These properties of the model will become useful in the subsequent empirical analysis. Since calculating the general equilibrium of the model is not necessary for the empirical analysis, we choose to close the model later in the paper (see Section 5).

Compared to the previous literature (e.g. Melitz, 2003, Chaney, 2008 and Eaton, Kortum and Kramarz 2010), the model has two innovations. First, we introduce additive trade costs. Second, we have two layers of heterogeneity, demand shocks and productivity, that are potentially correlated.[8] Heterogeneity in demand shocks can be interpreted as heterogeneity in quality: higher values of the demand shock, resulting in higher demand for a given price, can be interpreted as being associated with higher quality (Khandelwal, 2011, Sutton, 1991). By allowing for a (positive) correlation between demand shocks and prices, we can account for the possibility that the largest exporters are not necessarily the lowest price firms.

2.1 The Basic Environment

We consider a world economy comprising N asymmetric countries. Each country n is populated by a measure L_n of workers. The economy consists of a differentiated goods

[8] In Eaton, Kortum and Kramarz (2010), demand shocks are uncorrelated with the productivity draws. We do not introduce entry shocks in the model, in contrast to Eaton, Kortum and Kramarz (2010), since the extensive margin is largely irrelevant for the identification of trade costs (see Section 4).

sector and a transport services sector (described in the next section). For expositional ease we do not label sectors, and present the model for a generic unspecified sector.[9]

Preferences across varieties of the differentiated product have the standard CES form with an elasticity of substitution $\sigma > 1$. Each variety enters the utility function with its own exogenous country-specific weight η_n. These weights represent firm- and destination-specific demand shocks. These preferences generate a demand function $A_n (p_n/\eta_n)^{1-\sigma}$ in country n for a variety with price p_n and demand shock η_n. The demand level $A_n \equiv \mu Y_n P_n^{\sigma-1}$ is exogenous from the point of view of the individual supplier and depends on total expenditure Y_n and the consumption-based price index P_n.

Finally, we assume that workers are immobile across countries, but mobile across sectors and that market structure in the differentiated sector is monopolistic competition.

2.2 Variable Trade Costs

Unlike much of the earlier trade literature, firms also have to incur an additive cost t_{in}, per unit output, in order to transfer a good from i to market n. In other words, technology is assumed to be Leontief, so additive trade costs are proportional to the quantity produced (not proportional to value).[10] In Section 5, we model how wages w_i and t_{in} are determined and assign a numeraire. For now it suffices to take as given the matrix of trade costs across countries. Additionally, the economic environment consists of a standard iceberg cost τ_{in}, so that τ_{in} units of the final good must be shipped in order for one unit to arrive. The presence of iceberg costs ensures that any positive correlation between product value and shipping costs is captured by the model.[11]

[9] In the econometric section, a sector is interpreted as a product group according to the harmonized system (HS) nomenclature, at the 8 digit level (HS8). A differentiated good within a sector is interpreted as a firm observation within an HS8 code.

[10] This is similar to Burstein, Neves and Rebelo (2003) and Corsetti and Dedola (2005), who assume that production and retailing are complements.

[11] Hummels, Lugovskyy, and Skiba (2009) find evidence for market power in international shipping. An extension of our model with increasing returns in shipping could generate lower additive trade costs for more efficient firms. In other words, additive trade costs would become more like iceberg costs, since they would be correlated with the price of the good shipped.

2.3 Prices

Firms are heterogeneous in terms of both their technology, associated with productivity z, and their set of destination-specific demand shocks $\{\eta_n\}_{n=1,\ldots,N}$. A firm in country i can access market n only after paying a destination-specific fixed cost f_{in}, in units of the *numéraire*. Given labor costs w_i and the variable trade costs t_{in} and τ_{in}, profits are[12]

$$x_{in}\left[p_{in} - w_i\tau_{in}/z - t_{in}\right] - f_{in},$$

where $x_{in} = A_n\eta_n^{\sigma-1}p_{in}^{-\sigma}$ is the *quantity* demanded. Given market structure and preferences, a firm with efficiency z maximizes profits by setting its consumer price as a constant markup over total marginal production cost,

$$p_{in} = \frac{\sigma}{\sigma-1}\left(\frac{w_i\tau_{in}}{z} + t_{in}\right). \tag{1}$$

Exploiting the relationship between consumer prices, p_{in}, and producer (f.o.b.) prices, $\tilde{p}_{in}$,

$$p_{in} = \tau_{in}\tilde{p}_{in} + t_{in}, \tag{2}$$

the producer price can be written as

$$\tilde{p}_{in} = \frac{\sigma}{\sigma-1}\left(\frac{w_i}{z} + \frac{t_{in}}{\sigma\tau_{in}}\right).$$

Note that the markup over production costs is no longer constant. All else equal, a more efficient firm will charge a higher markup, since the perceived elasticity of demand that such a firm faces is lower. In other words, the markup is higher for more efficient firms since, due to the presence of additive trade costs, a larger share of the consumer price does not depend on the producer price.

2.4 Model Implications

In this Section, we summarize a few properties of the theoretical framework. Among these, Proposition 1 will become particularly useful in the subsequent empirical analysis. The first two propositions describe the relationship between demand and producer prices and demand and additive trade costs. The last two propositions describe how relative prices across and within markets are affected by additive trade costs.

[12] As a convention, we assume that additive trade costs are paid on the "melted" output.

Proposition 1 *The (absolute value of the) elasticity of demand, with respect to the f.o.b. price, E, is dampened when additive trade costs constitute a large share of the price. Moreover, the elasticity is dampened more among low-price firms than high price firms as additive trade costs increase.*

The first part of the proposition can be seen analytically from

$$E = \left| \frac{\partial \ln x_{in}}{\partial \ln \widetilde{p}_{in}} \right| = \left| \frac{\partial \ln x_{in}}{\partial \ln p_{in}} \frac{\partial \ln p_{in}}{\partial \ln \widetilde{p}_{in}} \right| = \sigma \left(1 + \widetilde{t}_{in}/\tilde{p}_{in}\right)^{-1},$$
$$\text{where } \widetilde{t}_{in} \equiv t_{in}/\tau_{in}.$$

Due to CES preferences, the first elasticity is $\partial \ln x_{in}/\partial \ln p_{in} = -\sigma$. Due to the relationship between the consumer and producer prices, $p_{in} = \tau_{in}\tilde{p}_{in} + t_{in}$ (equation 2), the second elasticity is $\partial \ln p_{in}/\partial \ln \widetilde{p}_{in} = \left(1 + \widetilde{t}_{in}/\tilde{p}_{in}\right)^{-1}$. Without additive trade costs, the second elasticity is one. With positive additive trade costs, the elasticity is decreasing in $\widetilde{t}_{in}/\tilde{p}_{in}$. In other words, if additive trade costs constitute a large share of the price (t_{in} relative to $\tau_{in}\widetilde{p}_{in}$), the demand elasticity with respect to the f.o.b. price is low. The economic intuition is that, since additive trade costs constitute a larger share of the consumer price for low-price goods, a given percentage increase in the producer price translates into a smaller percentage increase in the consumer price and consequently a smaller percentage decrease in consumption.

The second part of the proposition can be seen analytically from

$$\frac{\partial E}{\partial \widetilde{t}_{in}} \frac{\tilde{t}_{in}}{E} = -\left(1 + \tilde{p}_{in}/\tilde{t}_{in}\right)^{-1} < 0,$$

holding producer prices constant.[13] In other words, the elasticity falls as $\widetilde{t}_{in}$ rises, and the decline is larger when f.o.b. prices are low. The economic intuition is that, for high price firms, a marginal increase in $\widetilde{t}_{in}$ has a small impact on the consumer price (see Proposition 2) and the impact on the elasticity is therefore small. In the empirical Section below we identify the magnitude of additive trade costs by exploiting this mechanism.

[13] Allowing producer prices to change in response to a rise in additive trade costs (due to endogenous markups) does not change this conclusion. Specifically, we get $\left(\partial E/\partial \widetilde{t}_{in}\right)\left(\widetilde{t}_{in}/E\right) = -\left[1 - (\sigma - 1)^{-1} \widetilde{t}_{in}/\tilde{p}_{in}\right]\left(1 + \widetilde{t}_{in}/\tilde{p}_{in}\right)^{-1}\left(\widetilde{t}_{in}/\tilde{p}_{in}\right)$, which, after inserting the optimal $\tilde{p}_{in}$, also turns out to be negative.

Proposition 2 *The (absolute value of the) elasticity of demand with respect to additive trade costs is higher for low price than high price firms.*

Analytically, we see this from

$$\left|\frac{\partial \ln x_{in}}{\partial \ln \widetilde{t}_{in}}\right| = \left|\frac{\partial \ln x_{in}}{\partial \ln p_{in}} \frac{\partial \ln p_{in}}{\partial \ln \widetilde{t}_{in}}\right| = \sigma \left(1 + \tilde{p}_{in}/\widetilde{t}_{in}\right)^{-1},$$

holding producer prices constant.[14] The second elasticity is now $\partial \ln p_{in}/\partial \ln \widetilde{t}_{in} = \left(1 + \tilde{p}_{in}/\widetilde{t}_{in}\right)^{-1}$. Hence, if additive trade costs constitute a large share of the price (which will be the case when prices are low), a percentage increase in additive trade costs has a big negative percentage impact on quantity sold. The economic intuition is simply that an increase in additive trade costs translates into a larger percentage increase in the consumer price for low price firms.

Proposition 3 *Relative consumer prices within a market are distorted in the presence of additive trade costs, but not in the presence of iceberg costs.*

Consider two different varieties with producer prices $\tilde{p}_{in} > \tilde{p}'_{in}$ sold in market n. Then

$$\frac{p_{in}}{p'_{in}} = \frac{\tilde{p}_{in} + \widetilde{t}_{in}}{\tilde{p}'_{in} + \widetilde{t}_{in}} > 1$$

and, holding producer prices constant,

$$\frac{\partial p_{in}/p'_{in}}{\partial \widetilde{t}_{in}} = -\frac{\tilde{p}_{in} - \tilde{p}'_{in}}{\left(\tilde{p}'_{in} + \widetilde{t}_{in}\right)^2} < 0$$

In other words, an increase in t_{in} reduces the consumer price of the high price variety relative to the low price variety. Under some regularity conditions about demand (see e.g. Hummels and Skiba, 2004), an increase in t_{in} raises relative consumption of the high price variety relative to the low price variety. This is the well-known Alchian-Allen effect. On the contrary, if $t_{in} = 0$ and $\tau_{in} > 0$, relative consumer prices equals relative producer prices, $p_{in}/p'_{in} = \tilde{p}_{in}/\tilde{p}'_{in}$, so that changes in iceberg costs do not affect relative demand.

[14] Allowing producer prices to change in response to a rise in additive trade costs (due to endogenous markups) does not change this conclusion. Specifically, we get $\left|\partial \ln x_{in}/\partial \ln \widetilde{t}_{in}\right| = \left[\sigma^2/\left(\sigma - 1\right)\right]\left[1 + \tilde{p}_{in}/\widetilde{t}_{in}\right]^{-1}$.

