The Tip of the Iceberg: Modeling Trade Costs and Implications for Intra-industry Reallocation[*]

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

May 2010

Abstract

International economics has overwhelmingly relied on Samuelson's (1954) assumption that trade costs are proportional to value. We build a general equilibrium heterogeneous firms model of trade that allows for both ad valorem and per-unit costs. Using a novel minimum distance estimator we are able to identify per unit trade costs from the distribution of foreign sales across markets. Estimated average per-unit costs are substantial being, on average, between 35 and 45 percent of the average consumer price. This leads us to reject the pure ad valorem cost assumption. An important theoretical finding is that non-ad valorem trade costs create an additional channel of gains from trade through within-industry reallocation. Thus, we show that standard welfare assessments of trade liberalization may be understated.

JEL Classification: F10 *Keywords*: Trade Costs, Heterogeneous Firms, Intra-Industry Reallocation, Exports, Trade Liberalization

[*]Acknowledgements: We would like to thank Costas Arkolakis, Gregory Corcos, Samuel Kortum, Ralph Ossa, Arvid Raknerud, 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, 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.

[†]University of Oslo, Department of Economics, alfonso.irarrazabal@econ.uio.no.

[‡]University of Oslo, Department of Economics, andreamo@econ.uio.no.

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

1 Introduction

The costs of international trade are the costs associated with the exchange of goods and services across borders.[1] 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 variable trade costs as an ad valorem tax equivalent (iceberg costs), implying that pricier goods are also 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 and model trade costs as comprising both an ad valorem part and a per-unit part. Even though more expensive varieties of a given product might be costlier to ship, shipping costs are presumably not proportional to product price. For example, a $200 pair of shoes will typically face much lower ad valorem costs than a $20 pair of shoes.[2] A significant share of tariffs is also per-unit: According to WTO's tariff database, the great majority of member governments (96 out of the 131 included in the database) apply non-ad valorem duties. Among these, Switzerland is the country with the highest percentage of non-ad valorem tariff lines: 83 percent in 2008. The percentage of non-ad valorem active tariff lines in the European Union, the U.S., and

[1] In this paper 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. Among others this includes transportation costs (both freight costs and time costs), policy barriers (tariffs and non-tariff barriers), information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs, and local distribution costs (wholesale and retail) " (Anderson and van Wincoop, 2004).

[2] 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. Given that each pair of shoes weighs 0.2 kg, the ad-valorem shipping costs are in this case 126 and 13.5 percent for the $20 and $200 pair of shoes respectively.

Norway is 10.1, 13.2, and 55, respectively, in 2008.[3] Quotas are also non-ad valorem and can be treated as per-unit.[4] In the European Union, U.S., and Norway 15.1, 9.5, and 39.7 percent of Harmonized System six-digit subheadings in the schedule of agricultural concession is covered by tariff quotas (partial coverage is taken into account on a pro rata basis).

This modeling choice has important consequences when firms are heterogeneous as in Melitz (2003), Chaney (2008), or Eaton et al. (2008). When trade costs are incurred per-unit, trade costs not only alter relative prices across markets but also relative prices within markets.[5] Hence, we identify an additional channel of gains from trade through within-industry intensive margin reallocation. The intuition is that more efficient firms, characterized by lower production costs, will be hit harder by (per-unit) trade costs than less efficient firms, since trade costs will account for a larger share of their final consumer price.[6] As a consequence, per-unit costs tend to wash out the relationship between firm productivity and prices. On the other hand, when trade costs are of the iceberg type exclusively, relative prices within markets are independent of trade costs.

The first contribution of this paper is therefore to present a stylized theory of international trade with heterogeneous firms that encompasses both iceberg costs and per-unit costs. We emphasize the different welfare implications of per-unit versus ad valorem trade frictions. In the special case where iceberg is the only type of trade

[3] Data come from the WTO Integrated Database (IDB) (see http://tariffdata.wto.org). This source reports information, supplied annualy by member governments, on tariffs applied normally under the non-discrimination principle of most-favored nation (MFN). The share of so-called NAVs (non-ad valorem duties) is calculated as the number of NAVs relative to the total number of active tariff lines.

[4] 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.

[5] Relative prices within markets are independent of iceberg costs when per-unit costs are zero and markups do not depend on an interaction between firm characteristics and iceberg costs.

[6] Say that the prices of two varieties are 1 and 10 and that per-unit costs are 2. The relative domestic price is 10, while the relative export price is 4.

cost, our model collapses to the model of Chaney (2008). The second contribution is to structurally fit the model to Norwegian firm-product-destination level export data, using a novel minimum distance estimator. Using the model as our guide, we show that the magnitude of per-unit trade costs is identified by using higher-order moments of the distribution of exports.

Several strong results emerge from the analysis. First of all, per-unit costs are pervasive. The grand mean of per-unit trade costs, expressed relative to the consumer price, is $35-45\%$, depending on the elasticity of substitution. The pure iceberg model is therefore rejected. Second, we show that the costs of per-unit frictions are much higher than the costs of iceberg frictions. Specifically, we check what level of government revenue would be obtained by (a) imposing a per-unit tariff or (b) imposing a welfare-neutral iceberg tariff.[7] Using plausible parameter values, we find that revenue in (b) is much higher than revenue in (a). The flip side is that welfare gains from reducing per-unit frictions are substantially higher than gains from reducing iceberg frictions. Therefore, we conclude that the somewhat technical issue of the functional form of trade costs is quantitatively important for our assessment of the effects of trade distortions. We therefore ask whether the benefit of the iceberg model, in terms of analytical tractability, is worth the costs, in terms of severely biased welfare effects.

More flexible modeling of trade costs is not new in international economics. Alchian and Allen (1964) pointed out that per-unit costs imply that the relative price of two qualities 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

[7] I.e. obtaining the same level of welfare as in case (a). The condition of welfare neutrality makes the two cases comparable.

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 per-unit costs, as we do here. Also, our methodology identifies all kinds of trade costs, whereas their paper is concerned with shipping costs exclusively. Furthermore, Lugovskyy and Skiba (2009) introduce 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. However, the existing literature has not addressed the crucial combination of per-unit costs and heterogeneous firms, which are the two ingredients that drive the results in our model. Also, although we acknowledge that the relationship between trade costs and quality is an important one, in this paper we bypass this question and instead focus on what we think is the core issue: that trade costs alter within-market relative demand.[8] Whether the level of relative demand is due to quality, productivity, or taste differences is of less importance. Bypassing quality is also convenient in estimation, since quality is unobserved in the data.

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.[9] Our approach of using within-market dispersion in exports is conceptually different and provides an alternative approach to inferring trade barriers from data. This

[8] In the Alchian-Allen framework demand for a high-quality relative to low-quality good is increasing in trade costs. In our model demand for a high-price relative to low-price good is increasing in trade costs.

[9] 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.

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 (per-unit) trade costs.

Furthermore, this paper relates to the extensive literature on gains from trade. Most recently, Arkolakis, Costinot, and Rodríguez-Clare (2010) show that gains from trade can be expressed by a simple formula that is valid across a wide range of trade models. Specifically, the total size of the gains from trade is pinned down by the expenditure share on domestic goods and the import elasticity with respect to trade costs. Gains from trade in the presence of per-unit costs are, however, not discussed in their paper. A set of other papers such as Broda and Weinstein (2006), Hummels and Klenow (2005), Kehoe and Ruhl (2009), Klenow and Rodríguez-Clare (1997) and Romer (1994) emphasize welfare gains due to increased imported variety. Although variety gains are present in our model as well, we focus our discussion on the gains from trade due to relative price movements among incumbents.

Finally, our work relates to a recent paper by Berman, Martin, and Mayer (2009). They also introduce a model with heterogeneous firms and per-unit costs, but in their model the per-unit component is interpreted as local distribution costs that are independent of firm productivity. Their research question is very different, however, as their paper analyzes 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.

The rest of the paper is organized as follows. Section 2 presents the model and proposes a simple method for comparing the effect of per-unit costs and iceberg costs on welfare. Section 3 lays out the econometric strategy and presents the baseline estimates as well as a number of robustness checks. Finally, Section 4 concludes.

2 Theory

In this section, we present a stylized theory of international trade that encompasses both iceberg and per-unit 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.[10]

2.1 The Basic Environment

We consider a world economy comprising N asymmetric countries and multiple final goods sectors indexed by $k = 1, ..., K$. Each country n is populated by a measure L_n of workers. Each sector k consists of a continuum of differentiated goods.[11]

Preferences across varieties within a sector k have the standard CES form with an elasticity of substitution $\sigma > 1$.[12] Each variety enters the utility function symmetrically. These preferences generate, in country n, for every variety within a sector k, a quantity demanded function $x_{in}^{k} = \left(p_{in}^{k}\right)^{-\sigma} \left(P_{n}^{k}\right)^{\sigma-1} \mu_k Y_n$, where p_{in}^{k} is the consumer price of a variety produced in country i, P_{n}^{k} is the consumption-based price index in sector k, Y_n is total expenditure, and μ_k is the share of expenditure in sector k. We assume that workers are immobile across countries, but mobile across sectors, firms produce one variety of a particular product, and technology is such that all cost functions are linear in output. Finally, market structure is monopolistic competition.

[10] In the special case where iceberg is the only type of variable trade cost, our model collapses to Chaney (2008).

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

[12] Following Chaney (2008), preferences across sectors are Cobb-Douglas.

2.2 Variable Trade Costs

Unlike much of the earlier trade literature (e.g. Melitz, 2003, Chaney, 2008, Eaton et al., 2008),[13] the economic environment also consists of a transport sector, whose services are used as an intermediate input in final goods production, in order to transfer the goods from a firm's plant to the consumer's hands. Transport services are freely traded and produced under constant returns to scale.

$\phi_m \widetilde{t}_{in}^k$ units of labor are necessary for transferring one unit of a sector-k 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 $w_m \phi_m \widetilde{t}_{in}^k$ for each product and route, where w_m is the wage in country m.[14] 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 $w_i/w_n = \phi_n/\phi_i$. By normalizing the price on a particular shipping route to one, say from i to n for product k, all nominal wages are pinned down.

Firms need labor and transport services in production. Technology is assumed to be Leontief, so demand for the shipping service is proportional to the quantity produced (not proportional to value).

Additionally, the economic environment consists of a standard iceberg cost τ_{in}^k, so that τ_{in}^k units of the final good must be shipped in order for one unit to arrive. The presence of iceberg costs ensures that any correlation between product value and shipping costs is captured by the model.

[13] Hummels and Skiba (2004) and Lugovskyy and Skiba (2009) introduce more general trade costs functions.

[14] Hummels, Lugovskyy, and Skiba (2009) find evidence for market power in international shipping. An extension of our model with increasing returns in shipping would generate lower per-unit trade costs for more efficient firms. In other words, per-unit trade costs would become more like ad valorem costs, since they would be correlated with the price of the good shipped. We focus on perfect competition here in order to isolate the pure per-unit cost case (although the model also allows for pure iceberg costs).

2.3 Prices and Quantities

A firm owns a technology associated with productivity z. A firm in country i, operating in sector k, can access market n only after paying a sector- and destination-specific fixed cost f_{in}^k, in units of the *numéraire*. For notational convenience, let $t_{in}^k \equiv \phi_i \widetilde{t}_{in}^k$, i.e. t_{in}^k is the labor unit requirement of the shipping service if using a domestic shipping company.[15] Profits are then[16]

$$x_{in}^k(z)\left[p_{in}^k(z) - w_i\left(\tau_{in}^k/z + t_{in}^k\right)\right] - f_{in}^k.$$

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,[17]

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

Relative prices within markets are now altered as long as $t_{in}^k > 0$. Specifically, the relative price of two varieties with efficiencies z_1 and z_2 within a sector k is $p_{in}^k(z_1)/p_{in}^k(z_2) = \left(\tau_{in}^k/z_1 + t_{in}^k\right) / \left(\tau_{in}^k/z_2 + t_{in}^k\right)$. In general, both iceberg and per-unit costs will affect within-market relative prices. Relative prices are unaffected by trade frictions only in the special case with $t_{in}^k = 0$.

As in many of the previous trade models, the quantity sold by a firm is linear (in logs) in the price charged to the consumer. Specifically, using (1), the quantity sold by a firm with efficiency z is

$$x_{in}^k(z) = \left(\frac{\sigma}{\sigma-1} w_i\right)^{-\sigma} \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)^{-\sigma} \left(P_n^k\right)^{\sigma-1} \mu_k Y_n.$$

[15] The firm is indifferent between using a domestic or foreign shipping supplier, since the costs are the same, $w_m \phi_m \widetilde{t}_{in}^k = w_i \phi_i \widetilde{t}_{in}^k = w_i t_{in}^k$.

[16] As a convention, we assume that per unit costs are paid on the "melted" output.

[17] The corresponding producer price is $\tilde{p}_{in}^k(z) = \left(p_{in}^k - w_i t_{in}^k\right)/\tau_{in}^k = \sigma/(\sigma-1)\left[1 + z t_{in}^k/\left(\sigma \tau_{in}^k\right)\right] w_i/z$. 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 per-unit trade costs, a larger share of the consumer price does not depend on the producer price. This mechanism is explored theoretically and empirically in Berman et al. (2009).

However, while in previous models the sensitivity of quantity sold (and value of sales) to iceberg trade cost depended *only* on the elasticity of substitution σ, in our model the effect is more complex. The elasticity of the quantity sold to each type of variable trade cost *also* depends on the per-unit trade cost, on the iceberg trade cost, and on the efficiency of the firm itself. The elasticity of the quantity sold by a firm with efficiency z with respect to per-unit and ad valorem trade cost is,[18]

$$\begin{aligned}
\varepsilon_{t_{in}^k} &= -\sigma\left(\frac{\tau_{in}^k}{z t_{in}^k}+1\right)^{-1} \quad \text{and} \\
\varepsilon_{\tau_{in}^k-1} &= -\sigma\left(\frac{t_{in}^k z}{\tau_{in}^k}+1\right)^{-1}\frac{\tau_{in}^k-1}{\tau_{in}^k}.
\end{aligned}$$

The following proposition summarizes a series of important properties of the model.

Proposition 1 *When per-unit trade costs are positive,*

- $|\varepsilon_{t_{in}^k}|$ *is increasing in* z *while* $|\varepsilon_{\tau_{in}^k-1}|$ *is decreasing in* z *and* $|\varepsilon_{t_{in}^k}| > |\varepsilon_{\tau_{in}^k-1}|$ *if* $z > \left(\tau_{in}^k - 1\right)/t_{in}^k$,
- $|\varepsilon_{t_{in}^k}|$ *is increasing in* $t_{in}^k/\left(\tau_{in}^k-1\right)$ *while* $|\varepsilon_{\tau_{in}^k-1}|$ *is decreasing in* $t_{in}^k/\left(\tau_{in}^k-1\right)$,
- *both* $|\varepsilon_{t_{in}^k}|$ *and* $|\varepsilon_{\tau_{in}^k-1}|$ *have an upper bound equal to* σ.

Proof. *See Appendix A.1.* ■

The first statement in Proposition 1 emphasizes an asymmetry that affects most of the results in this paper. The higher the efficiency of a firm, the higher the elasticity of quantities to per-unit trade costs, and the lower the elasticity of quantities to iceberg costs. In other words, a reduction in t_{in}^k will benefit the high efficiency firms disproportionately more than the low productivity firms, in terms of increased quantities sold. As a consequence, factors of production are reallocated from low to high efficiency firms. The intuition is simple once we consider the optimal price (1).

[18] The following elasticities are computed without accounting for changes in the price index.

The share of per-unit trade costs in the consumer price is greater for efficient firms than for less efficient ones.[19] The opposite holds for iceberg costs.

Moreover, when per-unit costs are initially greater than iceberg costs (i.e. when t_{in}^{k} is greater than $\left(\tau_{in}^{k}-1\right)/z$, the iceberg cost converted to labor units for a firm with efficiency z) *a given firm* is more sensitive to changes in per-unit costs than to changes in iceberg costs. This is the case for all the firms lying to the right of the $\left(\tau_{in}^{k}-1\right)/t_{in}^{k}$ thresholds in Figure (1).[20]. The second statement in Proposition 1 points out that the asymmetric effect outlined above becomes stronger when per-unit costs are initially high relative to iceberg costs. Finally, the third statement says that the limit sensitivity of quantity sold to per-unit and ad valorem trade cost is the same and it equals the sensitivity (to ad valorem trade costs) in a model without per-unit trade cost. Figure (1) summarizes the qualitative relationships between the elasticities, firm's efficiency, and the variable trade costs.

2.4 Entry and Cutoffs

We assume that the total mass of potential entrants in country i is proportional to w_iL_i so that larger and wealthier countries have more entrants. 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_iL_i and of the dividends they earn from their portfolio $w_iL_i\pi$, where π is the dividend per share of the global mutual fund.

[19] In the limit, infinitely efficient firms charge a consumer price that is a fixed markup on per unit trade costs only.

[20] In this respect, our model enriches the predictions about sorting of firms that characterize the heterogeneous trade literature. Less efficient firms are more sensitive to ad valorem trade costs while more efficient firms are more sensitive to per-unit costs.

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 $\bar{z}_{in}^k$ from $\pi_{in}^k(\bar{z}_{in}^k) = 0$ as the lowest possible productivity level consistent with non-negative profits in export markets,

$$\bar{z}_{in}^k = \left[\lambda_1^k \left(\frac{f_{in}^k}{Y_n} \right)^{1/(1-\sigma)} \frac{P_n^k}{w_i \tau_{in}^k} - \frac{t_{in}^k}{\tau_{in}^k} \right]^{-1}, \tag{2}$$

with λ_1^k a constant.[21]

2.5 Closing the Model

Following Chaney (2008) and others, we assume that productivity shocks are drawn from a Pareto distribution with density $dF(z)$, shape parameter γ, and support $[1, +\infty)$.[22] The price index in sector k of country n is then

$$\left(P_n^k\right)^{1-\sigma} = \sum_i w_i L_i \int_{\bar{z}_{in}^k}^{\infty} \left(p_{in}^k\right)^{1-\sigma} dF(z)\, dz.$$

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

$$\begin{aligned}
P_n^k &= h\left(\bar{z}_{1n}^k, \bar{z}_{2n}^k, .., \bar{z}_{Nn}^k\right) \ \forall\, n \\
\bar{z}_{in}^k &= f\left(P_n^k, \pi\right) \ \forall\, i\,, n \\
\pi &= g\left(\bar{z}_{11}^k, .., \bar{z}_{1N}^k, \bar{z}_{21}^k, .., \bar{z}_{2N}^k, .., \bar{z}_{NN}^k\right)
\end{aligned}$$

The first equation states that the price index is a function the endogenous entry cutoffs (all exogenous variables are suppressed). The second states that the cutoffs are a function of the price index and the dividend share π (π is part of income Y_n). The third states that the dividend share is a function of all entry cutoffs. We show why this is so in Appendix A.2.1.

[21] Specifically, $\lambda_1^k = (\sigma/\mu_k)^{1/(1-\sigma)} (\sigma - 1)/\sigma$.

[22] Unlike in earlier models (e.g. Chaney, 2008), 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 per-unit trade costs are positive. When $t > 0$ even the most productive firms have finite revenue.

All in all, this constitutes a system of $N(N+1)+1$ equations and $N(N+1)+1$ unknowns. It is not possible to find a closed-form solution for the price index when $t_{in}^k > 0$. In Appendix A.2 we show how to solve the model numerically.

2.6 Welfare and Trade Costs

In this section we show that per-unit frictions lead to higher welfare losses than comparable iceberg frictions. The fundamental problem in comparing welfare effects is that changes in per-unit trade costs are not directly comparable to changes in iceberg costs. E.g. it makes little sense to compare a one percent increase in τ_{in} to a one percent increase in t_{in}. One way to deal with this is the following.[23]

- Start with a frictionless equilibrium and impose either (A) a per-unit import barrier t on imports from m to n or (B) an ad valorem barrier τ on imports from m to n.

- Make (A) and (B) comparable by requiring that welfare in n is identical in the two situations. This amounts to requiring that the price index in country n is identical.[24] From this, we obtain a function $\tilde{\tau}(t)$ that maps per-unit costs to welfare-neutral iceberg costs.

- Now ask how much is collected in import tariffs in (A) and (B). We suspect that revenue is higher in (B), as the price distortions are less severe in this case. If this is the case, the flip side is that welfare gains from reducing per-unit frictions are higher than gains from reducing iceberg frictions.

[23] For simplicity, in this subsection, we consider one sector only and drop the k subscript.

[24] Welfare also depends on income from profits, but as in Chaney (2008), income from profits is constant.

In Appendix A.3 we derive the function $\tilde{\tau}(t)$ and relative tariff revenue G_B/G_A. They are[25]:

$$\tilde{\tau}^{\gamma} = (1+\bar{z}t)^{\gamma-\sigma+1} \frac{\int_1^{\infty} z^{\sigma-1} dF(z)}{\int_1^{\infty} (1/z+\bar{z}t)^{1-\sigma} dF(z)}$$

and

$$\frac{G_B}{G_A} = \frac{\sigma}{\sigma-1} \frac{\int_1^{\infty} (1/z+\bar{z}t)^{1-\sigma} dF(z)}{\int_1^{\infty} (1/z+\bar{z}t)^{-\sigma} dF(z)} \frac{\tilde{\tau}-1}{\bar{z}t} \tag{3}$$

where $\bar{z}$ is the entry hurdle in equilibrium A.[26]

Relative revenue is related to only three variables: $\bar{z}t$, σ, and γ. $\bar{z}t$ is simply per-unit costs relative to production-unit costs of the least efficient exporter, and serves as a convenient measure of the distortion imposed by per-unit costs.[27]

We now show how the ratio relates to different combinations of these variables. In Figure 2 we plot $\bar{z}t$ on the x-axis and G_B/G_A as well as $\tilde{\tau}$ on the y-axis. The four graphs show four different choices of σ and γ, $\sigma = 4, 6, 8$; and $\gamma = \sigma - 1, \sigma + 1$.[28]

In all cases, $G_B/G_A > 1$, implying that iceberg tariffs generate more revenue than per-unit tariffs, holding welfare constant.[29] For example, if per-unit costs are 40 percent of production costs for the least efficient exporter, then the alternative strategy of imposing a welfare-neutral iceberg cost will increase tariff revenue by roughly 200 percent, on average (taking the mean over all σ and γ values).[30] The

[25] Compared to the expressions in Appendix A.2, we simplify notation by calling the variable of integration z, instead of $\tilde{z}_A$ and $\tilde{z}_B$. Also, we define $\bar{z} = \bar{z}_{Akn}$.

[26] Here we assume that $\gamma > \sigma - 1$, since equilibrium (B) is undefined otherwise.

[27] The fact that we can express the G ratio as a function of $t\bar{z}_{Akn}$ instead of t and $\bar{z}_{Akn}$ separately makes the calculations much simpler, since we do not have to evaluate the effect of t on $\bar{z}_{Akn}$ (which involves calculating the price index).