Proposition 4 *Relative consumer prices across markets are distorted in the presence of additive trade costs, and as product prices fall the distortion becomes larger.*

Consider two varieties, one produced and sold locally in n, the other exported from i to n, with consumer prices p_{nn} and p_{in}. Given that the producer price is $\tilde{p}$ for both varieties, we can write the relative consumer price

$$\frac{p_{in}}{p_{nn}} = \tau \left(1 + \widetilde{t}_{in}/\tilde{p}\right) > 1$$

A fall in the producer price $\tilde{p}$, e.g. due to technological improvements in the manufacturing sector, will magnify the relative price disadvantage of the imported variety. As a consequence, falling nominal prices in the manufacturing sector increases effective trade costs, if not accompanied by falling prices in the transport sector (or falling nominal tariffs). In other words, what matters for the degree of economic integration is technological progress in transport *relative to* other activities.[15]

3 Empirical Regularities

In this section, we present some empirical patterns that are suggestive of the presence of additive trade costs. In the next section, we move on to estimating them formally.

3.1 Data

The data consist of an exhaustive panel of Norwegian non-oil exporters in 2004. Data come from customs declarations. Every export observation is associated with a firm r, a destination n and product k, the quantity transacted x_{knr} and the total value.[16] We calculate f.o.b. prices $\widetilde{p}_{knr}$ by dividing total value by quantity (unit value). The product id k is based on the Harmonized System 8-digit (HS8) nomenclature, and there are $5,391$ active HS8 products in the data. 203 unique destinations are recorded in the data set.

[15] In Paul Krugman's blog post "A Globalization Puzzle" (http://krugman.blogs.nytimes.com/2010/02/21/a-globalization-puzzle), he hypothesizes that technological progress biased against transport can help explain the fall in trade in the inter-war period.

[16] Firm-product-year observations are recorded in the data as long as the export value is NOK 1000 ($\approx$ USD 148) or higher. The unit of measurement is kilos.

In 2004, 17,480 firms were exporting and the total export value amounted to NOK 232 billion ($\approx$ USD 34.4 billion), or 48 percent of the aggregate manufacturing revenue. On average, each firm exported 5.6 products to 3.4 destinations for NOK 13.3 million ($\approx$ USD 2.0 million). On average, there are 3.0 firms per product-destination (standard deviation 7.8). As we will see in section 4, our quantitative framework utilizes the relationship between f.o.b. price and export quantity across firms *within a product-destination pair*. In the formal econometric model, we therefore choose to restrict the sample to product-destinations where more than 40 firms are present.[17] In the robustness section, we evaluate the effect of this restriction by estimating the model on an expanded set of destination-product pairs. Extreme values of quantity sold, defined as values below the 1^{st} percentile or above the 99^{th} percentile for every product-destination, are also eliminated from the data set. All in all, this brings down the total number of products to 121 and the number of destinations to 21.[18]

3.2 Regularities

Our empirical strategy is to check the theoretical prediction in Proposition 1, namely that (the absolute value of) the demand elasticity is dampened by trade costs, and more so for firms charging low prices. The theoretical mechanism is that, since trade costs constitute a larger share of the consumer price for low-price goods, consumers respond less to changes in the producer price of low price goods than high price goods, and this effect is exacerbated when trade costs are higher. To this end, we perform a simple exercise to verify if the association between prices and quantities (in logs) is dampened more among low than high price firms as trade costs increase.[19] We regress export volume (x_{knr}) on a full set of interactions between f.o.b. price ($\widetilde{p}_{knr}$), distance ($Dist_n$) (as a proxy for trade costs)

[17] Also, the likelihood function is relatively CPU intensive, and this restriction saves us a significant amount of processing time.

[18] Exports to all possible combinations of these products and destinations amount to 26.2% of total export value. In the robustness Section below we consider an alternative sample that covers about 58.9% of total export value.

[19] As noted below, we identify the correlation between prices and quantities and not the demand elasticity. In Section 4 we show that the identification of additive trade costs does not rely on the identification of the true demand elasticity.

and a dummy equal to one if the price is above the product-destination median, $M_{knr} \equiv 1\left[\widetilde{p}_{knr} > median_r\left(\widetilde{p}_{knr}\right)\right]$,

$$\ln x_{knr} = \alpha + \left[\ln \widetilde{p}_{knr} \times \times \ln Dist_n \times \times M_{knr}\right]\beta + \varepsilon_{knr}$$

where $\times\times$ denotes the full set of interactions and β is the vector of coefficients. The relationship between f.o.b. price and quantity exported is

$$\frac{\partial \ln x_{knr}}{\partial \ln \widetilde{p}_{knr}} = \beta_1 + \beta_2 \ln Dist_n + \beta_3 M_{knr} + \beta_4\left(\ln Dist_n \times M_{knr}\right)$$

which is allowed to vary depending on distance from Norway (β_2), between low- and high price firms (β_3), and the interaction between the two (β_4). The important coefficient is the triple interaction term $\beta_4\left(\ln \widetilde{p}_{knr} \times \ln Dist_n \times M_{knr}\right)$, since this captures whether the change in elasticity as distance is increasing ($\partial^2 \ln x_{knr}/\partial \ln \widetilde{p}_{knr} \partial \ln Dist_n$, the empirical counterpart to $\partial E/\partial \widetilde{t}_{in}$ from Proposition 1) is different between low- and high price firms. Given that $\partial \ln x_{knr}/\partial \ln \widetilde{p}_{knr} < 0$, our theory suggests that $\beta_4 < 0$, so that (the absolute value of) the elasticity is dampened *more among low-price firms than high price firms* as trade costs increase (recall that $M_{knr} = 1$ denotes high price firms). We review the results for the triple interaction term in Table 1.[20]

Columns (1)-(4) use different sets of fixed effects to control for unobserved heterogeneity. Column (1) only includes destination fixed effects, column (2) also includes product fixed effects, column (3) instead includes firm-product fixed effects, while column (4) has product-destination fixed effects. The triple interaction term is negative and significant and is not varying much across specifications. Even when we only use variation within a given firm-product pair (column (3)), i.e. only compare prices within the same firm and same product, across destinations, the relationship is negative. We also tried including GDP and GDP/capita interactions in the regression, which yielded very similar results.

Since ε_{knr} is presumably correlated with prices, the estimated coefficients will not reflect the true demand elasticity. In the formal econometric model in section 4 we show that identification of additive trade costs does not rely on identifying the true demand elasticity. Here, we simply state that the association between prices and quantities (in logs) is dampened more among low than high price firms as trade costs increase.

[20] Results for all interaction terms available upon request.

Table 1: The association between f.o.b. price and export volume.

	(1)	(2)	(3)	(4)
$\ln \widetilde{p}_{knr} \times \ln Dist_n \times M_{knr}$	-0.04^a	-0.04^b	-0.03^b	-0.04^a
	(0.02)	(0.02)	(0.01)	(0.01)
Destination FE	Y	Y	Y	N
Product FE	N	Y	N	N
Firm-product FE	N	N	Y	N
Product-destination FE	N	N	N	Y
R^2	0.56	0.61	0.59	0.60
N	66,403	66,403	66,403	66,403

Notes: Robust standard errors in parentheses, clustered by (1) destination, (2) product, (3) firm-destination, (4) product-destination. Only product-destinations with more than 10 firms are included in the sample. Significance levels: a 1%; b 5%.

4 Estimating Trade Costs

In this section we structurally estimate the magnitude of trade costs, for every destination and every product in our dataset. We showed in Proposition 1 that variation in f.o.b. prices leads to relatively less variation in exports when prices are low, and that this pattern is magnified for higher levels of additive trade costs. It is this magnification mechanism that provides identification and that allows us to recover estimates of trade costs consistent with our model. The methodology is reminiscent of a difference-in-differences approach, where trade costs are identified by comparing the difference in the elasticity of the volume of exports to f.o.b. prices between low- and high-price firms, for a particular product, across destinations.

The econometric strategy consists of finding the expected export volume conditional on the producer price charged, and then minimizing the sum of squared residuals by nonlinear least squares.[21] This strategy has at least two merits. First, we are not required to simulate

[21] We choose to use data for export volume (quantities) instead of export sales for the following reasons. First, using quantities instead of sales avoids measurement error due to imperfect imputation of transport/insurance costs. Second, we avoid transfer pricing issues when trade is intra-firm (Bernard, Jensen and Schott 2006).

the full general equilibrium in order to obtain estimates of trade costs. Second, our estimator is more general than our theory. In particular, in the model, our assumption about CES preferences implies that mark-ups are constant. In the econometrics, however, there is no constraint on the mark-ups, since we always condition on observed f.o.b. prices.

Our methodology for estimating trade costs is very different from the earlier literature.[22] First, most studies model trade costs as iceberg exclusively, omitting the presence of additive costs. A notable exception is Hummels and Skiba (2004), who distinguish between them and find evidence for the presence of additive shipping costs.[23] However, they are not able to identify the magnitude of additive shipping costs. Also, compared to our work, they study freight costs exclusively, whereas we consider all types of international trade costs. Second, our methodology utilizes within product-destination, across firms, variation in exports and f.o.b. prices to achieve identification, whereas earlier studies typically utilize cross-country variation in aggregate (or product-level) trade. Third, whereas the traditional approach can only identify trade costs relative to some benchmark, usually domestic trade costs, our method identifies the absolute level of trade costs.[24]

4.1 Estimation

We employ a simple nonlinear least squares (NLS) estimator where the objective is to minimize the squared difference between expected export volume and actual export volume (in logs). We use the volume of exports instead of sales because using sales complicates the estimating equation considerably.[25] Export volume in the model is $x_{in} = A_n \eta_n^{\sigma-1} p_{in}^{-\sigma}$. Taking this to the data, we modify the expression in two ways. First, since the data is differentiated by products k, we make the demand shifter A product-destination-specific and the elasticity of substitution σ product-specific. Second, we allow for deviations from log linearity in the demand function by introducing a squared price term. The reason for

[22] Anderson and van Wincoop (2004) provide a comprehensive summary of the literature.

[23] They find an elasticity of freight rates with respect to price around 0.6, well below the unitary elasticity implied by the iceberg assumption on shipping costs.

[24] As will become clear below, we identify $w_i t_{in}/\tau_{in}$. Our preferred measure of additive trade costs is $w_i t_{in}/\tau_{in}$ relative to the observed median f.o.b. price.

[25] This occurs because f.o.b. sales are $\widetilde{p}_{in} x_{in} = A_n \eta_n^{\sigma-1} p_{in}^{-\sigma} \widetilde{p}_{in}$.

doing so will become clear in Section 4.2. The export volume expression then becomes

$$\ln x_{knr} = A_{kn} + \sigma_{1k} \ln p_{knr} + \sigma_{2k} \left(\ln p_{knr}\right)^2 + \ln \widetilde{\eta}_{knr}, \tag{3}$$

where σ_{1k} and σ_{2k} denote the polynomial price coefficients.[26] Subscripts k, n, and r denote HS-8 product, destination and firm, respectively (subscript i is dropped since Norway is always the source country). The demand shifter A_{kn} captures total expenditure and the price index of product k in market n. The demand shocks $\widetilde{\eta}_{knr} \equiv (\sigma - 1)\, \eta_{knr}$ can be systematically correlated with prices, as discussed in the theory section. We assume that this relationship is also approximated by a second order polynomial (in logs) plus statistical noise ε_{knr},

$$\ln \widetilde{\eta}_{knr} = a_{kn} + \phi_{1k} \ln p_{knr} + \phi_{2k} \left(\ln p_{knr}\right)^2 + \varepsilon_{knr}. \tag{4}$$

In sectors with a high degree of quality heterogeneity, we expect $\partial \ln \widetilde{\eta}_{knr} / \partial \ln p_{knr} > 0$, so that high-price firms on average get better demand shocks. We can then rewrite the demand equation as

$$\ln x_{knr} = A_{kn} + a_{kn} + \left(\sigma_{1k} + \phi_{1k}\right) \ln p_{knr} + \left(\sigma_{2k} + \phi_{2k}\right) \left(\ln p_{knr}\right)^2 + \varepsilon_{knr}. \tag{5}$$

The c.i.f. price p_{knr} is unobserved, but the f.o.b. price $\widetilde{p}_{knr}$ is observable in our data. We therefore substitute p_{knr} with $\widetilde{p}_{knr}$ using $p_{knr} = \tau_{kn}\widetilde{p}_{knr} + t_{kn}$. We also employ the approximation $\ln(1+x) \approx x$, which is reasonably accurate for $\widetilde{t}_{kn}/\widetilde{p}_{knr} \in [0, 1/2]$ (recall that $\widetilde{t}_{kn} \equiv t_{kn}/\tau_{kn}$). This allows us to difference out the product-destination specific intercept term.[27] Removing this nuisance parameter is important since the cost of minimizing the objective function, in terms of processing time, is prohibitive when nuisance parameters are present.[28] The resulting estimating equation is

$$\widehat{\ln x}_{knr} = \widetilde{\phi}_{1k} \left(\widehat{\ln \widetilde{p}_{knr}} + \widetilde{t}_{kn} \widehat{\widetilde{p}_{knr}^{-1}} \right) + \widetilde{\phi}_{2k} \left[\widehat{\left(\ln \widetilde{p}_{knr}\right)^2} + 2\widetilde{t}_{kn} \widehat{\widetilde{p}_{knr}^{-1} \ln \widetilde{p}_{knr}} + \widetilde{t}_{kn}^2 \widehat{\widetilde{p}_{knr}^{-2}} \right] + \widehat{\varepsilon}_{knr}, \tag{6}$$

[26] In the model, we had $\sigma_{1k} = -\sigma$ and $\sigma_{2k} = 0$.

[27] The constant term is $A_{kn} + a_{kn} + \left(\sigma_{1k} + \phi_{1k}\right) \ln \tau_{kn} + \left(\sigma_{2k} + \phi_{2k}\right) \left(\ln \tau_{kn}\right)^2$.

[28] As we show in the robustness section, the log approximation is also useful because the estimating equation encompasses the case where demand shocks are a function of c.i.f. prices (as in the baseline model), and the case where demand shocks are a function of f.o.b. prices.

where $\widetilde{t}_{kn}$ is our coefficient of interest, $\widetilde{\phi}_{1k} = \sigma_{1k} + \phi_{1k} + 2\widetilde{\phi}_{2k} \ln \tau_{kn}$, $\widetilde{\phi}_{2k} = \sigma_{2k} + \phi_{2k}$, and hats denote each variable's deviation from its mean. e.g. $\widehat{\ln x}_{knr} = \ln x_{knr} - (1/R_{kn}) \sum_r \ln x_{knr}$ with R_{kn} being the number of exporters in product-destination pair kn.

Finally, we decompose $\widetilde{t}_{kn}$ into product- and destination-specific fixed effects, $\widetilde{t}_{kn} = \beta_k b_n$, and normalize $\beta_1 = 1$.[29][30] This decomposition enables us to identify trade costs that are due to product and market characteristics, respectively. We then minimize the sum of squared residuals

$$O(\mathbf{\Psi}) = \sum_k \sum_{n \in S_k^1} \sum_{r \in S_{kn}^2} \widehat{\varepsilon}_{knr}^2$$

where S_k^1 is the set of destinations present for product k and S_{kn}^2 is the set of firms exporting to product-destination pair kn. The coefficient vector is then $\mathbf{\Psi} = \left(\beta_k, b_n, \widetilde{\phi}_{1k}, \widetilde{\phi}_{2k}\right)$, in total $3K + N - 1$ parameters.

A potential concern is that prices and quantities are determined simultaneously, so that the error term is correlated with the explanatory variables. Our estimator for $\widetilde{t}_{kn}$ is, however, robust to any supply side mechanisms that make $\widetilde{p}_{knr}$ endogenous. For example, assume that firms facing favorable demand shocks (ε_{knr}s) also charge higher prices. We could approximate this with the polynomial $\varepsilon_{knr} = \gamma_{1k} \ln p_{knr} + \gamma_{2k} (\ln p_{knr})^2 + v_{knr}$ where v_{knr} is an error term. In that case, the estimating equation would be similar to equation (6), the only difference being the interpretation of the slope parameters, which would take the form $\widetilde{\phi}_{1k} + \gamma_{1k}$ and $\widetilde{\phi}_{2k} + \gamma_{2k}$. In sum, even though the interpretation of the slope parameters would change, the estimate of $\widetilde{t}_{kn}$ would not. Intuitively, the slope coefficients are a mixture of various structural supply and demand side parameters and any particular element is not separately identified (e.g. σ_{1k}). Identification of the trade cost coefficient is instead based on systematic nonlinear deviations from this equilibrium relationship between

[29] The normalization is similar to the one adopted in the estimation of two-way fixed effects in the employer-employee literature (Abowd, Creecy, and Kramarz 2002). Note that even though β_k is estimated relative to some normalization, the estimate of $\widetilde{t}_{kn}$ is invariant to the choice of normalization. We also need to ensure that all products and destinations belong to the same mobility group. The intuition is that if a given product is sold only in a destination where no other products are sold, then one cannot separate the product from the destination effect.

[30] In the robustness Section below we check whether our estimates are sensitive to the trade cost decomposition $\widetilde{t}_{kn} = \beta_k b_n$. by estimating $\widetilde{t}_{kn}$ directly for all possible product-destination pairs.

price and quantity.

4.2 Identification of trade costs

The intuition behind identification can be explained by the following example. Assume that we have two products, feather (F) and stone (S) sold in two different destinations, Sweden (SE) and Japan (JP). Figure 1 shows f.o.b. prices and quantities for one particular numerical example.[31] $\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$ are identified by fitting the empirical model to the data (for each product) among high-price firms. For high-price firms, the slopes are roughly similar in both markets, as additive trade costs constitute a negligible share of their c.i.f. price. In other words, we get information about $\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$ by looking at high-price intervals where the slopes are fairly similar across markets.[32]

The product and destination fixed effects β_k and b_n are identified by the differences in the slopes for low-price firms across products (comparing F and S) and across markets (comparing SE and JP). For a given product (e.g. S), the elasticity may be nonconstant across the price interval for reasons other than trade costs (i.e. $\widetilde{\phi}_{2k} \neq 0$).[33] But, as we move to more remote markets, any dampening of the elasticity *that is specific to low-price firms* will be attributed to trade costs. The methodology is therefore reminiscent of a difference-in-differences approach, where trade costs are identified from the change in the difference in elasticities between low- and high price firms, as we move to more remote destinations. Defining E^m_{kn} the absolute value of the elasticity with respect to the f.o.b. price, for product-destination kn and for $m = H$igh or Low, we can express this double difference as $E^H_{kn} - E^L_{kn} - \left(E^H_{kn'} - E^L_{kn'}\right)$ for destinations n and n'.

In addition, identification is helped by the fact that the impact of additive trade costs

[31] We used the following values for the parameters: $\widetilde{\phi}_{1k} = 1$, $\widetilde{\phi}_{2k} = 0.1$, $b_{JP}/b_{SE} = 10$, $\beta_F/\beta_S = 5$.

[32] This can be easily seen by letting $\widetilde{t}_{kn} \longrightarrow 0$ in equation (6),

$$\widehat{\ln x}_{knr} = \widetilde{\phi}_{1k}\left(\widehat{\ln \widetilde{p}_{knr}}\right) + \widetilde{\phi}_{2k}\left[\widehat{(\ln \widetilde{p}_{knr})^2}\right] + \widehat{\varepsilon}_{knr}.$$

[33] Note that the inclusion of $\widetilde{\phi}_{2k}$ in the empirical model is important in order to allow non-constant slope coefficients. Without $\widetilde{\phi}_{2k}$, any deviation from log-linearity among low price firms would be attributed to additive trade costs. In practice though, the estimates of trade costs are fairly similar when estimating under the restriction that $\widetilde{\phi}_{2k} = 0$.

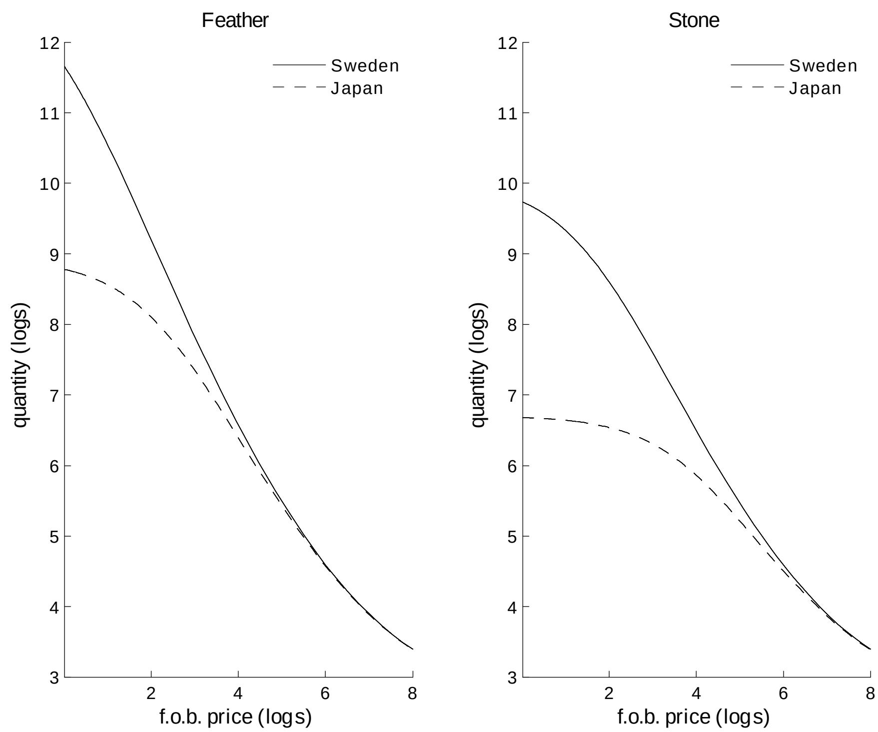

Figure 1: Identification.

on the f.o.b. price elasticity is highly non-linear. Specifically, an increase in trade costs for stone β_S, given the costs of exporting to JP relative to SE b_{JP}/b_{SE}, produces a larger percentage decline in the demand elasticity in JP than in SE. In other words, shipping stone to Japan instead of Sweden has a larger percentage impact on the elasticity than shipping feather to Japan instead of Sweden. Analytically, as we saw in Proposition 1,

$$\frac{\partial E}{\partial \widetilde{t}_{kn}} \frac{\tilde{t}_{kn}}{E} = -\frac{\tilde{t}_{kn}/\tilde{p}_{kn}}{1 + \widetilde{t}_{kn}/\tilde{p}_{kn}} < 0.$$

This shows that a one percent increase in additive trade costs produces a larger percentage fall in E if additive trade costs are already high ($\widetilde{t}_{kn}/\tilde{p}_{kn}$ high). This helps identification since, given information about the magnitude of b_{JP}/b_{SE}, a small change in e.g. β_S may make the fit of the model much better in JP without making it much worse in SE.