[28] The values for γ chosen here are higher than the ones estimated in the next section (see Table 1). The reason is that the G ratio is defined only when $\gamma > \sigma - 1$. Figure 2 shows that the G ratio increases for lower γ, so using the estimated value for γ instead would presumably increase the G ratio even more.

[29] We have tried a range of combinations of parameter values, and we always find that $G_B/G_A > 1$.

[30] In this case tariff revenue (in equilibrium B) is slightly less than the value of imports.

intuition is that per-unit costs will hit the more productive firms especially hard (their prices will increase by more than less efficient firms, in percent), which will translate into lower welfare compared to iceberg costs. In order to stay on the same welfare level, iceberg costs must increase, raising the G_B/G_A ratio.

One concern is that quality heterogeneity might weaken these theoretical findings. For example, if efficient firms are producing high quality goods at high prices, then per-unit costs will hit low quality goods harder than high quality ones.[31] The welfare effect of per-unit costs in such a model is not clear cut. We explore this possibility in Appendix A.4. We find that the G_B/G_A ratio can indeed become less than 1, but that for the large majority of parameter values $G_B/G_A > 1$. Our simulations suggest that the additional distortion generated by per-unit frictions are quantitatively important, even in the presence of quality heterogeneity, and that reducing trade frictions may give higher welfare gains than in standard models.

2.7 The Export Volume Distribution

In this section we examine some properties of the distribution of exports, across firms for a given product-destination pair. We will make extensive use of these properties further below when we estimate the model. We first derive the theoretical exports volume distribution for every destination n and product k. Source country subscripts are dropped because Norway is always the source in the data. Given that productivity among potential entrants is distributed Pareto, the productivity distribution among exporters of product k to destination n is also Pareto with cumulative distribution function $F\left(z|\bar{z}_n^k\right) = 1 - \left(z/\bar{z}_n^k\right)^{-\gamma}$. The Pareto shape coefficient γ is assumed to be equal across products and destinations. Then the exports volume cumulative

[31] Johnson (2009) proposes a model where firms are heterogeneous both in terms of unit costs and quality.

distribution function (CDF), conditional on $z > \bar{z}_n^k$, is[32]

$$\begin{aligned} Q\left(x|\bar{z}_n^k\right) &= \Pr\left(X < x|Z > \bar{z}_n^k\right) \\ &= 1 - \left(A_n^k x^{-1/\sigma} - B_n^k\right)^{\gamma}, \end{aligned} \tag{4}$$

where A_n^k and B_n^k are two clusters of parameters,

$$\begin{aligned} A_n^k &= \frac{\sigma-1}{\sigma}\bar{z}_n^k\left(P_n^k\right)^{(\sigma-1)/\sigma}\mu_k^{1/\sigma}\frac{Y_n^{1/\sigma}}{\tau_n^k w}, \\ B_n^k &= \frac{t_n^k}{\tau_n^k/\bar{z}_n^k}. \end{aligned}$$

2.7.1 Properties of the Distribution

As with the scale parameter for the Pareto distribution, A_n^k will affect the location of the distribution. For example, an increase in market size Y_n will shift the probability density function to the right, so that it becomes more likely to sell greater quantities.

Since $B_n^k = t_n^k/\left(\tau_n^k/\bar{z}_n^k\right)$, B_n^k simply measures per-unit trade costs (t_n^k) relative to the unit costs of the least efficient firm, inclusive ad valorem costs ($\tau_n^k/\bar{z}_n^k$). When $t_n^k = 0 \implies B_n^k = 0$, the distribution is identical to Pareto with shape parameter γ/σ. This is similar to Chaney (2008), where the sales distribution preserves the shape of the underlying efficiency distribution and the sales distribution is identical across markets. When $t_n^k > 0$, B_n^k will affect the dispersion of quantity sold. This can be seen by finding the inverse CDF:

$$x_n^k(\vartheta) = Q^{-1}(\vartheta) = \left[\frac{(1-\vartheta)^{1/\gamma} + B_n^k}{A_n^k}\right]^{-\sigma}.$$

Dispersion, as measured by the ratio between the ϑ_2^{th} and ϑ_1^{th} percentiles ($0 < \vartheta_1 < \vartheta_2 < 1$) is then

$$D\left(\vartheta_2, \vartheta_1; B_n^k, \gamma, \sigma\right) \equiv \frac{x_n^k(\vartheta_2)}{x_n^k(\vartheta_1)} = \left[\frac{(1-\vartheta_1)^{1/\gamma} + B_n^k}{(1-\vartheta_2)^{1/\gamma} + B_n^k}\right]^{\sigma}. \tag{5}$$

[32] The cdf is well-behaved when $\left(\frac{1+B_n^k}{A_n^k}\right)^{-\sigma} - x_{\min n}^k < 0$ and $x_{\max n}^k - \left(\frac{B_n^k}{A_n^k}\right)^{-\sigma} < 0$ where $x_{\min n}^k$ is the minimum export volume and $x_{\max n}^k$ is maximum export volume.

When $t_n^k = 0$, this ratio is constant across destinations. When $t_n^k > 0$, the ratio declines as B_n^k goes up. That is, exports volume becomes less dispersed with higher per-unit costs. The intuition is that higher per-unit costs will hit the high productivity/low cost firms harder than firms with low productivity/high cost, since more trade costs will force the high productivity firms to increase their price by more than the low productivity firms, in percentage terms. This will translate into a larger reduction in quantity sold for the high productivity firms relative to the low productivity firms, so that dispersion will decrease. The following proposition summarizes our findings:

Proposition 2 *When per-unit costs are positive ($t_{in}^k > 0$), dispersion, as measured by the ratio between the ϑ_2^{th} and ϑ_1^{th} percentiles, is decreasing in t_{in}^k and increasing in τ_{in}^k. Moreover, when per-unit costs are zero ($t_{in}^k = 0$), then dispersion is invariant to a change in variable trade costs τ_{in}^k.*

Proof. See Appendix A.5. ■

In Appendix A.5 we prove this proposition allowing for trade costs to alter the entry cutoffs and the price index.

In Figure 3 we plot the theoretical exports volume complementary cumulative distribution function, on log scales. The value of the complementary CDF is on the horizontal axis, while quantity exported is on the vertical axis. The solid line represents the case when $B_n^k = t_n^k / \left(\tau_n^k / \bar{z}_n^k\right) = 0$. The gradient is then equal to $-\sigma/\gamma$. The dotted line represents the case when per-unit costs are positive. As B_n^k increases, the complementary CDF becomes more and more concave. The concavity reflects the fact that large firms (low cost firms) are hit harder by per-unit costs than small firms (high cost firms).

The properties of the exports volume distribution also survive, under some assumptions, in a framework where firms are heterogeneous both in terms of unit costs and quality.[33] In Appendix A.6 we also investigate whether departures from the CES

[33] More specifically, the result that dispersion decreases with per-unit trade costs carries through as long as the relationship between production unit costs and quality is not too convex (as we show in

framework in models with ad valorem costs can generate predictions similar to those of our model. We show that for a popular class of linear demand systems (and with zero per-unit costs), dispersion in exports will increase in ad valorem costs - the opposite of the case with per-unit costs.[34] We provide additional information about the properties of the distribution in Section 3.4 (identification).

3 Estimating the Model

In this section we structurally estimate the magnitude of per-unit trade costs. We showed in the theory section that per-unit costs introduce curvature in the export volume CDF, leading to less dispersion in exports volume as per-unit trade costs increase. When per-unit trade costs are zero, dispersion in exports volume is unaffected by (ad valorem) trade costs. This is the identifying assumption that allows us to recover estimates of trade costs consistent with our model.[35] The econometric strategy consists of using a minimum distance estimator that matches the empirical distribution of exports volume (per product-destination) to the theoretical distribution.[36]

Our approach of estimating trade costs from an economic model is very different from the earlier literature.[37] First, most studies model trade costs as ad valorem exclusively, omitting the presence of per-unit costs. A notable exception is Hummels

Appendix A.5.1). In our particular dataset, as we will show below, this is the case for an overwhelming majority of product-destination pairs.

[34] In Appendix A.6.1, we also consider an extension of our model where firms have to sustain marketing costs in order to promote their products and reach consumers, following Arkolakis (2008). It turns out that, in the extended model, as long as the market penetration effect is not too strong compared to the per-unit trade cost effect, we can interpret our results as a lower bound on the true magnitude of the ad valorem equivalent of per-unit trade costs.

[35] In Section 3.2 below, we provide evidence that is consistent with the identifying assumption.

[36] We choose to use data for export volume (quantities) instead of export sales for the following reasons. First, a closed-form solution for the sales distribution does not exist. Second, using quantities instead of sales avoids measurement error due to imperfect imputation of transport/insurance costs. Third, we avoid transfer pricing issues when trade is intra-firm (Bernard, Jensen and Schott 2006).

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

and Skiba (2004), who distinguish between them and find evidence for the presence of per-unit shipping costs.[38] Compared to our work, they study freight costs exclusively, whereas we consider all types of international trade costs. Second, our methodology utilizes within-country dispersion in exports volume to achieve identification of trade costs, whereas earlier studies utilize cross-country variation in 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 (although conditional on a value of the elasticity of substitution). Fourth, we do not impose trade cost symmetry (t_{in}^k and τ_{in}^k can differ from, respectively, t_{ni}^k and τ_{ni}^k).

3.1 Data

The data consist of an exhaustive panel of Norwegian non-oil exporters in the 1996-2004 period. Data come from customs declarations. Every export observation is associated with a firm, a destination and a product *id* and for every export observation we observe the quantity transacted and the total value.[39] Since identification in the empirical model is based solely on cross-sectional variation, we chose to work, in our baseline specification, with the 2004 cross-section, the most recent available to us. The product *id* 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 dataset.

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

[38] 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.

[39] 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 for all the products. Moreover, 27.5% of the products are additionally measured in quantities, while 4.7% are additionally measured in other units (m^3, carat, etc.). In the baseline estimation we use kilos unless a different unit is available.

(standard deviation 7.8). As we will see, we utilize the distribution of export quantity across firms within a product-destination in the econometric model. We therefore choose to restrict the sample to product-destinations where more than 40 firms are present.[40] In the robustness section, we evaluate the effect of this restriction by estimating the model on an expanded set of destination-product pairs. Below, extreme values of quantity sold, defined as values below the 1^{st} percentile or above the 99^{th} percentile for every product-destination, are eliminated from the dataset. All in all, this brings down the total number of products to 121 and the number of destinations to 21.[41]

Before presenting the formal econometric model, we show some descriptive statistics that suggest how dispersion is related to trade costs. In Figure 4, we first calculated the ratio between the 90^{th} and the 10^{th} percentile of exports quantity for each product-destination. Second, we averaged the ratios across products for every destination, using exports value for each product as weights.[42] Third, we plotted the mean ratio against distance (a proxy for trade costs), in logs. The relationship is clearly negative, indicating that trade costs tend to narrow the dispersion in exports quantity. Regressions that include the usual gravity-type right hand side variables and product fixed effects will give the same result. The relationship is also robust to other measures of dispersion, such as the Theil index or the coefficient of variation.[43]

We also investigate why the percentile ratios are falling with trade costs (as proxied

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

[41] Exports to all possible combinations of these products and destinations amount to 26.2% of total export value.

[42] In order to show the pattern for as many destinations as possible, we have based these calculations on the unrestricted sample, i.e. using all product-destinations with more than one firm present.