The empirical model controls for the degree of quality heterogeneity within an HS-8 product category. For instance, in sectors characterized by firms producing high quality and charging high prices, $\widetilde{\phi}_{1k}$ is a small negative number or possibly positive (so that higher prices are associated with more sales). As long as we control for differences in $\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$ across products, our estimate of $\widetilde{t}_{kn}$ is unbiased.

The essential identifying assumption is that the parameters governing the intersection between supply and demand ($\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$) are product specific and not product-destination specific (but the intercepts are allowed to be product-destination specific). Even though these two assumptions are not directly testable, it is difficult to explain the findings in this paper by product-destination specific variation in $\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$ exclusively. In particular, alternative theoretical explanations, e.g. demand side explanations, would need to reproduce the empirical finding that the f.o.b. price elasticity for low price firms falls faster than the elasticity for high price firms as trade costs (as proxied by distance) increases.

Furthermore, a model with firms varying their level of quality across markets (for a given product), perhaps due to country income differences such as in Verhoogen (2008), would not be able to reproduce the findings in this paper. In our framework, quality differences across markets would be captured by the constant term in the demand shock equation (4), which is differenced out in the estimating equation (6).

We also emphasize that, although our $\widetilde{t}_{kn}$ is assumed to be constant within an HS-8 product, across firms (e.g. same \$20 trade cost for all pairs of shoes exported to the U.S.), our framework allows for varying total trade costs across firms, within a product-destination pair. Recall that iceberg costs τ_{kn} is controlled for (subsumed into the intercept terms), even though not separately identified. Hence, any mechanism that would make $\widetilde{t}_{kn}$ vary systematically with product value would be subsumed into the intercept terms. This just shows that the $\widetilde{t}_{kn}$ that we identify is, by definition, the cost that is constant across all firms within a product-destination pair.

Finally, a comment about the interpretation of the results. Our methodology only allows identification of $\widetilde{t}_{kn} \equiv wt_{kn}/\tau_{kn}$ (and not wt_{kn}). When commenting on the magnitude of additive trade costs in Section 4.3, we divide the estimates of $\widetilde{t}_{kn}$ by the observed median f.o.b. price in product-destination kn, i.e. $TC_{kn} = (wt_{kn}/\tau_{kn})/\overline{\overline{p}}_{kn} = wt_{kn}/\left(\tau_{kn}\overline{\overline{p}}_{kn}\right)$. In other words, we measure additive trade costs relative to the f.o.b. price multiplied by the iceberg cost. As a consequence, our estimates of additive trade costs would be higher if we were to report $wt_{kn}/\overline{\overline{p}}_{kn}$ (and had information about τ_{kn}).

Table 2: Estimates of additive trade costs relative to f.o.b. price

	Weighted mean	Unweighted mean	Median	Std. deviation
Trade costs TC_{kn}	0.33	0.08	0.02	0.16
$\widetilde{\phi}_{1k}$	-2.76	-2.02	-1.40	3.06
$\widetilde{\phi}_{2k}$	-0.01	0.01	0.02	0.38
Criterion f	51,992			
# of Countries (N)	21			
# of Products (K)	121			

Note: The mean, median, and standard deviation of trade cost estimates are computed only over product-destination pairs where the f.o.b. price is non-missing. The weighted average is computed using export value weights.

4.3 Results

Given the estimates of b_n and β_k, we calculate trade costs relative to f.o.b. prices, $TC_{kn} = \widetilde{t}_{kn}/\overline{\overline{p}}_{kn}$, where $\overline{\overline{p}}_{kn}$ is the median f.o.b. price in product-destination pair kn. In Table 2 below, we report various moments of TC_{kn}.[34] The weighted average of additive trade costs is 0.33. The unweighted mean and median are smaller, indicating that many product-destination pairs with low point estimates of $\widetilde{t}_{kn}$ have small export volumes.

81 and 88 percent of the b_n and β_k coefficients (the destination and product fixed effects) are significantly different from zero at the 0.05 level. This suggests that, for the large majority of product-destination pairs, the null hypothesis of zero additive trade costs (i.e. a model with iceberg costs exclusively) is rejected.[35]

Figure 2 shows b_n for every destination on the vertical axis and distance (both in logs) on the horizontal axis. The left figure includes all destinations, whereas the right figure excludes destinations with insignificant b_n. Estimated trade costs are clearly increasing in actual trade costs, as proxied by distance.[36] Note that our two-way fixed effects approach means

[34] The estimates of β_k and b_n are available on the authors' homepages.

[35] We also test the hypothesis that all $\widetilde{t}_{kn} = 0$ formally. Let n_T be the number of observations, $\mathbf{\Psi}^{res}$ the vector of restricted coefficients (all $\widetilde{t}_{kn} = 0$), and $\mathbf{\Psi}^{unres}$ the vector of unrestricted coefficients. Then the likelihood ratio statistic $2n_T \left[O\left(\mathbf{\Psi}^{res}\right) - O\left(\mathbf{\Psi}^{unres}\right)\right]$, is $\chi^2(r)$ distributed under the null, where r is the $K + N - 1$ restrictions. The null is rejected at any conventional p-value.

[36] Freight costs are known to increase with distance, see e.g. Hummels and Skiba (2004).

that b_n is an index of trade costs in n that does not depend on the set of products traded in n. This implies that there is no selection bias in Figure 2 (e.g. that low $\widetilde{t}_{kn}$ products are sold in one destination and high $\widetilde{t}_{kn}$ products in another destination). According to our estimates, trade costs to e.g. the U.S. are roughly $3\frac{1}{2}$ times higher than trade costs to the Netherlands ($\exp(\ln b_{US} - \ln b_{NL})$). The robust relationship between distance and trade costs also emerges when regressing estimated trade costs b_n on a set of gravity variables (distance, GDP, and GDP per capita, all in logs). The distance elasticity is then 0.49 (s.e. 0.18).[37]

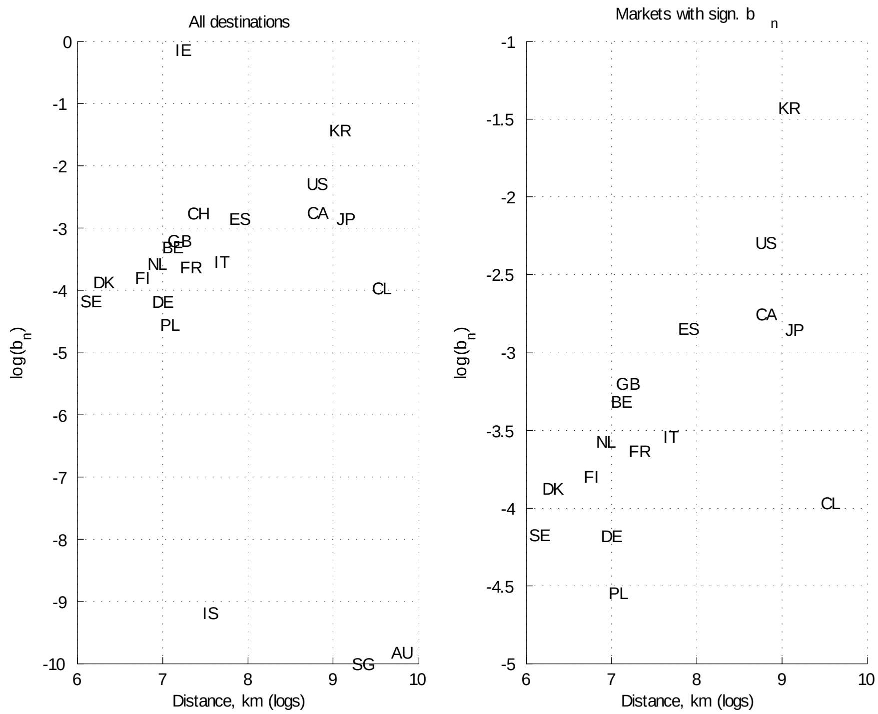

Figure 2: Estimates of b_n and distance (logs).

The top graphs in Figure 3 show the kernel densities of $\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$. The densities are centered around -1 and 0 respectively, suggesting that for the large majority of products, lower prices are associated with higher export volumes. The bottom graph in Figure 3 shows the kernel density of the product fixed effects $\ln \beta_k$. As expected, trade costs are

[37]The GDP and GDP/capita elasticities are not significantly differant from zero at the 0.05 level. The full set of results is available upon request.

quite heterogeneous. The 75/25 percentile ratio of β_k is 17.6.

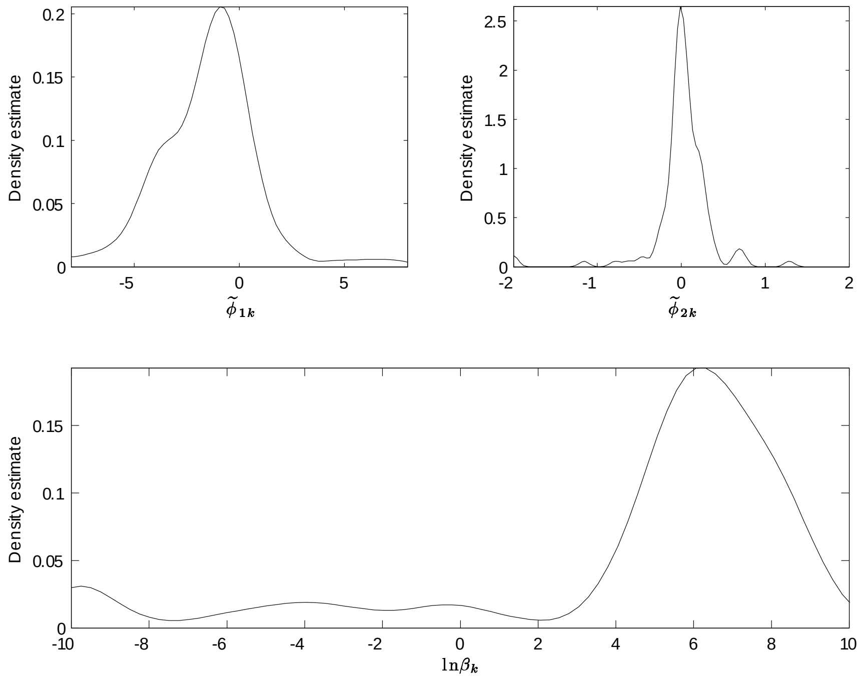

Figure 3: Estimates $\widetilde{\phi}_{1k}$, $\widetilde{\phi}_{1k}$ and $\ln\beta_k$. Kernel densities.

Figure 4 shows the relationship between TC_{kn}, averaged across destinations, and actual average weight/value (both in logs).[38] Since weight/value is presumably positively correlated with actual trade costs, we expect to see a positive relationship between these measures and estimated trade costs.[39] Indeed, the scatter indicate an upward sloping relationship, especially for high TC_{kn} products. The correlation is 0.18 (p-value 0.05).

Most of the estimates in the product dimension also make intuitive sense. For example, certain types of wooden furniture (HS 94016119) is among the products with estimated β_k above the 95^{th} percentile. Certain types of fish (HS 3022106) and computer accessories (HS 84716005) are among the products with estimated β_k below the 5^{th} percentile.[40]

[38] Average weight/value is obtained by dividing total weight (summed over firms and destinations) over total value (summed over firms and destinations). Average trade costs per product is unweighted. Using a weighted average instead produces similar results.

[39] Note that average TC_{kn}, and not β_k, is the proper measure of trade costs relative to price.

[40] Specifically, HS 94016119 = upholstered seats, with wooden frames (excl. convertible into beds), HS 3022106 = fresh or chilled lesser or greenland halibut, atlantic halibut and pacific halibut, HS 84716005 = input or output units for digital automatic data-processing machines, whether or not containing storage

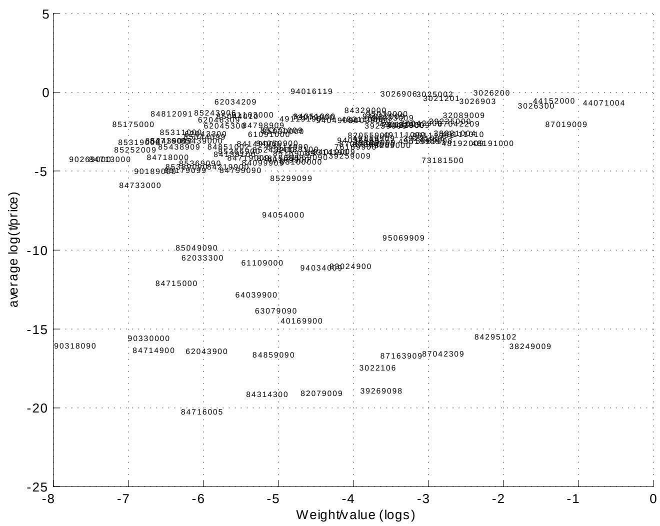

Figure 4: Estimates of average TC_{kn} and weight/value (in logs).

The decomposition of product and destination effects also allows us to study whether costly destinations are associated with products with lower trade costs. Or in other words, that the product mix in a given destination is a selected sample influenced by the costs of shipping to that market. A simple indicator is the correlation between the destination fixed effect b_n and the product fixed effect, averaged over the products actually exported there. Formally, we correlate b_n with $(1/K_n)\sum_{k\in\Omega_n}\beta_k$, where K_n is the number of products exported to destination n and Ω_n is the set of products exported to n. The results indicate that there is not much support for the hypothesis. The correlation is roughly zero.

The overall fit of the model is adequate, with an R^2 of 0.44. As a further check on the performance of the model, we plot normalized actual export volume and prices ($\widehat{\ln x}_{knr}$ and $\widehat{\ln \widetilde{p}_{kn}}$) as well as the conditional expectation of export volume for a few product-destination pairs. In Figure 5, we have chosen all export destinations for product HS 73269000, one of the top products in terms of export value.[41] The solid markers represent the conditional expectation whereas 'x' markers represent the data. F.o.b. prices are on the horizontal axis and export volume on the vertical axis (in logs). We observe that the model is able to capture a substantial share of the variation in the data, and especially the fact that the

units in the same housing.

[41] Articles of iron or steel, excl. cast articles or articles of iron or steel wire.

slope flattens out when prices are low (exactly what we would expect from Proposition 1).

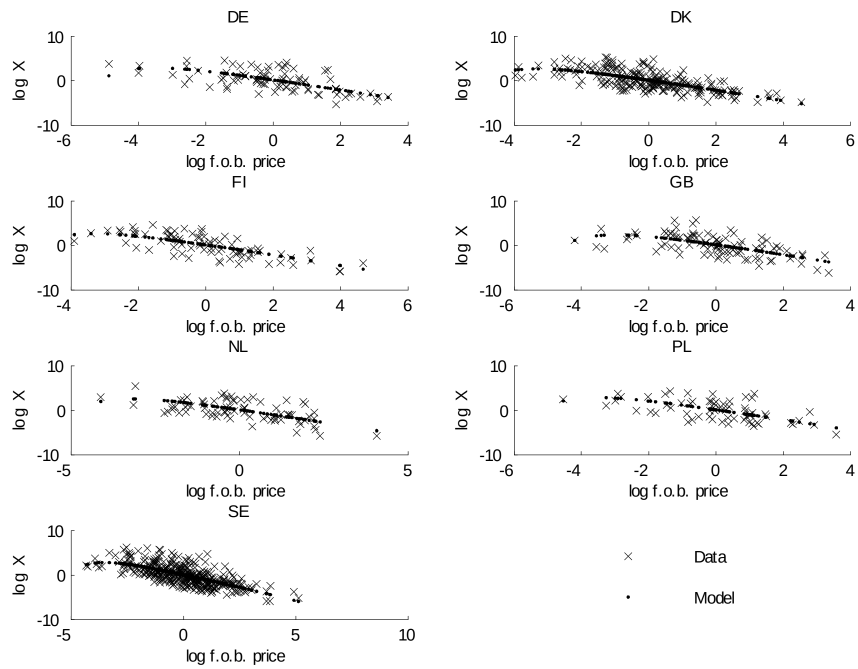

Figure 5: Predicted and actual export volume (normalized). HS73269000.

4.4 Robustness

In our baseline specification, we model demand shocks as a function of c.i.f. price. One implication of this modeling choice is that higher trade costs will, on average, produce better demand shocks. Alternatively, we could assume that demand shocks are a function of f.o.b. price. Here we show that the resulting econometric model in this case remains largely unchanged, except for a slight change of interpretation of the parameters. If $\ln\widetilde{\eta}_{knr} = \phi_{1k}\ln\widetilde{p}_{knr} + \phi_{2k}\left(\ln\widetilde{p}_{knr}\right)^2 + \varepsilon_{knr}$, then the estimating equation (6) can be rewritten as

$$\widehat{\ln x_{knr}} = \widetilde{\phi}_{1k}\left(\widehat{\ln\widetilde{p}_{knr}} + \frac{\sigma_{1k}}{\widetilde{\phi}_{1k}}\widetilde{t}_{kn}\widehat{\widetilde{p}_{knr}^{-1}}\right) + \widetilde{\phi}_{2k}\left[\widehat{(\ln\widetilde{p}_{knr})^2} + 2\frac{\sigma_{2k}}{\widetilde{\phi}_{2k}}\widetilde{t}_{kn}\widehat{\widetilde{p}_{knr}^{-1}\ln\widetilde{p}_{knr}} + \frac{\sigma_{2k}}{\widetilde{\phi}_{2k}}\widetilde{t}_{kn}^{2}\widehat{\widetilde{p}_{knr}^{-2}}\right] + \varepsilon_{knr}$$

The only difference compared to the baseline specification in equation (6) is that the trade cost coefficient is now multiplied by the factors $\sigma_{1k}/\widetilde{\phi}_{1k}$ and $\sigma_{2k}/\widetilde{\phi}_{2k}$. Given a guess of σ_{1k}, and assuming that $\sigma_{1k}/\widetilde{\phi}_{1k} = \sigma_{2k}/\widetilde{\phi}_{2k}$, we can easily recalculate average trade costs by multiplying our baseline estimate with $\widetilde{\phi}_{1k}/\sigma_{1k}$. In Table 3, column R1, we report the

Table 3: Robustness: Alternative specifications

	Shocks as a function of f.o.b. prices (R1)	Separate estimations for each product (R2)	Product-destinations with ≥ 20 firms (R3)
Trade costs,			
weighted mean	0.42	0.42	0.66
unweighted mean	0.08	0.12	0.25
median	0.01	0.04	0.06
std. deviation	0.31	0.39	0.78
# product-destinations	270	270	917
# of countries (N)	21	21	33
# of products (K)	121	121	378

mean, median and standard deviation of trade costs under $\sigma_{1k} = -4$ for all k. Weighted average trade costs are in this case 42 percent of the median f.o.b. price. Decreasing σ_{1k} to -8 lowers the average to 21 percent. The relative magnitude of trade costs (i.e. across destinations or products) is not affected by this change of the model.

In the next columns of Table 3 we present some re-estimations of the model that address several issues. First, we check whether our estimates are sensitive to the trade cost decomposition $\widetilde{t}_{kn} = \beta_k b_n$. We instead estimate $\widetilde{t}_{kn}$ directly for all possible product-destination pairs. Since there are no longer any interlinkages between different products, we minimize the objective function product by product. As shown in column R2, the results are fairly similar compared to the baseline case.

We also investigate whether the choice of truncating the data set to only product-destinations with more than 40 firms affects the results. We choose product-destinations with more than 20 firms present, resulting in 33 destinations and 378 products.[42] The increase in product-destination pairs now makes joint estimation infeasible, so we proceed by estimating product by product, as above. The estimate of average weighted trade costs increases to 66 percent, as shown in column (R3). The unweighted average increases more moderately from 12 to 25 percent.

[42] Exports to all possible combinations of these products and destinations amount to 58.9% of total export value.

Finally, firms are not randomly entering into different product-destinations and this can create a correlation between prices and the error term. We hypothesize that the correlation is positive, since firms with both bad demand shocks and high prices are not exporting. Analogous to the case with endogenous prices, described in the identification section, such a selection effect would only affect the slope parameters $\widetilde{\phi}_{1k}$ and $\widetilde{\phi}_{2k}$, and not the estimates of trade costs. We refer the reader to the appendix for further details.

5 Equilibrium

Our exposition has, so far, highlighted the relationship between producer prices and sales, taking input costs w_i and t_{in}, consumer prices p_{in}, as well as the CES price index P_i and the set of entry hurdles $z_{in}(\eta_n)$ as given. In this section, we determine the full equilibrium. This will become useful in the last part of the paper, when we calibrate the model and calculate simulated trade flows, in order to assess the importance of additive trade costs in shaping aggregate trade flows.

5.1 Input Costs

First we turn to the determination of the additive costs t_{in}. The economic environment consists of a transport sector, whose services are used as an intermediate input in final goods production. Similar to the assumption about a frictionless homogeneous good sector in e.g. Chaney (2008), transport services are freely traded and produced under constant returns to scale.

$\varphi_m T_{in}$ units of labor are necessary for transferring one unit of a good from a plant in i to its final destination in n, using shipping services from country m. The sector is perfectly competitive, so there is a global shipping service price $w_m \varphi_m T_{in}$ for each route, where w_m is the wage in country m. Relative wages between any two pair of countries i and n are then pinned down in all markets, as long as each country produces the shipping service, and are equal to $w_i/w_n = \varphi_n/\varphi_i$. By normalizing the price on a particular shipping route to one, say from i to n, all nominal wages are pinned down. The additive trade cost is then defined as $t_{in} \equiv w_l \varphi_l T_{in} = w_m \varphi_m T_{in}$, $\forall\ l, m$ (i.e. same cost irrespective of the nationality of the shipping supplier).

5.2 Entry and Cutoffs

We assume that the total mass of potential entrants in country i is $\kappa w_i L_i$ so that larger and wealthier countries have more entrants.[43] This assumption, as in Chaney (2008), greatly simplifies the analysis and it is similar to Eaton and Kortum (2002), where the set of goods is exogenously given. Without a free entry condition, firms generate net profits that have to be redistributed. Following Chaney, we assume that each consumer owns w_i shares of a totally diversified global fund and that profits are redistributed to them in units of the *numéraire* good. The total income Y_i spent by workers in country i is the sum of their labor income $w_i L_i$ and of the dividends they earn from their portfolio $w_i L_i \pi$, where π is the dividend per share of the global mutual fund.