[43] Specifically, a regression that includes distance, population, real GDP per capita (all in logs) and an indicator function for contiguity, together with product fixed effects yields an estimate of the distance coefficient of $-.667$ (t-stat $= -27.10$), with standard errors clustered at the product level. Alternative regressions where the dependent variable is the coefficient of variation or the Theil index yield estimates of $-.204$ (t-stat $= -23.99$) and $-.406$ (t-stat $= -23.34$), respectively.

by distance). Our theory predicts that firms in the top of the distribution are hit harder than firms in the bottom of the distribution as per-unit trade costs increase. An alternative way of checking the validity of this prediction, is to investigate how export percentiles (not ratios) are correlated with trade costs (as proxied by distance). We construct export percentiles (from P03 to P99 with intervals of 3) for every product-destination pair and regress each of them on a product fixed effect, distance, GDP, and GDP per capita (all in logs).[44] The product fixed effect ensures that we are only using variation within a product category, across markets, to identify the distance effect. Figure 5 shows the results. The solid line shows the estimated distance coefficient (indicated on the y-axis) when the dependent variable is the n^{th} export percentile (indicated on the x-axis). The dotted lines show the 95 percent confidence interval. All the export percentiles are negatively related to distance, but as we move from the lower to the higher percentiles, the distance estimate is decreasing. Hence, there is a tendency of the upper percentiles to be hit harder by trade costs, just as predicted by the model.

3.2 Prices and Quantities Within a Market

The theoretical prediction of a negative correlation between per-unit trade costs and export dispersion relies on the assertion that firms in the top of the export quantity distribution charge lower prices than firms in the bottom of the distribution.[45] We show in Appendix A.5.1 that this pattern will emerge even in models of quality heterogeneity, under certain parameter restrictions. The identifying assumption is therefore consistent with recent findings, e.g. Baldwin and Harrigan (2007) and Johnson (2009), that more able firms also charge higher prices. From an empirical

[44] Standard errors are clustered by HS8 product. Product-destinations are included if the number of firms per product-destination > 40, as in the baseline model.

[45] To fix ideas, this would be equivalent to IKEA selling a higher number of beds than Crate and Barrel. We emphasize that we are referring to a negative correlation between the price charged and the *quantity/volume* of items sold as opposed to the value of sales. The former is much more empirically likely than the latter.

point of view, the correlation between price and quantity sold is something we can easily check in the data, as prices can be approximated by unit values. In the data, we find that the average correlation between unit value and (quantity) market share is $-.59$ (the unweighted average over product-destination pairs, using only the pairs that we estimate on) and that 96.7 percent of the correlations are negative.[46] The histogram of the correlation coefficients is shown in Figure 6. If the negative relationship between quantity and price is weaker than our model suggests, possibly due to quality heterogeneity within an HS8 category, then the link between per-unit trade costs and dispersion will also be weaker. In that case, our estimate of per-unit trade costs will be biased downward, since our model will interpret high dispersion as low trade costs. However, the strong negative correlations for most of the products in our sample indicate that the bias is relatively small.

3.3 Estimation

We use a minimum distance estimator that matches empirical dispersion in exports volume (per product-destination) to simulated dispersion in exports volume. Specifically, denote the empirical ratio between the ϑ_2^{th} and ϑ_1^{th} percentiles for product k in destination n as $\widetilde{D}_n^k(\vartheta_2, \vartheta_1)$ and stack a set of $(\vartheta_2, \vartheta_1)$ ratios in the $M \times 1$ column vector $\widetilde{\mathbf{D}}_n^k$. Denote its simulated counterpart $D\left(\vartheta_2, \vartheta_1; B_n^k, \gamma, \sigma\right)$, as defined in equation (5), and stack a set of $(\vartheta_2, \vartheta_1)$ ratios in the $M \times 1$ column vector $\mathbf{D}\left(B_n^k, \gamma, \sigma\right)$. Define the criterion function as the squared difference between $\ln \mathbf{D}\left(B_n^k, \gamma, \sigma\right)$ and $\ln \widetilde{\mathbf{D}}_n^k$:

$$d(\mathbf{\Psi}) = \sum_{n}^{N} \sum_{k \in \Omega_n} \left[\ln \mathbf{D}\left(B_n^k, \gamma, \sigma\right) - \ln \widetilde{\mathbf{D}}_n^k\right]' \left[\ln \mathbf{D}\left(B_n^k, \gamma, \sigma\right) - \ln \widetilde{\mathbf{D}}_n^k\right],$$

where $\mathbf{\Psi}$ is the vector of coefficients to be estimated, N is the total number of destinations and Ω_n is the set of products sold in market n. We minimize $d(\mathbf{\Psi})$ with

[46] Using an alternative dataset (see Section 3.6) of Portugal-based exporters, we find that the average correlation between unit value and export quantity is $-.36$ and 97.5 percent of the correlations are negative. Manova and Zhang (2009), using data on Chinese trading firms, also find a negative correlation between firm f.o.b. export prices and quantity sold (see Table 4, column 2).

respect to $\mathbf{\Psi}$ and denote $\widehat{\mathbf{\Psi}}$ the equally weighted minimum distance estimator.[47]

We model B_n^k as the product of sector and destination fixed effects,

$$B_n^k = \beta_k b_n,$$

and normalize $\beta_1 = 1$.[48] This decomposition enables us to identify the share of trade costs that is due to product characteristics and the share that is due to market characteristics. Also, note that even though β_k is estimated relative to some normalization, the estimates of the B values are invariant to the choice of normalization. Finally, we condition the criterion function on a guess of σ (see next section). The coefficient vector then consists of $\mathbf{\Psi} = (\beta_k, b_n, \gamma)$, in total $K + N$ parameters.

We choose the following percentile ratio moments: $(.95, .05)$, $(.90, .10)$, $(.75, .25)$, $(.60, .40)$, $(.20, .10)$, $(.30, .20)$, $(.40, .30)$, $(.50, .40)$, $(.60, .50)$, $(.70, .60)$, $(.80, .70)$, $(.90, .80)$; in total $M = 12$ moments per product-destination.[49]

As the covariance matrix of the vector of empirical percentile ratios ($\widetilde{\mathbf{D}}_n^k$) is unknown, the standard error of the estimator is not available using standard formulas. Instead, we employ a nonparametric bootstrap (empirical distribution function bootstrap). Specifically, we sample with replacement within each product-destination pair, obtaining the same number of observations as in the original sample. After performing 500 bootstrap replications, we form the standard errors by calculating the standard deviation for each coefficient in $\mathbf{\Psi}$.

[47] Theory suggests that for overidentified models it is best to use optimal GMM. In implementation, however, the optimal GMM estimator may suffer from finite-sample bias (Altonji and Segal 1996). Furthermore, it is difficult to calculate the optimal weighting matrix in our context, as it would necessitate evaluating the variance of the percentile ratios for every product-destination (see e.g. Cameron and Trivedi, 2005, section 6.7).

[48] The normalization is similar to the one adopted in the estimation of two-way fixed effects in the employer-employee literature (see Abowd, Creecy, and Kramarz 2002). 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.

[49] We experimented with other combinations of moments as well and the results remained largely unchanged.

3.4 Identification

Recall that Figure 3 showed that per-unit costs (or B_n^k) introduced curvature in the complementary CDF. The set of percentile ratio moments enables us to trace out this curvature, which will pin down B_n^k. The Pareto shape parameter γ is identified by the gradient of the CDF that is common across markets for a given product. We emphasize that with only one moment, B_n^k and γ are not separately identified, as one percentile ratio will give information only about the slope of the CDF. As a consequence, variability in dispersion alone (e.g. variability in the P90/P10 ratio) will not generate a positive estimate of per-unit costs. Instead, per-unit costs are identified by the non-linearity in the complementary CDF - that large (low price) firms are hit harder by per-unit costs (embedded in B_n^k) than small (high price) firms.[50]

As is usual in trade models, the elasticity of substitution σ is not identified. The criterion function $d(\mathbf{\Psi})$ is therefore conditional on a guess of σ. In the results section we report estimates based on different values of σ. In the baseline specification, we assume a common γ across all products, but a model with product-specific γ values is also identified. We estimate a model with heterogeneity in γ and σ in the robustness section.

As noted above, $B_n^k = t_n^k / \left(\tau_n^k / \bar{z}_n^k\right)$ simply measures per-unit trade costs (t_n^k) relative to the unit costs of the least efficient exporter, inclusive of ad valorem costs ($\tau_n^k / \bar{z}_n^k$). A more common measure of trade costs is trade costs relative to price. First, using the first-order condition from the firm's maximization problem, we can re-express firm-level consumer prices as

$$p_n^k(\tilde{z}) = \frac{\sigma w t_n^k}{\sigma - 1} \left(\frac{1}{\tilde{z} B_n^k} + 1 \right), \tag{6}$$

where $\tilde{z}$ is productivity measured relative to the cutoff ($z = \tilde{z}\bar{z}_n^k$).[51] Second, consider

[50] An implication is that a linear empirical CDF (in logs) will result in an estimate of zero per-unit trade costs.

[51] Note that $\tilde{z}$ is distributed like a Pareto with scale parameter 1 and shape parameter γ.

the average price (charged by Norwegian exporters) of product k in destination n:

$$\begin{aligned}\bar{p}_n^k &= \frac{1}{1-F\left(\bar{z}_n^k\right)}\int_{\bar{z}_n^k} p_n^k\left(z\right)dF\left(z\right)\\ &= \left(\bar{z}_n^k\right)^{\gamma}\int_{\bar{z}_n^k} p_n^k\left(z\right)\gamma z^{-\gamma-1}dz\\ &= \int_1 p_n^k\left(\tilde{z}\right)dF\left(\tilde{z}\right),\end{aligned}$$

where we used the substitution $z=\tilde{z}\bar{z}_n^k$ in the last equality. Third, inserting equation (6) and solving for $wt_n^k/\bar{p}_n^k$ yields:

$$\frac{wt_n^k}{\bar{p}_n^k}=\frac{\sigma-1}{\sigma}\frac{B_n^k\left(\gamma+1\right)}{\gamma+B_n^k\left(\gamma+1\right)}.$$

The ratio $wt_n^k/\bar{p}_n^k$ measures (per-unit) trade costs relative to the average consumer price.[52] Given our estimate of B_n^k and γ, the expression on the right hand side can be easily computed. Note that integrating over productivities allows us to express trade costs only as a function of B_n^k, γ and σ. This is due to the fact that a Pareto density is parameterized only by the cutoff ($\bar{z}_n^k$) and the shape parameter (γ). Our estimates of B_n^k and γ are therefore sufficient to obtain a meaningful measure of per-unit trade costs.

3.5 Results

Table 1 summarizes the results.[53] We apply the methodology described in the previous section in order to back out a simple measure of per-unit costs from the model. Estimated per-unit trade costs $wt_n^k/\bar{p}_n^k$, measured relative to the consumer price, averaged over products and destinations, are 0.36 (s.e. 0.01), conditional on $\sigma=6$, which we use as our baseline case.[54] Estimated trade costs drop to 0.35 for $\sigma=4$

[52] Note that this measure of the relevance of per-unit trade costs does not require us to disentagle t_n^k from τ_n^k.

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

[54] σ is estimated to 3.79 in Bernard, Eaton, Jensen, and Kortum (2004). In summarizing the literature, Anderson and van Wincoop (2004) conclude that σ is likely to be in the range of five to ten.

and rise to 0.45 for $\sigma = 8$. These estimates are similar to those found in the existing literature, where international trade barriers are typically estimated in the range of $40 - 80$ percent for a $5 - 10$ elasticity estimate (Anderson and van Wincoop 2004).[55]

Furthermore, 99 and 95 percent of the β_k and b_n coefficients respectively (the product and destination fixed effects embedded in B_n^k) are significantly different from zero at the 0.05 level. Since B_n^k is greater than zero only when per-unit costs $t_n^k > 0$, our findings suggest that the standard model with only iceberg costs is rejected.[56] The estimate of γ, the Pareto coefficient, is 1.31 (s.e. 0.03) in the baseline case.

Figure 7 shows $wt_n^k/\bar{p}_n^k$ for every destination, averaged over products, on the vertical axis and distance (in logs) on the horizontal axis. Estimated trade costs are clearly increasing in distance. Note that our two-way fixed effects approach enables us to construct $wt_n^k/\bar{p}_n^k$ even for product-destination pairs that are not present in the data. This implies that there is no selection bias in Figure 7, since all products are included in every destination. The robust relationship between distance and trade costs also emerges when regressing trade costs on a product fixed effect and a set of gravity variables (distance, contiguity, GDP, and GDP per capita, all in logs).[57] The distance coefficient is then 0.07 (s.e. 0.001), meaning that doubling distance yields a 7% increase in trade costs.

Figure 8 shows the distribution of estimated trade costs from Norway to the U.S. across products. This figure exploits the variability retrieved from the β_k variables. As expected, per-unit trade costs are heterogeneous, with values ranging from roughly 10 to 70 percent of the product value.[58] Figure 9 shows the relationship between es-

[55] Earlier estimates of international trade barriers are not directly comparable to our estimate of $wt_n^k/\bar{p}_n^k$, however, as earlier studies define trade barriers as the ratio of total (ad valorem) trade barriers relative to domestic trade barriers τ_{in}/τ_{ii}.

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

[57] The full set of results is available upon request.

[58] Densities for $\ln B_k^n$ for other markets are simply shifted left or right compared to the density for

timated trade costs and actual average weight/unit and weight/value in logs.[59] Since weight/unit and weight/value should be positively correlated with actual trade costs, we expect to see a positive relationship between these measures and estimated trade costs. Indeed, the figures indicate an upward sloping relationship. The correlation between weight/unit (weight/value) and trade costs $wt^k/\bar{p}^k$ (averaged over destinations) is 0.59 (0.22).

Most of the estimates in the product dimension also make intuitive sense. For example, tractors (HS 87019009) and self-propelled front-end shovel loaders (HS 84295102) are among the products with estimated trade costs above the 95^{th} percentile (taking the average over all destinations). Toys (HS 95039000) and CDs (HS 85243901) are among the products with estimated trade costs below the 5^{th} percentile (taking the average over all destinations).

It is also of interest to study the importance of product and destination characteristics on trade costs. Since the expression for $wt_n^k/\bar{p}_n^k$ is a monotonically increasing function of B_n^k, a straightforward indicator of the importance of product and destination characteristics is the dispersion in β_k and b_n respectively. In the baseline case, the 90-10 percentile ratio of β_k and b_n is 5.40 and 1.63, respectively, suggesting that product characteristics are $3-4$ times as important for trade costs compared to destination characteristics.

Furthermore, the decomposition of product and destination effects allows us to study whether costly destinations are associated with products with lower transport 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,

the U.S. This is by construction, since it is only the destination fixed effect b_n that is different in the construction of the density for alternative markets.

[59]Since only a subset of products has quantities measured in units, the number of products in the graph is lower than what is used in the estimation. Average weight/unit and weight/value are obtained by dividing total weight (summed over firms and destinations) over total units, or value (summed over firms and destinations).

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 slightly positive but not significantly different from zero.

We also investigate whether the unweighted average of trade costs is different from the weighted average.[60] When using export values per product-destination as weights, the weighted average of trade costs is 0.27. This suggests that product-destinations associated with high costs have below average exports.

Finally, Figure 10 shows actual and simulated percentile ratios (95/05, 90/10, 75/25, and 60/40), for all product-destination pairs. Most observations lie close to the 45 degree line, although the fit of the model is declining closer to the median. Overall, this leads us to conclude that the model is able to fit the data quite well.

3.6 Robustness

3.6.1 Variations of the Baseline Model

Next we present some re-estimations of the model that address several issues. The results are summarized in Table 2. One concern is our reliance on the Pareto distribution. Even though the Pareto is known to approximate the US firm size distribution quite well (e.g. Luttmer 2007), one could argue that dispersion is decreasing with trade costs due to extensive margin effects. As is well known, the fractal nature of the Pareto distribution implies that a percentile ratio is independent of truncation. As a consequence, movements in the entry cutoff do not affect the percentile ratio (when $t = 0$). However, under other distributions this may no longer be the case. For example, with the lognormal distribution and $t = 0$, dispersion will decrease with higher entry hurdles simply because the density is truncated from below, not due to intensive margin reallocation. One way of controlling for this, is to estimate the

[60] We focus mainly on the unweighted average because otherwise we would have a selection problem when comparing trade costs across destinations.

model on a subsample where the set of exporters is identical in every destination (for each product). Specifically, we take the three most popular destinations, Sweden, Denmark, and Germany, and extract the firms that are present in all three markets, for each product. Since this procedure reduces the sample size dramatically, we are forced to lower the threshold of firms present in a product-destination pair, from 40 to 10, resulting in three destinations and 11 products. With only 10 or more firms per product-destination, we must also reduce the number of estimating moments per product-destination, so we choose four moments: The $(95, 05)$, $(90, 10)$, $(75, 25)$ and $(60, 40)$ percentile ratios. The results are shown in column (R1) in Table 2. Although the sample size is much smaller, the trade cost estimates are nearly unchanged and estimated with high precision. Therefore, we conclude that truncation of the export distribution is not driving our results, nor are other types of selection effects.

Second, a concern is that quality heterogeneity within an HS8 product group might bias our results, something we briefly discussed in Section 3.2. One way to check whether this is an issue, is to re-estimate the model on homogeneous goods exclusively. If the results become very different, there is reason for concern. We use the Rauch (1999) classification of differentiated and homogeneous goods and choose products that are classified as non-differentiated.[61] The number of products declines quite substantially in this case, so as in (R1), we set the firm threshold to more than 10 firms per product-destination and restrict the set of moments to four. Estimates are reported in column (R2), and the results are reassuring. The grand mean of trade costs (conditional on $\sigma = 6$) is identical to the baseline case.

Third, we investigate whether the choice of truncating the dataset to only product-destinations with more than 40 firms affects the results. We choose product-destinations with more than 10 firms present, resulting in 42 destinations and 826 products. Again, the lower threshold forces us to reduce the number of moments used in estimation,

[61] Goods classified as "goods traded on an organized exchange" or "reference priced" in the Rauch (1999) database. We use Jon Haveman's concordances at http://www.macalester.edu/research/economics/page/haveman/Trade.Resources/TradeData.html to convert SITC to HS codes.

so as in (R1), we choose four moments: The $(95, 05)$, $(90, 10)$, $(75, 25)$, and $(60, 40)$ percentile ratios. The estimate of trade costs increases slightly, to 0.43, as shown in column (R3).[62] This suggests that although the product-destination pairs we estimate on are not a random sample, the selection bias resulting from truncation is not very large.

Fourth, we investigate whether the choice of units affects the results. The high share of products that are measured in kilos might bias the results if weight per-unit is varying across both destinations and firms. For example, if high productivity firms are able to reduce unit weight in remote markets, while low productivity firms are not, then dispersion will decrease. We address this issue by selecting the subsample of products that are measured in units, not kilos. This truncates the dataset to 40 products and six markets. Again, the results do not change much, as shown in column (R4) in the table.

Fifth, we re-estimate the model on the 2003 cross-section instead of the 2004 cross-section. The results in column (R5) show that the grand mean of trade costs is identical to the baseline result.

Sixth, we estimate the model on a dataset of Portuguese exporters. The data have the same structure as the Norwegian one.[63] The results in column (R6) show that the mean of per-unit trade costs for Portugal is 0.34, very close to the Norwegian estimates (for $\sigma = 6$).

Finally, we also check the sensitivity of the results to heterogeneity in the elasticity of substitution σ and the Pareto coefficient γ. First, we take estimates of the σ from Broda and Weinstein (2006), and take the unweighted average of their HS 10-digit estimates for every 4-digit product.[64] Second, we allow for product-specific γ

[62] Our estimation algorithm spent 15 hours minimizing the objective function in this case. Therefore, we do not report standard errors in column (R2), as bootstrapping becomes prohibitively costly.

[63] A detailed description of the dataset of Portuguese exporters can be found in Martins and Opromolla (2009). The average correlation between unit value and export quantity is $-.36$ (the unweighted average over product-destination pairs, using only the pairs that we estimate on) and 97.5 percent of the correlations are negative.

[64] We average up to the 4-digit level because i) only the first six digits are internationally comparable

values, so that the theoretical percentile ratios become $D\left(\vartheta_2, \vartheta_1; B_n^k, \gamma_k, \sigma_k\right)$ and the coefficient vector to be estimated becomes $\mathbf{\Psi} = (\beta_k, b_n, \gamma_k)$, in total $2K + N - 1$ coefficients. The results are reported in column (R7).[65] Again, per-unit costs are large and significant, although the point estimate falls somewhat compared to the baseline case.

3.6.2 Other Robustness Checks

We have shown that the concavity of the export distribution is systematically related to market characteristics and that per-unit trade costs are a reasonable explanation for this fact. However, we would like the model to fit other aspects of the data as well, in dimensions that are not easily explained in competing models. One such feature is f.o.b. export prices. Our model predicts that f.o.b. prices are increasing in per-unit trade costs (see theory section).[66] The reason is that the perceived elasticity of demand is lower when per-unit costs are high, leading firms to charge a higher markup.[67] In Table 3 we provide evidence that this is indeed the case in the data. We regress firm-product-destination level unit values on a firm-product fixed effect, distance, GDP, and GDP per capita. Standard errors are clustered on HS8 products. The distance coefficient is identified purely from the variation within a firm-product pair across destinations. The coefficient is positive and significant, using both Norwegian and Portuguese firm-level data, suggesting that a given exporter charges a higher price for the same HS8 good in more remote markets. This result is difficult to reconcile with existing heterogeneous firms trade models, as noted by Manova and Zhang

and (ii) not all products are jointly present in the Norwegian and U.S. data.

[65] We do not report standard errors in this case, because bootstrapping is prohibitively time-consuming.

[66] A model with endogenous quality choice and per-unit costs will also generate a positive correlation between f.o.b. prices and per-unit trade costs.

[67] Note that, as in Berman et al. (2009), our model is consistent with incomplete exchange rate pass-through. The elasticity of the f.o.b. price to a real depreciation is less than one as long as per-unit trade costs are finite. Our model predicts that this elasticity (and therefore the degree of exchange rate pass-through) is increasing in t/τ.

(2009). In standard models with CES demand, e.g. Melitz (2003), firms charge the same f.o.b. price in all locations. In standard models with linear demand, e.g. Melitz and Ottaviano (2008), firms charge lower markups in remote markets. In standard quality-sorting models, e.g. Baldwin and Harrigan (2007), firms are assumed to sell the same quality worldwide.

Eaton, Kortum, and Kramarz (2008) argue that sales and entry shocks are needed in order to explain the entry and sales patterns of French exporters. Our model, on the other hand, has only variability along the productivity dimension. Although additional error components would certainly increase the fit of the model, we decided to choose a somewhat simpler setup in this paper.[68] First, the defining feature of the data we have attempted to explain is the varying dispersion in exports across destinations. A model with entry and sales shocks but without per-unit costs cannot explain this,[69] unless one assumes that the variances and/or covariances of the shocks are correlated with distance. Second, our econometric model is expressed in closed-form, even though analytical expressions for many key relationships do not exist. This helps to keep the run-time of the estimation algorithm down to an acceptable level.