Firms will enter market n only if they can earn positive profits there. Some low productivity firms may not generate sufficient revenue to cover their fixed costs. We define the productivity threshold $\bar{z}_{in}(\eta_n)$ from $\pi_{in}(\bar{z}_{in}, \eta_n) = 0$, as the lowest possible productivity level consistent with non-negative profits in export markets, conditional on a demand draw η_n,

$$\bar{z}_{in}(\eta_n) = \begin{cases} w_i \tau_{in} \left[\lambda_1 \left(\frac{f_{in}}{\eta_n Y_n} \right)^{1/(1-\sigma)} P_n - t_{in} \right]^{-1} & \text{if } t_{in} < \bar{t}_{in}, \\ \infty & \text{if } t_{in} \geqslant \bar{t}_{in}, \end{cases} \tag{7}$$

with $\bar{t}_{in} = \lambda_1 \left[f_{in} / \left(\eta_n Y_n \right) \right]^{1/(1-\sigma)} P_n$, and λ_1 a constant.[44]

In the presence of finite additive trade costs, even the most productive firm receives finite revenues that may not be sufficient to cover the entry cost in market n. Therefore, under some parameter values, the entry hurdle can be infinite, opening up the possibility of zero trade flows between country-pairs. Note that, unlike in Helpman, Melitz, and Rubinstein (2008), zero trade flows will emerge without imposing an upper bound on productivity levels. Also note that, unlike Eaton, Kortum, and Sotelo (2011), zero trade flows will emerge without assuming a finite integer number of firms.

[43] $\kappa > 0$ is a proportionality constant.

[44] Specifically, $\lambda_1 = (\sigma/\mu)^{\frac{1}{1-\sigma}} (\sigma - 1)/\sigma$.

5.3 Price Levels

Productivity and demand shocks in market n are drawn from a joint distribution with density $f(z, \eta_n)$. We do not impose any particular assumptions on $f(.)$ now. E.g. z and η_n may be negatively correlated, so that high cost firms (low z firms) on average draw better demand shocks.[45] The price index is then

$$P_n^{1-\sigma} = \sum_i \kappa w_i L_i \int \int_{\bar{z}_{in}(\eta_n)}^{\infty} \left(p_{in}(z) / \eta_n\right)^{1-\sigma} f(z, \eta_n)\, dz d\eta_n. \tag{8}$$

We can summarize an equilibrium with the following set of equations:

$$\begin{aligned} P_n &= g(P_n, \pi) \ \forall\, n \\ \pi &= h(\pi, P_1, .., P_N) \end{aligned}$$

The first equation states that the price index is a function of itself (since $\bar{z}_{in}(\eta_n)$ is a function of P_n) and the dividend share π (since $\bar{z}_{in}(\eta_n)$ is a function of Y_n which is a function of π). The second equation states that the dividend share is a function of itself and all price indices. We show why this is so in the Appendix.

6 Implications for Aggregate Trade Flows

In this section, we ask what our trade cost estimates imply for aggregate trade flows. Specifically, we ask to what extent our micro-level estimates are able to explain the macro trade elasticity, i.e. the aggregate impact of trade barriers on trade flows. This enables us to assess the importance of additive trade costs in shaping aggregate trade flows. Moreover, we can quantify the relative importance of additive versus multiplicative (iceberg) trade costs in shaping trade flows. For instance, if the macro trade elasticity is fully explained by our micro-level estimates of additive trade costs, then the role of multiplicative trade costs in explaining the trade elasticity must be limited.

Our methodology is as follows. First, from aggregate trade data, we calculate the *actual* elasticity of trade flows with respect to variable trade barriers (proxied by distance). Second,

[45] Baldwin and Harrigan (2011) and Johnson (2010) find evidence for a positive correlation between costs and demand shocks.

we calculate the general equilibrium from our model, given a set of parameters (some to be calibrated, others based on our micro-level estimates as well as on the previous literature). Third, we estimate the elasticity of *simulated* trade flows with respect to variable trade barriers. Our objective is to match the simulated and actual trade elasticity.

We make one simplification compared to the more general model, by assuming, as in Baldwin and Harrigan (2008) and Johnson (2010), that demand shocks are related to prices according to $\eta_{in} = p_{in}^{\psi}$. By linking demand shocks and prices in this manner, we can account for the possibility that the largest exporters are not necessarily the lowest price firms.[46] The function is simply intended to reflect a reduced form relationship that is observable in the data, and we show in the next paragraph how we can infer ψ from our micro estimates.[47] The model is then effectively recast to one dimension of heterogeneity (productivity), and we follow the literature and assume that productivities are distributed Pareto, with shape parameter γ and support $[1; +\infty)$. In the appendix, we derive the expressions for the price index, cutoffs and quantity sold under the restriction that $\eta_{in} = p_{in}^{\psi}$.

Next, we choose some baseline parameters from our micro estimates and from the previous literature. The parameters are summarized in Table 5. Specifically, the Pareto shape coefficient relative to the elasticity of substitution, $\gamma/(\sigma-1)$, is 2.46, as in Eaton, Kortum and Kramarz (2010). Fixed costs f are \$700,000 in 2004-prices, as in Das, Roberts and Tybout (2008). In the baseline specification, we simulate the model under three different values of the elasticity of substitution $\sigma = \{5, 7, 9\}$. From our micro estimates of Section 4, we have $mean\left(\widetilde{\phi}_{1k}\right) = -2.02$ and $mean\left(\widetilde{\phi}_{2k}\right) \approx 0$ (see Table 2). In the model, the elasticity of quantity sold with respect to c.i.f. prices is $\psi(\sigma-1) - \sigma$ (see the Appendix). Since $\widetilde{\phi}_{1k}$ (in Section 4) is the micro estimate of this elasticity, the average ψ is then $\psi = (mean(\widetilde{\phi}_{1k}) + \sigma)/(\sigma-1)$.[48] Finally, we assume that productivity in the transport sec-

[46] Additive trade costs have a larger negative impact on sales for low price firms. If low price firms have the largest market share, then variation in additive trade costs will have a larger impact on aggregate trade flows compared to when low price firms have a smaller market share.

[47] I.e. we do not model the possibility of firms choosing higher quality subject to a cost, but $\eta_{in} = p_{in}^{\psi}$ may be a reduced form outcome of such a process.

[48] $\widetilde{\phi}_{1k}$ is defined as $\sigma_{1k} + \phi_{1k} + 2\widetilde{\phi}_{2k} \ln \tau_{kn}$. Given $\widetilde{\phi}_{2k} = 0$, we get $\widetilde{\phi}_{1k} = \sigma_{1k} + \phi_{1k}$, which is the elasticity with respect to c.i.f. price (see equation 5).

tor φ_i^{-1} is identical across countries, so that wages are also identical, normalized to 1. This assumption will have a negligible impact on our results, since all country-specific variation will be controlled for by country fixed effects (see below).

We follow the literature (e.g. Anderson and van Wincoop, 2004), and let $\tau_{in} = Dist_{in}^{\rho_1}$, where ρ_1 is a parameter to be estimated and $Dist_{in}$ is distance in kilometers from i to n, normalized relative to $\min_n (Dist_{in})$.[49] Our micro estimates suggest that the elasticity of t_{in}/τ_{in} with respect to distance is $\rho_2 = 0.49$, with GDP and GDP/capita insignificant (see Section 4.3). We therefore let $t_{in}/\tau_{in} = \theta Dist_{in}^{\rho_2} \iff t_{in} = \theta Dist_{in}^{\rho_1+\rho_2}$, where θ is a parameter that we will calibrate. Finally, the simulation relies on data on income $w_i L_i$, as proxied by PPP-based GDP from Penn World Table 6.2, and data on distance, from CEPII (2008). We denote the set of fixed parameters $\Theta = \{\sigma, \gamma, \psi, f, \rho_2\}$.

Before calibrating the model, we estimate the actual trade elasticity, from a standard gravity equation with exporter and importer fixed effects,

$$\ln S_{in} = a_i + b_j + \epsilon^{actual} \ln Dist_n + \varepsilon_{in} \tag{9}$$

where S_{in} is aggregate trade from i to n in 2004. We estimate (and simulate the model) on the same set of 22 countries that our micro estimates are based on. Trade data are gathered from CEPII (2008). The estimated ϵ^{actual} is -0.92 (s.e. 0.05), close to what is typically found in the literature (e.g. Anderson and van Wincoop, 2003).

Finally, we calibrate the model. There are three unknown variables, κ (the constant determining the number of potential entrants), θ (the constant determining the level of additive trade costs), and ρ_1 (the elasticity of τ_{in} with respect to distance). Calibration proceeds as follows:

1. Given Θ, choose some initial $\left(\theta^0, \rho_1^0, \kappa^0\right)$ and pin down the matrix of additive and multiplicative trade costs using $t_{in} = \theta Dist_{in}^{\rho_1+\rho_2}$ and $\tau_{in} = Dist_{in}^{\rho_1}$. Then simulate the full general equilibrium.

2. Calculate the simulated counterpart to the average trade cost estimate $\overline{TC}_{kn}$from Section 4.3. Specifically, the equilibrium unweighted average additive trade costs

[49] As is usual in gravity models, only the relative magnitude of τ_{in} can be identified, not the absolute magnitude. We therefore choose the normalization that $\tau_{in} = 1$ for the country pair in with $\min(Dist_{in})$.

(divided by τ_{in}) relative to the simulated median f.o.b. price for Norwegian exporters,

$$\overline{TC}_{NO}^{sim} = \frac{1}{N-1} \sum_{n \neq NO} \frac{t_{NOn}/\tau_{NOn}}{median\left(\widetilde{p}_{NOn}^{sim}\right)}$$

The median f.o.b. price in n is the simulated median f.o.b. price charged by Norwegian firms entering export market n.

3. Calculate the simulated trade elasticity ϵ^{sim} by estimating equation (9) on simulated trade data $\ln S_{in}^{sim}$.

4. Iterate (index r) over $(\theta^r, \rho_1^r, \kappa^r)$ until $\overline{TC}_{NO}^{sim} = \overline{TC}_{NO}^{actual} = 0.08$ (see Table 2), $\epsilon^{sim} = \epsilon^{actual} = 0.92$, and $\min\left(\overline{z}_{in}\right) = 1$.[50]

The results are summarized in Table 4. Under $\sigma = 5$, the calibrated value of ρ_1 is 0.055, suggesting that a doubling of distance increases τ_{in} by only 5.5 percent. This stands in sharp contrast to conventional gravity studies, where $\rho_1(\sigma - 1)$ is typically around 1.0 (Anderson and van Wincoop, 2004), giving $\rho_1 = 0.25$ when $\sigma = 5$. In other words, our results suggest that the impact of distance on multiplicative trade costs is roughly one fifth of what conventional estimates suggest. Increasing the elasticity of substitution σ to 7 lowers ρ_1 even more, and when $\sigma = 9$ our estimates show that ρ_1 is only about one tenth of the magnitude found in conventional gravity studies $(0.013/(1/8))$.

We also calibrate the model under the assumption that $\rho_1 = 0 \implies \tau_{in} = 1$ for all country pairs, and stop matching the trade elasticity moment ϵ^{actual} (but keep the other two moments). The goal is to understand to what extent additive trade costs alone can explain the macro trade elasticity. The results are shown in row 4 of Table 4. The actual trade elasticity is -0.92, while the simulated trade elasticity is in the range of -0.35 to -0.64, depending on the choice of the elasticity of substitution. This means that additive trade costs alone can explain 40 to 70 percent of the observed aggregate trade elasticity (0.35/0.92 to 0.64/0.92).[51] Varying the level of fixed costs f_{in}, the demand shock parameter ψ, or the Pareto shape parameter γ will only change ϵ^{sim} slightly. Under the baseline with $\sigma = 5$,

[50] κ will affect the $\overline{z}_{in}$ matrix since more potential entrants reduce the price index and increase the entry hurdles. The requirement that $\min\left(\overline{z}_{in}\right) = 1$ ensures that the extensive margin will be active in all markets (if $\overline{z}_{in} < 1$, changes in trade costs will not necessarily change the number of entrants).