4 Conclusions

In this paper we have first explored theoretically the implications of introducing more flexible trade costs in an otherwise standard Melitz (2003) heterogeneous firm model of international trade. An important finding is that we identify an additional channel of gains from trade through intensive margin reallocation compared to the standard model. The mechanism behind the result is that the more productive firms are hit harder by trade costs compared to the less productive firms when trade costs are independent of efficiency (and price). It is thus the marriage of per-unit costs and heterogeneity in efficiency that drives the theoretical results in this paper.

[68] In an earlier paper (Irarrazabal, Moxnes, and Opromolla 2009) we estimated demand and fixed cost shocks in a model with heterogeneous firms, exports, and multinational production.

[69] Abstracting from possible effects of truncation of the distribution on dispersion.

We tie the stylized model to a rich firm-level dataset of exports, by product and destination. By using the identifying assumption from theory that higher-order moments from the export quantity distribution (for a given product-destination pair) are systematically related to (per-unit) trade costs, we are able to back out a structural estimate of trade costs. Our empirical results indicate that per-unit costs are not just a theoretical possibility: they are pervasive in the data, and the grand mean of trade costs, expressed relative to the consumer price, is between 35% and 45%. We therefore conclude that pure iceberg is 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: per-unit costs.

A broader implication of our work is related to the skill premium. To the extent that more productive firms demand more high-skill labor (e.g. as in Verhoogen 2008), lowering trade barriers will increase aggregate demand for high skill labor through the intensive margin reallocation channel emphasized in this paper. As a consequence, our model makes clear an additional link between trade (the decline in international transportation costs) and the skill premium.

Finally, we explore the welfare implications of per-unit versus iceberg frictions. We introduce either a per-unit friction or an iceberg friction and impose that they are welfare neutral. We then check what the government would collect in tariff revenue in both cases. Using plausible parameter values, we find that iceberg frictions generate much higher revenue than per-unit frictions. The flip side is that welfare gains from reducing per-unit frictions are substantially higher than gains from reducing iceberg frictions. This suggests that existing estimates of the potential for gains from trade may be too low. A fairly robust policy implication of our work is that, if governments are determined to raise revenue through import duties, they should impose ad valorem duties rather than per-unit duties, due to the additional distortions associated with per-unit duties.

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

[9] Berman, Nicolas, Philippe Martin, and Thierry Mayer, '´How Do Different Exporters React to Exchange Rate Changes? Theory, Empirics and Aggregate Implications," mimeo, 2009.

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

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

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

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

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

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

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

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

[18] 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.

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

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

[21] 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.

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

[23] 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.

[24] 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.

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

[26] Johnson, Robert C., "Trade and Prices with Heterogeneous Firms," mimeo, 2009.

[27] 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.

[28] Klenow, Pete, and Andrés Rodríguez-Clare, "Quantifying variety gains from trade liberalization," mimeo, 1997.

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

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

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

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

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

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

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

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

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

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

[39] 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.

[40] Verhoogen, Eric, "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 2(2008), 489-530.

A Appendix. Not for publication.

A.1 Elasticity of Quantity Sold to Trade Costs (Proposition 1)

In this subsection we summarize (and extend) the conditions outlined in Proposition 1 and Figure 1.

Derivatives with respect to z. The relevant derivatives are

$$\frac{\partial|\varepsilon_{t_{in}^k}|}{\partial z} = \frac{\tau_{in}^k}{\sigma t_{in}^k z^2}\varepsilon_{t_{in}^k}^2 > 0 \ ; \quad \frac{\partial^2|\varepsilon_{t_{in}^k}|}{\partial z^2} = -\frac{2\tau_{in}^k}{\sigma t_{in}^k z^3}\left(\frac{\tau_{in}^k}{\sigma t_{in}^k z}\varepsilon_{t_{in}^k}+1\right)\varepsilon_{t_{in}^k}^2 < 0$$
$$\frac{\partial|\varepsilon_{\tau_{in}^k-1}|}{\partial z} = -\frac{t_{in}^k}{\sigma\tau_{in}^k-\sigma}\varepsilon_{\tau_{in}^k-1}^2 < 0; \quad \frac{\partial^2|\varepsilon_{\tau_{in}^k-1}|}{\partial z^2} = -2\left(\frac{t_{in}^k}{\sigma\tau_{in}^k-\sigma}\right)^2\varepsilon_{\tau_{in}^k-1}^3 > 0,$$

showing that $|\varepsilon_{t_{in}^k}|$ is increasing and concave in z while $|\varepsilon_{\tau_{in}^k-1}|$ is decreasing and convex in z. The signs of the derivatives rely upon $\sigma > 1$, $z \geq 1$, $\tau_{in}^k \geq 1$ and $t_{in}^k \geq 0$.

Derivatives with respect to $t_{in}^k/\left(\tau_{in}^k-1\right)$. The relevant derivatives are

$$\frac{\partial|\varepsilon_{t_{in}^k}|}{\partial t_{in}^k/(\tau_{in}^k-1)} = \frac{\left(\tau_{in}^k\right)^2}{\sigma z\left(t_{in}^k\right)^2}\varepsilon_{t_{in}^k}^2 > 0 \text{ and}$$
$$\frac{\partial|\varepsilon_{\tau_{in}^k-1}|}{\partial t_{in}^k/(\tau_{in}^k-1)} = -\frac{\left(1+zt_{in}^k\right)}{\sigma t_{in}^k}\varepsilon_{\tau_{in}^k-1}^2 < 0,$$

showing that $|\varepsilon_{t_{in}^k}|$ is increasing in $t_{in}^k/\left(\tau_{in}^k-1\right)$ while $|\varepsilon_{\tau_{in}^k-1}|$ is decreasing in $t_{in}^k/\left(\tau_{in}^k-1\right)$.The signs of the derivatives rely upon $\sigma > 1$, $z \geq 1$, $t_{in}^k \geq 0$.

The statement that $z > \left(\tau_{in}^k-1\right)/t_{in}^k \implies |\varepsilon_{t_{in}^k}| > |\varepsilon_{\tau_{in}^k-1}|$ and the upper bounds for $|\varepsilon_{t_{in}^k}|$ and $|\varepsilon_{\tau_{in}^k-1}|$ can be easily derived from the expressions for $|\varepsilon_{t_{in}^k}|$ and $|\varepsilon_{\tau_{in}^k-1}|$.

A.2 Simulating the Model

In this subsection we show how to simulate the model numerically. For simplicity we restrict the number of products to one and suppress the product index. The numerical approximation of the equilibrium consists of the following steps.

1. Choose a starting value of the price index P_n^0.[70]
2. Solve the equilibrium cutoffs and global profits simultaneously, conditional on P_n^0. The cutoffs and global profits are

$$\bar{z}_{in} = f\left(P_n,\pi\right)\ \forall i,j$$
$$\pi = g\left(\bar{z}_{11},..,\bar{z}_{1N},\bar{z}_{21},..,\bar{z}_{NN}\right),$$

   where only the endogenous arguments in functions f and g are explicitly shown. The expression for π is shown further below. The system consists of N^2+1 equations and N^2+1 unknowns and can be solved by choosing a candidate π, solving $\bar{z}_{in}$ using f, inserting the solution back into g, etc., until the system converges.

[70] Superscripts denote the round of iteration.

3. Given the solutions $\bar{z}_{in}^{0}$, a new candidate price index $P_n^1 = h(\bar{z}_{1n}, \bar{z}_{2n}, .., \bar{z}_{Nn})$ is calculated.

4. Iterate over 2 and 3. When $\left|P_n^r - P_n^{r-1}\right|$ is sufficiently small, the equilibrium $\{P_n, \bar{z}_{in}, \pi\}$ is found.

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

$$P = \int_Z^{\infty} p(z)^{1-\sigma} g(z)\, dz$$

is then approximated by taking the mean of $p(z)^{1-\sigma}$ conditional on $z^r > Z$, and (iii) adjusting by multiplying with the share of observations above Z,

$$P \approx mean\left(p(z)^{1-\sigma} | z^r > Z\right) \times \frac{\#obs > Z}{1e+5}.$$

A.2.1 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

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

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

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

Solving for π yields

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

Note that since $n_{in} = w_i L_i \int_{\bar{z}_{in}} dF_1(z) = w_i L_i \bar{z}_{in}^{-\gamma}$, π is only a function of the endogenous variables $\bar{z}_{in}$. That is why we expressed $\pi = g(\bar{z}_{11}, .., \bar{z}_{1N}, \bar{z}_{21}, .., \bar{z}_{NN})$ in the section above.

A.3 Welfare Effects of Changes in τ and t

Here we derive the expressions for G_B/G_A and $\tau(t)$ stated in the main text. Define A as the equilibrium with a t barrier on imports from m to n and B as the equilibrium with a welfare-neutral τ barrier on imports from m to n. All other trade flows are assumed to be frictionless. The price indices in (A) and (B) are:

$$P_{An}^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left[\sum_{i\neq m} w_i L_i \int_{\bar{z}_{Ain}}^{\infty} (w_i/z)^{1-\sigma}\, dF(z) + w_m L_m \int_{\bar{z}_{Amn}}^{\infty} \left[w_m(1/z+t)\right]^{1-\sigma} dF(z)\right],$$

and

$$P_{Bn}^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\left[\sum_{i\neq m} w_i L_i \int_{\bar{z}_{Bin}}^{\infty} (w_i/z)^{1-\sigma}\, dF(z) + w_m L_m \tau^{1-\sigma} \int_{\bar{z}_{Bmn}}^{\infty} (w_m/z)^{1-\sigma} dF(z)\right],$$

where $\bar{z}_{Ain}$ and $\bar{z}_{Bin}$ are the entry hurdles in the two cases. Let $P_{An} = P_{Bn}$. Then

$$\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\int_{\bar{z}_{Amn}}^{\infty} \left[w_m(1/z+t)\right]^{1-\sigma} dF(z) = \left(\frac{\tau\sigma}{\sigma-1}\right)^{1-\sigma}\int_{\bar{z}_{Bmn}}^{\infty} (w_m/z)^{1-\sigma} dF(z).$$

Here we used the fact that $\bar{z}_{Ain} = \bar{z}_{Bin}$, $\forall i \neq m$, since these cutoffs are determined by the same price level. Rearranging, the welfare-neutral iceberg barrier $\tilde{\tau}$ is

$$\tilde{\tau}^{1-\sigma} = \frac{\int_{\bar{z}_{Amn}}^{\infty} (1/z+t)^{1-\sigma}\, dF(z)}{\int_{\bar{z}_{Bmn}}^{\infty} z^{\sigma-1} dF(z)}. \tag{7}$$

Let us calculate government revenue in the two cases,

$$\begin{aligned} G_A &= w_m L_m \int_{\bar{z}_{Amn}}^{\infty} w_m x_A(z) t\, dF(z) \\ &= \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} w_m^{2-\sigma} L_m t\, (P_{An})^{\sigma-1} \mu Y_n \int_{\bar{z}_{Amn}}^{\infty} (1/z+t)^{-\sigma} dF(z) \end{aligned}$$

and

$$\begin{aligned} G_B &= w_m L_m \int_{\bar{z}_{Bmn}}^{\infty} x_B(z)(\tilde{\tau}-1)\, p_B(z)\, dF(z) \\ &= \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} w_m^{2-\sigma} L_m (\tilde{\tau}-1)\, \tilde{\tau}^{1-\sigma} (P_{An})^{\sigma-1} \mu Y_n \int_{\bar{z}_{Bmn}}^{\infty} z^{\sigma-1} dF(z), \end{aligned}$$

using $x_s = p_s^{-\sigma}(P)^{\sigma-1}\mu Y$, $s = A, B$. Note that the demand shifter is the same in both equilibria because the price index is identical. Relative tariff revenue is then

$$\begin{aligned} \frac{G_B}{G_A} &= \frac{\tilde{\tau}^{1-\sigma}\int_{\bar{z}_{Bmn}}^{\infty} z^{\sigma-1} dF(z)}{\int_{\bar{z}_{Amn}}^{\infty} (1/z+t)^{-\sigma} dF(z)} \frac{\sigma\tilde{\tau}-\sigma}{\sigma t - t} \\ &= \frac{\int_{\bar{z}_{Amn}}^{\infty} (1/z+t)^{1-\sigma} dF(z)}{\int_{\bar{z}_{Amn}}^{\infty} (1/z+t)^{-\sigma} dF(z)} \frac{\sigma\tilde{\tau}-\sigma}{\sigma t - t}, \end{aligned} \tag{8}$$

where we used equation (7) in the last equality.

We can proceed further by using the expression for productivity cutoffs and the Pareto assumption. The relationship between the cutoffs is

$$\bar{z}_{Bmn} = \frac{\tilde{\tau}}{\bar{z}_{Amn}^{-1} + t}. \tag{9}$$

Since the support of the Pareto is $[1, +\infty]$, the cutoff is well defined when $\tilde{\tau}/\left(\bar{z}_{Amn}^{-1} + t\right) > 1$. Inserting $dF(z) = \gamma z^{-\gamma-1}dz$ and using integration by substitution, $\tilde{z}_A = z/\bar{z}_{Amn}$ and $\tilde{z}_B = z/\bar{z}_{Bmn}$, we can re-express equation (7) as

$$\tilde{\tau}^{1-\sigma} = \left(\frac{\bar{z}_{Amn}}{\bar{z}_{Bmn}}\right)^{\sigma-\gamma-1} \frac{\int_1^\infty \left(1/\tilde{z}_A + \bar{z}_{Amn}t\right)^{1-\sigma} dF\left(\tilde{z}_A\right)}{\int_1^\infty \tilde{z}_B^{\sigma-1} dF\left(\tilde{z}_B\right)}.$$

Eliminating $\bar{z}_{Bmn}$ using (9) then yields

$$\tilde{\tau}^{\gamma} = \left(1 + \bar{z}_{Amn}t\right)^{\gamma-\sigma+1} \frac{\int_1^\infty \tilde{z}_B^{\sigma-1} dF\left(\tilde{z}_B\right)}{\int_1^\infty \left(1/\tilde{z}_A + \bar{z}_{Amn}t\right)^{1-\sigma} dF\left(\tilde{z}_A\right)}.$$

This is the expression for welfare-neutral iceberg costs used in the main text.

Finally, we can re-express the revenue ratio in the same fashion,

$$\frac{G_B}{G_A} = \frac{\int_1^\infty \left(1/\tilde{z}_A + \bar{z}_{Amn}t\right)^{1-\sigma} dF\left(\tilde{z}_A\right)}{\int_1^\infty \left(1/\tilde{z}_A + \bar{z}_{Amn}t\right)^{-\sigma} dF\left(\tilde{z}_A\right)} \frac{\sigma\tilde{\tau} - \sigma}{\sigma\bar{z}_{Amn}t - \bar{z}_{Amn}t}.$$

This is the expression for relative revenue used in the main text.

A.4 Extension: Quality heterogeneity

In this section, we extend our model to account for quality heterogeneity across firms. We show that under very plausible conditions the welfare result from Subsection 2.6 still applies, and we discuss the conditions under which the exports volume distribution retains the properties derived in Subsection 2.7.1. In order to deal with quality heterogeneity, we consider an extended CES utility function of the following type,

$$U = \left(\int_{\omega\in\Omega_n} \left[x_{in}(\omega)\, q_{in}(\omega)\right]^{(\sigma-1)/\sigma} d\omega\right)^{\sigma/(\sigma-1)},$$

where $q_{in}(\omega)$ is the quality of variety ω measured in *utils* per physical unit and Ω_n is the set of available varieties. Quantity demanded is given by

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

and the CES price index is

$$P_n^{1-\sigma} = \sum_i \int_{\omega\in\Omega_{in}} \left[p_{in}(\omega)/q_{in}(\omega)\right]^{1-\sigma} d\omega.$$

Assume the following relationship between quality and cost: $q(z) = (1/z)^{\eta}$.[71] If $\eta = 0$, there is no quality heterogeneity. When $0 < \eta < 1$, higher unit costs are associated with higher quality, but with decreasing returns, so that doubling unit costs implies less than twice the quality. When $\eta > 1$, higher unit costs are associated with higher quality under increasing returns.[72] The relationship between unit costs and quality-adjusted prices depends on the value of η. For example, when $\eta > 1$, higher unit costs are associated with higher quality $q(z)$ and lower quality-adjusted prices,

$$p_{in}(z)/q(z) = \frac{\sigma}{\sigma - 1} w_i \left[\tau_{in} \left(\frac{1}{z} \right)^{1-\eta} + t_{in} \left(\frac{1}{z} \right)^{-\eta} \right].$$

A.4.1 Quality and Welfare

With this framework, it is relatively straightforward to replicate the analysis shown in Subsection A.2. This yields the welfare neutral $\tilde{\tau}$,

$$\tilde{\tau}^{\gamma/(1-\eta)} = (1 + t\bar{z})^{\gamma/(1-\eta)-\sigma+1} \frac{\displaystyle\int_1^{\infty} z^{(1-\eta)(\sigma-1)} dF(z)}{\displaystyle\int_1^{\infty} z^{-\eta(\sigma-1)} \left(1/z + \bar{z}t\right)^{1-\sigma} dF(z)}$$

and the government revenue ratio

$$\frac{G_B}{G_A} = \frac{\displaystyle\int_1^{\infty} z^{-\eta(\sigma-1)} \left(1/z + \bar{z}t\right)^{1-\sigma} dF(z)}{\displaystyle\int_1^{\infty} z^{-\eta(\sigma-1)} \left(1/z + \bar{z}t\right)^{-\sigma} dF(z)} \frac{\sigma\tilde{\tau} - \sigma}{\sigma\bar{z}t - \bar{z}t}$$

where G_A and G_B are defined as in the main text.

We can now evaluate the robustness of our welfare result in the presence of quality heterogeneity. Specifically, we would like to check the conditions under which $G_B/G_A < 1$. This can occur if η is sufficiently large: A high η means that firms charging a high nominal price have a low quality-adjusted price (because of superior quality). Consequently, introducing per-unit costs does not greatly affect these firms. Presumably, then, per-unit costs will have a small effect on welfare, since these firms have a large market share. In order to understand how likely this case is, we replicate the simulation in the main text and calculate the G_B/G_A ratio when $\eta \in [0, 10]$. We find that, for the full range of η, $G_B/G_A > 1$ when $\bar{z}t > 0.1$. This means that, as long as per-unit costs are 10 percent or more of production costs for the least efficient exporter, the alternative strategy of imposing a welfare-neutral iceberg cost will increase tariff revenues. All in all, we conclude that although quality heterogeneity makes the theoretical mechanism less clear cut, simulations show that this is not quantitatively important. For a majority of parameter values, per-unit costs still reduce welfare more than comparable iceberg costs.

[71] This is similar to the functional form used in e.g. Baldwin and Harrigan (2007) and Johnson (2009).

[72] An additional, theoretically possible (but empirically unlikely), case is $\phi < 0$. In this case, higher unit costs are associated to lower quality.

A.5 Dispersion and Trade Costs (Proposition 2)

In this subsection we prove Proposition 2. The first part of the subsection is about the relationship between dispersion in the exports volume distribution and per-unit costs. The second part is about the relationship between dispersion in the exports volume distribution and iceberg costs.

Proof that $D_{\vartheta_2,\vartheta_1}$ is decreasing in t_n^k, when $t_n^k > 0$. Recall from (5) that the percentile ratio is

$$D_{\vartheta_2,\vartheta_1} = \left[\frac{(1-\vartheta_1)^{1/\gamma} + B_n^k}{(1-\vartheta_2)^{1/\gamma} + B_n^k}\right]^{\sigma},$$

where $B_n^k = \bar{z}_n^k t_n^k / \tau_n^k$. Consider the impact on $D_{\vartheta_2,\vartheta_1}$ of a small change in t_n^k,[73]

$$\hat{D}_{\vartheta_2,\vartheta_1} = \sigma D_{\vartheta_2,\vartheta_1}^{-1/\sigma} \frac{(1-\vartheta_2)^{1/\gamma} - (1-\vartheta_1)^{1/\gamma}}{\left[(1-\vartheta_2)^{1/\gamma} + B_n^k\right]^2} \hat{B}_n^k B_n^k.$$

Note that the fraction is negative since $0 < \vartheta_1 < \vartheta_2 < 1$. It remains to evaluate the sign of $\hat{B}_n^k$,

$$\hat{B}_n^k = \hat{\bar{z}}_n^k + \hat{t}_n^k.$$

The percentage change in the cutoff is

$$\hat{\bar{z}}_n^k = \bar{z}_n^k \frac{t_{in}^k}{\tau_{in}^k} \hat{t}_{in}^k - \hat{P}_n^k \left(1 + \bar{z}_n^k \frac{t_{in}^k}{\tau_{in}^k}\right). \tag{10}$$

Inserting the previous expression back into $\hat{B}_n^k$ yields

$$\hat{B}_n^k = \left(1 + \bar{z}_n^k \frac{t_{in}^k}{\tau_{in}^k}\right)\left(\hat{t}_n^k - \hat{P}_n^k\right).$$

Since $\hat{t}_n^k > \hat{P}_n^k$ (see proof below), $\hat{B}_n^k > 0$ and $\hat{D}_{\vartheta_2,\vartheta_1} < 0$. Therefore, dispersion, measured by the $D_{\vartheta_2,\vartheta_1}$ percentile ratio, is declining in per-unit trade costs. ■

Proof that $\hat{t}_n^k > \hat{P}_n^k$. The price index in country n is defined as

$$P_n^k = \frac{\sigma}{\sigma - 1}\left[\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}^k} \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)^{1-\sigma} dF_1(z)\right]^{1/(1-\sigma)},$$

where $F_1(z)$ is the Pareto CDF with support $z \in [1, +\infty)$ and $f_1(z)$ is the corresponding PDF. Consider a percentage change in the price index due to a marginal change in t_{in}^k,[74] $\forall i$,

$$\begin{aligned}
\hat{P}_n^k &= \frac{1}{1-\sigma} \frac{\sum_i w_i L_i w_i^{1-\sigma} dI_{in}^k}{\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}^k} \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)^{1-\sigma} dF_1(z)} \\
&= \frac{\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}^k} \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)^{1-\sigma} dF_1(z) \left[\hat{I}_{in}^k / (1-\sigma)\right]}{\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}^k} \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)^{1-\sigma} dF_1(z)} \\
&= \sum_i \lambda_{in}^k \frac{\hat{I}_{in}^k}{1-\sigma},
\end{aligned} \tag{11}$$

[73] Let $\hat{X}$ be the percentage change in variable X due to a small change in t_n^k.