[51] If fixing τ to a level larger than one (i.e. $\rho_1 > 0$), additive trade costs would explain a bigger share of the

Table 4: Estimates of simulated and actual trade elasticity.

σ	5		7		9	
Baseline:						
ρ_1	0.055		0.027		0.013	
θ	0.0037		0.0040		0.0042	
κ	0.300		0.250		0.200	
ϵ^{sim} under $\rho_1 = 0$[1]	-0.35	(0.01)	-0.49	(0.01)	-0.64	(0.01)
Actual elasticity ϵ^{actual}			-0.92(0.05)			

Notes: Standard errors in parentheses. [1] Calibrated with same θ and κ as in the baseline.

doubling f produces $\epsilon^{sim} = -0.38$ while setting $\psi = 0$ produces $\epsilon^{sim} = -0.39$. Doubling $\gamma/(\sigma-1)$ produces $\epsilon^{sim} = -0.56$.[52]

The results suggest that our micro estimates of additive trade costs explain a substantial share of the variation in aggregate trade flows. There are at least two implications of our findings. First, the role of multiplicative trade costs must be limited, since additive trade costs alone explain 40 to 70 percent of the trade elasticity. Second, estimating trade costs from standard gravity models suffers from specification bias, since these models assume away the existence of additive trade costs.

7 Conclusions

In this paper we develop a quantitative analytical framework that features both additive and multiplicative (iceberg) trade costs, building on a model of international trade with heterogeneous firms. An important property of the model, which we use in the subsequent empirical analysis, is that variation in f.o.b. prices translates into less variation in exports when prices are low, and that this mechanism is magnified with high additive trade costs. It is thus the marriage of additive costs and price heterogeneity that drives the theoretical and empirical results in this paper.

observed trade elasticity. This is due to the fact that $t_{in} = \theta Dist_{in}^{\rho_1+\rho_2}$, so that increasing ρ_1 will produce a higher elasticity of t_{in} with respect to distance.

[52] Note that in a model without additive trade costs, as in Chaney (2008), the trade elasticity is a function of γ but independent of σ. In our model, the trade elasticity depends on both γ and σ.

Table 5: Parameter values and data used in the simulation

Notation	Description	Value	Source
$\gamma/(\sigma-1)$	Pareto shape parameter (sales)	2.46	Eaton, Kortum and Kramarz (2010)
f	Fixed costs	0.7m USD 2004	Das, Roberts and Tybout (2008)[1]
$\psi(\sigma-1)-\sigma$	Elasticity of quantity sold with respect to c.i.f. prices	-2.02	Own micro estimates (Section 4)
ρ_2	Elasticity of t_{in}/τ_{in} with respect to distance	0.49	Own micro estimates (Section 4)
w_iL_i	Income	GDP (PPP adj. 2004)	Penn World Table 6.2
d_{in}	Distance		CEPII (2008)
S_{in}	Trade flows		CEPII (2008)

Notes: 22 countries and 10,000 draws used in simulation. [1]In their paper, the average cost of foreign market entry is estimated to be 0.4m in 1986-USD which is approximately 0.7m 2004-USD. We abstract from differences in fixed costs between country-pairs.

We structurally estimate the magnitude of additive trade costs, for every product and destination in our dataset, exploiting the nonlinearity in the relationship between f.o.b. prices and exports as predicted by the model. Our findings indicate that additive trade costs are on average 33 percent, expressed relative to the median price. We therefore conclude that pure iceberg costs are rejected, and that empirical work, especially at this level of disaggregation, must account for both the tip of the iceberg, as well as the part of trade costs that are largely hidden under the surface: additive costs. Furthermore, we show that our micro estimates are able to explain most of the geographical variation in aggregate world trade flows, suggesting that the role of multiplicative (iceberg) costs must be limited. An implication of our work is that inferring trade costs from standard gravity models suffers from specification bias, since these models by assumption assume away the role of additive trade costs.

Our analytical framework is potentially useful in a number of applications. For example, our analysis points to the need for further research in understanding the geographic response of aggregate trade flows to additive trade costs. Furthermore, our theoretical finding that uneven productivity growth in the manufacturing versus the transport sector may in fact dampen economic integration, suggests that modeling additive trade costs may enhance our understanding of the growth of trade. Finally, we saw that additive trade costs not only distorts prices across markets, but also within markets. Hence, investigating the welfare consequences of additive barriers (which may be very different from the standard case in Arkolakis, Costinot and Rodriguez-Clare, 2010) might be a fruitful avenue for future research.

References

[1] Abowd, John M., Robert H. Creecy, and Francis Kramarz, "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Cornell University Working Paper, 2002.

[2] Alchian, Armen A., and William R. Allen, *University Economics* (Belmont, CA: Wadsworth Publishing Company, 1964).

[3] Altonji, Joseph G., and Lewis M. Segal, "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 3 (1996), 353-366.

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

[5] Anderson, James E., and Eric van Wincoop, "Trade Costs," *Journal of Economic Literature*, 3 (2004), 691-751.

[6] Arkolakis, Konstantinos, "Market Penetration Costs and the New Consumers Margin in International Trade," NBER working paper No. 14214, 2008.

[7] Arkolakis, Konstantinos, Arnaud Costinot, and A. Rodríguez-Clare, "New Trade Models, Same Old Gains?," mimeo, 2010.

[8] Arkolakis, Konstantinos, Arnaud Costinot, and A. Rodríguez-Clare, "Gains From Trade under Monopolistic Competition: A Simple Example with Translog Expenditure Functions and Pareto Distributions of Firm-Level Productivity", mimeo, 2010.

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

[10] Berman, Nicolas, Philippe Martin, and Thierry Mayer, ‘´How Do Different Exporters React to Exchange Rate Changes? Theory, Empirics and Aggregate Implications,” Forthcoming in the *Quarterly Journal of Economics*, 2011.

[11] Bernard, Andrew B., Jonathan Eaton, Bradford J. Jensen, and Samuel Kortum, “Plants and Productivity in International Trade,” *American Economic Review*, 4(2003), 1268-1290.

[12] Bernard, Andrew, Bradford J. Jensen, and Peter Schott, “Transfer Pricing by US-Based Multinational Firms,” NBER Working Paper 12493, 2006.

[13] Broda, Christian, and David Weinstein, “Globalization and the Gains from Variety,” *Quarterly Journal of Economics,* 2 (2006), 541-585.

[14] Burstein, A. and João C. Neves and Sergio Rebelo, "Distribution costs and real exchange rate dynamics during exchange-rate-based stabilizations", Journal of Monetary Economics, 2003.

[15] Cameron A. Colin, and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, (New York, NY: Cambridge University Press, 2005).

[16] Chaney, Thomas, “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, 98 (2008), 1707–1721.

[17] Corsetti, Giancarlo and Dedola, Luca, 2005. "A macroeconomic model of international price discrimination," Journal of International Economics, Elsevier, vol. 67(1), pages 129-155, September.

[18] Delipalla, Sofia and Keen, Michael (1992), "The Comparison between Ad Valorem and Specific Taxation under Imperfect Competition", *Journal of Public Economics.*

[19] Demidova, Svetlana, Hiau Looi Kee, and Kala Krishna, “Do Trade Policy Di¤erences Induce Sorting? Theory and Evidence From Bangladeshi Apparel Exporters,” mimeo, 2009.

[20] Eaton, Jonathan, and Samuel Kortum, “Technology, Geography, and Trade,” *Econometrica*, 70 (2002), 1741-1779.

[21] Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, “An Anatomy of International Trade: Evidence from French Firms," NBER Working Paper No. 14610, 2008.

[22] Keith Head, and John Ries, “Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade,” *American Economic Review,* 4 (2001), 858-876.

[23] Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein. 2008. “Estimating Trade Flows: Trading Partners and Trading Volumes.” *Q.J.E.*, 123 (May): 441-487.

[24] Hummels, David, “Transportation Costs and International Trade in the Second Era of Globalization,” *Journal of Economic Perspectives,* 21 (2007), 131-154.

[25] Hummels, David, and Alexandre Skiba, “Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture,” *Journal of Political Economy*, 6(2004), 1384-1402.

[26] Hummels, David and Peter J. Klenow, “The Variety and Quality of a Nation’s Exports,” *American Economic Review*, 3 (2005), 704-723.

[27] Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba, “The Trade Reducing Effects of Market Power in International Shipping,” *Journal of Development Economics,* 1 (2009), 84-97.

[28] Irarrazabal, Alfonso, Moxnes Andreas, and Luca David Opromolla, “The Margins of Multinational Production and the Role of Intra-firm Trade,” CEPR Discussion Paper 7145, 2008.

[29] Irwin, Douglas A. "Changes in U.S. Tariffs: The Role of Import Prices and Commercial Policies," *American Economic Review*, 4 (1998), 1015-1026.

[30] Jacks, David S., Christopher M. Meissner, and Dennis Novy, “Trade Costs, 1870–2000,” *American Economic Review Papers & Proceedings,* 2 (2008), 529–534.

[31] Johnson, Robert C., “Trade and Prices with Heterogeneous Firms,” mimeo, 2009.

[32] Kehoe, Timothy J., and Kim J. Ruhl, “How Important is the New Goods Margin in International Trade?,” Federal Reserve Bank of Minneapolis Staff Report 324, 2009.

[33] Khandelwal, Amit K., "The Long and Short (of) Quality Ladders", *Review of Economic Studies*, 77(4), 1450-1476, 2010.

[34] Khandelwal, Amit K., Schott, Peter K. and Wei, Shang-Jin, "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters," mimeo, 2011.

[35] Klenow, Pete, and Andrés Rodríguez-Clare, “Quantifying variety gains from trade liberalization,” mimeo, 1997.

[36] Lugovskyy, Volodymyr, and Alexandre Skiba, “Quality Choice: Effects of Trade, Transportation Cost, and Relative Country Size,” mimeo, 2009.

[37] Luttmer, Erzo G. J., “Selection, Growth, and the Size Distribution of Firms,” *The Quarterly Journal of Economics*, 3 (2007), 1103-1144.

[38] Manova, Kalina, and Zhiwei Zhang, “Quality Heterogeneity across Firms and Export Destinations,” NBER Working Paper 15342, 2009.

[39] Martins, Pedro S., and Luca David Opromolla, “Exports, Imports and Wages: Evidence from Matched Firm-Worker-Product Panels,” IZA Discussion Papers 4646, 2009.

[40] Melitz, Mark J., “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica,* 71(2003), 1695-1725.

[41] Melitz, Mark J., and Giammarco Ottaviano, “Market Size, Trade, and Productivity,” *Review of Economic Studies* 1(2008), 295-316.

[42] Obstfeld, Maurice, and Kenneth Rogoff, “The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?,” NBER Macroeconomics Annual, 15 (2000), 339-390.

[43] Ottaviano, Giammarco, Takatoshi Tabuchi and Jacques-François Thisse, “Agglomeration and Trade Revisited,” *International Economic Review*, 43(2002), 409-435.

[44] Rauch, James E., "Networks versus markets in international trade," *Journal of International Economics*, 1(1999), 7-35.

[45] Romer, Paul, "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions," *Journal of Development Economics*, XLIII(1994), 5–38.

[46] Samuelson, Paul A., "The Transfer Problem and the Transport Costs II: Analysis of Effects of Trade Impediments," *The Economic Journal*, June 1954, 64, 264-89; reprinted in Joseph Stiglitz, ed., *Collected Scientific Papers of Paul A. Samuelson*, Vol. 2, Cambridge, Mass., ch. 75.

[47] Suits, D.B. and Musgrave, R.A. (1953), "Ad Valorem and Unit Taxes Compared", *Quarterly Journal of Economics,* Vol. 67, pp 598-604.

[48] Sutton, J. (1991). Price Competition, Advertising, and the Evolution of Concentration. Cambridge: The MIT Press.

[49] WTO "World Tariff Profiles 2006," WTO Secretariat.

A Appendix

A.1 Simulating the Model

A.1.1 Numerical approximation

In this subsection we show how to simulate the model numerically. The numerical approximation of the equilibrium consists of the following steps.

1. Choose a starting value of the the dividend share and the price indices π^0 and P_n^0.[53]

2. For $k = 1$, solve the the system of $N + 1$ equations and $N + 1$ unknowns characterized by

$$
\begin{aligned}
P_n^k &= g\left(P_n^{k-1}, \pi^{k-1}\right) \ \forall \ n \\
\pi^k &= h\left(\pi^{k-1}, P_1^{k-1}, .., P_N^{k-1}\right)
\end{aligned}
$$

[53] Superscripts denote the round of iteration.

as shown in the main text. This involves solving the equilibrium cutoffs (7) and the expression for the price index (8).

3. Iterate over 2. When $\left|P_n^k - P_n^{k-1}\right|$ and $\left|\pi^k - \pi^{k-1}\right|$ are sufficiently small, the equilibrium $\left(\{P_n\}_{n=1}^N, \pi\right)$ is found.

Since the price index does not have a closed-form solution, we approximate it with Monte Carlo methods. Specifically, we take $R = 1e+5$ random draws z^r and η^r from the density $f(z, \eta_n)$. An integral of the form

$$\int\int_{\bar{z}_{in}(\eta_n)}^{\infty} k(z, \eta_n) f(z, \eta_n)\, dz d\eta_n,$$

for an arbitrary function $k()$, is approximated by taking the mean of $k(z, \eta_n)$ over (z^r, η^r) draws that satisfy $z^r > \bar{z}_{in}(\eta_n)$, and adjusting by multiplying with the share of observations that satisfy $z^r > \bar{z}_{in}(\eta_n)$,

$$mean\left[k(z^r, \eta_n^r)\,|z^r > \bar{z}_{in}(\eta_n)\right] \;\times\; \frac{\#obs \text{ where } z^r > \bar{z}_{in}(\eta_n)}{R}.$$

A.1.2 Global Profits

Following Chaney (2008), we assume that each worker owns w_n shares of a global fund. The fund collects global profits Π from all firms and redistributes them in units of the numéraire good to its shareholders. Dividend per share in the economy is defined as $\pi = \Pi / \sum w_i L_i$, and total labor income is $Y_n = w_n L_n (1+\pi)$. Profits for country i firms selling to market n are

$$\pi_{in} = \frac{S_{in}}{\sigma} - n_{in} f_{in},$$

where S_{in} denotes total sales from i to n, n_{in} is the number of entrants, and f_{in} is the entry cost. Global profits are then

$$\begin{aligned} \Pi &= \sum_i \sum_n \left(\frac{S_{in}}{\sigma} - n_{in} f_{in}\right) \\ &= \sum_n \mu Y_n / \sigma - \sum_i \sum_n n_{in} f_{in}. \end{aligned}$$

Note that $\sum_i S_{in}$ is simply $\mu_k Y_n$. Dividend per share is then:

$$\begin{aligned}\pi &= \frac{\Pi}{\sum_i w_i L_i} = \frac{(1/\sigma)\sum_n \mu Y_n - \sum_i \sum_n n_{in} f_{in}}{\sum w_i L_i} \\ &= \frac{(\mu/\sigma)(1+\pi)\sum_n w_n L_n - \sum_i \sum_n n_{in} f_{in}}{\sum w_i L_i}.\end{aligned}$$

Solving for π yields

$$\pi = \frac{\mu/\sigma - \frac{\sum_i \sum_n n_{in} f_{in}}{\sum w_i L_i}}{1-\mu/\sigma}.$$

Note that since $n_{in} = \kappa w_i L_i \int\int_{\bar{z}_{in}(\eta_n)} f(z,\eta_n)\, dz d\eta_n$, π is only a function of the entry hurdle function $\bar{z}_{in}(\eta_n)$. Replacing $\bar{z}_{in}(\eta_n)$ with the entry hurdle expression (7), π becomes a function of itself and the price indices, $\pi = h(\pi, P_1, .., P_N)$ (suppressing all exogenous variables).

A.1.3 Demand shocks and prices

In Section 6, we simulate the model under the restriction that $\eta_{in} = p_{in}^{\psi}$. The price index then becomes

$$P_n^{1-\sigma} = \sum_i \kappa w_i L_i \int_{\bar{z}_{in}}^{\infty} p_{in}(z)^{(1-\psi)(1-\sigma)} f(z)\, dz.$$

where $f(z)$ is the marginal pdf, assumed to be Pareto with shape parameter γ and support $[1;+\infty)$.[54]

Quantity demanded can now be re-expressed as

$$x_{in}(\omega) = \mu p_{in}(\omega)^{\psi(\sigma-1)-\sigma} P_n^{\sigma-1} Y_n$$

The entry hurdle to access market n becomes

$$\bar{z}_{in} = \begin{cases} \tau_{in} w_i \left[\widetilde{\lambda}_1 \left(\frac{f_{in}}{Y_n}\right)^{1/[(\sigma-1)(\psi-1)]} P_n^{1/(1-\psi)} - t_{in}\right]^{-1} & \text{if } t_{in} < \bar{\bar{t}}_{in}, \\ \infty & \text{if } t_{in} \geqslant \bar{\bar{t}}_{in}, \end{cases} \tag{10}$$

where $\bar{\bar{t}}_{in} \equiv \widetilde{\lambda}_1 \left(\frac{f_{in}}{Y_n}\right)^{1/[(\sigma-1)(\psi-1)]} P_n^{1/(1-\psi)}$ and

$$\widetilde{\lambda}_1 = \frac{\sigma-1}{\sigma}\left(\frac{\sigma}{\mu}\right)^{1/[(\sigma-1)(\psi-1)]}.$$

[54] Unlike in earlier models, we do not need to impose the condition $\gamma > \sigma - 1$ for the size distribution of firms to have a finite mean, as long as additive trade costs are positive. The reason is that even the most productive firms have finite revenue.

A.2 Selection bias

Firms are not randomly entering into different product-destinations and this can create a correlation between prices and the error term. In this section, we show that selection may bias the incidental slope coefficients, but not the trade costs coefficients.

According to the model, a firm with a demand shock η_n enters market n if its' productivity is above the theshold $\overline{z}_{kn}(\eta_{knr})$, i.e. $z_{knr} > \overline{z}_{kn}(\eta_{knr})$. Alternatively, we can re-express the entry hurle in terms of the highest price the firm can charge, conditional on a demand shock, $p_{knr} < \overline{p}_{kn}(\eta_{knr})$. Assuming we find a suitable log-linear approximation of the inequality, we write the entry condition as

$$\ln p_{knr} + f(\overline{p}_{kn}) + \ln \eta_{knr} > 0.$$

Export volume is, from equation (3),

$$\ln x_{knr} = A_{kn} + \sigma_{1k} \ln p_{knr} + \sigma_{2k} (\ln p_{knr})^2 + \ln \eta_{knr},$$

Since $\ln \eta_{knr}$ determines both entry and sales, the error term is correlated with the price $\ln p_{knr}$. Using standard methods, and assuming that $\ln \eta_{knr}$ is normal, we find the expectation of the error term in the export volume equation,

$$E[\ln \eta_{knr} | \ln p_{knr} + f(\overline{p}_{kn}) + \ln \eta_{knr} > 0] = \lambda [\ln p_{knr} + f(\overline{p}_{kn})]$$

where λ is the Mills ratio, $\lambda(z) = \phi(z)/\Phi(z)$. Heckman's two step procedure suggest the following regression,

$$\ln x_{knr} = A_{kn} + \sigma_{1k} \ln p_{knr} + \sigma_{2k} (\ln p_{knr})^2 + \lambda [\ln p_{knr} + f(\overline{p}_{kn})] + v_{knr}.$$

Approximating the Mills ratio with the polynomial $\lambda [\ln p_{knr} + f(\overline{p}_{kn})] = c_{kn} + d_k \ln p_{knr} + e_k (\ln p_{knr})^2$, we get

$$\ln x_{knr} = A_{kn} + c_{kn} + (\sigma_{1k} + d_k) \ln p_{knr} + (\sigma_{2k} + e_k)(\ln p_{knr})^2 + v_{knr}.$$

Hence, the incidental slope coefficients may suffer from selection bias, but the the parameter of interest $\widetilde{t}_{kn}$ remains unchanged.

A.3 The prevalence of non-ad valorem duties (NAVs)

A significant share of duties are non-ad valorem (NAVs). According to the WTO World Tariff Profiles (2006), "NAVs are applied by 68 out of the 151 countries shown in this publication including several LDCs..." Table 6 reports, for a set of countries, the share of Harmonized System six-digit subheadings (both for agricultural and non-agricultural products) subject to non-ad valorem duties. The share of products subject to NAVs is usually higher in the case of agricultural products but is also important for non-agricultural products. For example, in the United States, the 3.4% of non-agricultural products that are subject to NAVs account for 18.9% of imports. Still according to the WTO World Tariff Profiles (2006) "One of the peculiarities of NAVs resides in the fact that even if they are applied to a limited number of tariff lines, the products concerned are often classified as sensitive, either because governments collect significant tariff revenues, e.g. cigarettes and alcoholic drinks, or for protecting domestic products against lower priced imports. These highlight the importance of analysing NAVs."

Table 6: Non-ad Valorem Tariffs and Tariff Quotas

		NAV (in %)		Tariff quotas
		MFN Applied	Imports	(in %)
United States	AG	39.9	33.9	9.5
	NAG	3.4	18.9	
European Communities	AG	31.0	24.5	15.1
	NAG	0.6	0.5	
Russian Federation	AG	25.6	58.6	n.a.
	NAG	10.1	6.1	
China	AG	0.3	1.3	5.0
	NAG	0.4	0.1	
Switzerland	AG	73.0	80.3	24.7
	NAG	81.3	62.7	
Japan	AG	13.8	17.0	9.5
	NAG	2.1	2.0	

Note: NAV (in %) corresponds to the share of HS six-digit subheadings subject to non-ad valorem duties under the non-discrimination principle of most-favored nation (MFN). When only part of the HS six-digit subheading is subject to non-ad valorem duties, the percentage share of these tariff lines is used. Tariff quotas (in %) corresponds to the percentage of HS six-digit subheadings in the schedule of agricultural concession covered by tariff quotas. Partial coverage is taken into account on a pro rata basis. Only duties and imports recorded under HS Chapters 01-97 are taken into account. AG stands for "agricultural" while NAG for "non-agricultural" products. Source: WTO World Tariff Profiles 2006.