[74] Again, let $\hat{X}$ be the percentage change in variable X due to a small change in t_n^k.

where

$$I_{in}^{k} \equiv \int_{\bar{z}_{in}^{k}} \left(\frac{\tau_{in}^{k}}{z} + t_{in}^{k} \right)^{1-\sigma} dF_1(z),$$

and λ_{in}^{k} is the share of country n's total expenditure that is devoted to goods from country i (the import share),

$$\lambda_{in}^{k} = \frac{w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}^{k}} \left(\tau_{in}^{k}/z + t_{in}^{k} \right)^{1-\sigma} dF_1(z)}{\sum_j w_j L_j w_j^{1-\sigma} \int_{\bar{z}_{jn}^{k}} \left(\tau_{jn}^{k}/z + t_{jn}^{k} \right)^{1-\sigma} dF_1(z)}.$$

Applying Leibnitz's Rule, the percentage change in I_{in}^{k} is

$$\hat{I}_{in}^{k} = (1-\sigma)\, \hat{t}_{in}^{k} \chi_{in}^{k} - \bar{\chi}_{in} \bar{z}_{in}^{k} \hat{\bar{z}}_{in}^{k},$$

where

$$\chi_{in}^{k} = \frac{\int_{\bar{z}_{in}^{k}} t_{in}^{k} \left(\frac{\tau_{in}^{k}}{z} + t_{in}^{k} \right)^{-\sigma} dF_1(z)}{\int_{\bar{z}_{in}^{k}} \left(\frac{\tau_{in}^{k}}{z} + t_{in}^{k} \right)^{1-\sigma} dF_1(z)} \text{ and } \bar{\chi}_{in}^{k} = \frac{\left(\frac{\tau_{in}^{k}}{\bar{z}_{in}^{k}} + t_{in}^{k} \right)^{1-\sigma} f_1\left(\bar{z}_{in}^{k} \right)}{\int_{\bar{z}_{in}^{k}} \left(\frac{\tau_{in}^{k}}{z} + t_{in}^{k} \right)^{1-\sigma} dF_1(z)}.$$

Note that χ_{in}^{k} is always less than one.

Consider a change in per-unit costs (dt_n^k) from a particular country (ex. Norway) to country n. Denote with λ_n^k, I_n^k, χ_n^k, $\bar{\chi}_n^k$, $\bar{z}_n^k$ the values of the corresponding variables when associated with the source country under consideration. Disregard any possible second-order effects so that $d\bar{z}_{in}^{k} = 0$ whenever the source country is not Norway. Then

$$\hat{P}_n^k = \lambda_n^k \frac{\hat{I}_n^k}{1-\sigma} = \lambda_n^k \left(\chi_n^k \hat{t}_n^k + \frac{\bar{\chi}_n^k}{\sigma - 1} \bar{z}_n^k \hat{\bar{z}}_n^k \right).$$

The first term in the expression above captures the intensive margin effect on the price index, while the second term captures the extensive margin effect. Using (10) and solving for $\widehat{P}_n^k$ yields

$$\hat{P}_n^k = \frac{\lambda_n^k \chi_n^k + \lambda_n^k \frac{\bar{\chi}_n^k}{\sigma-1} \bar{z}_n^k \left(\bar{z}_n^k \frac{t_n^k}{\tau_n^k} \right)}{1 + \lambda_n^k \frac{\bar{\chi}_n^k}{\sigma-1} \bar{z}_n^k \left(1 + \bar{z}_n^k \frac{t_n^k}{\tau_n^k} \right)} \hat{t}_n^k$$

Since $\lambda_n^k \chi_n^k < 1$, the fraction is less than one and therefore $\hat{P}_n^k < \hat{t}_n^k$. ■

Proof that $D_{\vartheta_2,\vartheta_1}$ is increasing in τ_n^k, when $t_n^k > 0$. Consider the impact on $D_{\vartheta_2,\vartheta_1}$ of a small change in τ_n^k. The expression for $\hat{D}_{\vartheta_2,\vartheta_1}$ remains the same as above, but $\hat{B}_n^k$ now becomes

$$\hat{B}_n^k = \hat{\bar{z}}_n^k - \frac{\tau_n^k - 1}{\tau_n^k} \hat{\tau}_n^k,$$

and the percentage change in cutoff becomes

$$\hat{\bar{z}}_n^k = \frac{\tau_n^k - 1}{\tau_n^k} \hat{\tau}_{in}^k - \widehat{P}_n^k \left(1 + \bar{z}_n^k \frac{t_{in}^k}{\tau_{in}^k} \right).$$

Inserting this back into $\hat{B}_n^k$ yields

$$\hat{B}_n^k = -\hat{P}_n^k \left(1 + \bar{z}_n^k \frac{t_{in}^k}{\tau_{in}^k} \right).$$

Note that $\hat{B}_n^k < 0$ since $\hat{P}_n^k > 0$. Therefore $\hat{D}_{\vartheta_2,\vartheta_1} > 0$ and dispersion rises with iceberg costs. ■

Proof that D_{ϕ_2,ϕ_1} is independent of τ_n^k, when $t_n^k = 0$. Consider the impact on $D_{\vartheta_2,\vartheta_1}$ of a small change in τ_n^k when $t_{in}^k = 0$. The percentile ratio then collapses to

$$D_{\vartheta_2,\vartheta_1} = \left[\frac{(1-\vartheta_1)^{1/\gamma}}{(1-\vartheta_2)^{1/\gamma}}\right]^{\sigma},$$

clearly showing that dispersion is independent of variable trade costs in the Chaney (2008) model. ■

A.5.1 Quality and the Export Distribution

In this subsection, we explore how the exports volume distribution responds to an increase in per-unit trade costs in the presence of quality heterogeneity. In particular, we analyze the conditions under which the distribution becomes less dispersed as per-unit trade costs increase. This is equivalent to determining the circumstances under which firms charging low prices also sell higher quantities. The optimal quantity sold is $x_{in} = P_n^{\sigma-1}\mu Y_n q^{\sigma-1} p_{in}^{-\sigma}$. Using the expression for quality, we get $x_{in} = P_n^{\sigma-1}\mu Y_n z^{-\eta(\sigma-1)} p_{in}^{-\sigma}$, or equivalently,

$$x_{in} = P_n^{\sigma-1}\mu Y_n \left(\frac{\sigma}{\sigma-1} w_i \tau_{in}\right)^{-\eta(\sigma-1)} p_{in}^{-\sigma+\eta(\sigma-1)},$$

using the fact that $p_{in}(z) = \frac{\sigma}{\sigma-1} w_i \frac{\tau_{in}}{z} \iff z = \frac{\sigma}{\sigma-1} w_i \tau_{in} p_{in}(z)^{-1}$ under $t_{in} = 0$. As a consequence, low price firms sell more when $\eta < \sigma/(\sigma-1)$.[75] Higher per-unit costs will now reduce dispersion since the low price firms (in the top of the distribution) will be hit harder by per-unit costs than the high price firms (in the bottom of the distribution). The fundamental identifying assumption in the econometric model is therefore consistent with quality heterogeneity, as long as the (positive) relationship between unit costs and quality is not too convex.[76] In our particular dataset, we know that for an overwhelming majority of product-destination pairs, the correlation between quantity and price is negative. This suggests that η is less than $\sigma/(\sigma-1)$ (but not necessarily zero). As a comparison, Johnson (2009) finds that $\eta > 1$ in most of the SITC 4-Digit sectors in his data. Since $\sigma/(\sigma-1) > 1$, our results are fully compatible with his.

A.6 Extension: Departures from CES

In this subsection we consider two departures from standard CES preferences: marketing costs *à la* Arkolakis (2008) and linear demand with horizontal product differentiation *à la* Ottaviano, Tabuchi, and Thisse (2002).

[75] This condition is necessary and sufficient when $t_{in} = 0$ but only necessary when $t_{in} > 0$.

[76] Returning to a previous example, this would be the case if, for example, IKEA were selling *fewer* beds than Crate and Barrel.

A.6.1 Marketing Costs

We consider an extension of our model that includes marketing costs *à la* Arkolakis (2008). The problem of the firm is now the following:

$$\max_{h_{in}(z),p_{in}(z)} x_{in}(z)\left[p_{in}(z)-w_i\left(\frac{\tau_{in}}{z}+t_{in}\right)\right]-w_n^{\theta}w_i^{1-\theta}\frac{L_n^{\alpha}}{\psi}\frac{1-[1-h_{in}(z)]^{1-\beta}}{1-\beta},$$
$$\text{s.t.}\quad h_{in}(z)\in[0,1].$$

where demand is

$$x_{in}(z)=\frac{[p_{in}(z)]^{-\sigma}}{(P_n)^{1-\sigma}}y_n h_{in}(z)L_n,$$

y_n is per-capita spending in country n, L_n is population of country n, and $h_{in}(z)$ is the fraction of country n consumers reached by the firm. The remaining parameters and the functional form adopted to describe marketing costs are discussed extensively in Arkolakis (2008). The optimal price charged by an exporter to country n is the same as in our framework and equal to (1). The elasticity of the total volume of goods exported to country n to trade costs ($\varepsilon_{t_{in}}$ and $\varepsilon_{\tau_{in}-1}$) is equal to the sum of (i) the elasticity of the *per-consumer* volume to trade costs and (ii) the elasticity of the fraction of consumers reached to trade costs ($\varepsilon_{h_{in}(z),t_{in}}$ and $\varepsilon_{h_{in}(z),\tau_{in}-1}$). The elasticity of the per-consumer volume to trade costs in this modified model is equal to the elasticity of total volume to trade costs in our baseline model. Overall,

$$\begin{aligned}\varepsilon_{t_{in}} &= -\sigma\left(\frac{\tau_{in}}{zt_{in}}+1\right)^{-1}+\varepsilon_{h_{in}(z),t_{in}},\\ \varepsilon_{\tau_{in}-1} &= -\sigma\left(\frac{t_{in}z}{\tau_{in}}+1\right)^{-1}\frac{\tau_{in}-1}{\tau_{in}}+\varepsilon_{h_{in}(z),\tau_{in}-1}.\end{aligned}$$

We can now draw two conclusions. The first conclusion is that in a standard Arkolakis (2008) model (that is when $t_{in}=0$) the dispersion of the export volume distribution is increasing in ad valorem trade costs. To see this note that: (i) the elasticity of *per-consumer* volume to ad-valorem trade costs reduces to $-\sigma(\tau_{in}-1)/\tau_{in}$ and therefore does not depend on the productivity of the firm; (ii) the elasticity of the fraction of consumers reached to ad-valorem trade costs is negative and its absolute value is lower for more productive firms.[77] Therefore, a higher τ reduces the volume sold per consumer for all firms in the same proportion, but the fraction of consumers reached (the customer base) decreases less the higher is the productivity of the firm. The overall volume sold by more productive firms increases relative to that sold by less productive firms.

The second conclusion regards the effects of per-unit trade costs. It turns out that

$$\frac{\partial\varepsilon_{h_{in}(z),t_{in}}}{\partial z}>0\quad iff\quad h_{in}(z)<\frac{\sigma-1}{\beta}\equiv\bar{h},$$

so that, within the set of firms that reach a fraction of consumers lower than $\bar{h}$, the most efficient ones (those with a higher initial customer base) adjust proportionally less the "new consumer margin" than the less efficient firms. Therefore, in this extended model, while more efficient firms are still the ones that decrease the most the volume of goods sold to *each* customer, they are also the ones that reduce less, in percentage terms, their customer base in the event of a rise in per-unit trade costs. The overall effect on the total volume of exports depends on how strong the "marketing effect" is compared to the "per-unit trade costs" effect.

[77] Proof available upon request.

A.6.2 Endogenous Markups

The CES assumption in the main text ensures that markups (over production plus transportation costs) are constant. A model with non-CES preferences will typically generate endogenous markups, which may have an effect on the dispersion of exports. In this section we explore this case, and discuss whether departures from CES alone (with no per-unit costs) can generate the observed correlation between dispersion in exports and trade costs. Specifically, we examine the model of Melitz and Ottaviano (2008), who incorporate endogenous markups using the linear demand system with horizontal product differentiation developed by Ottaviano, Tabuchi, and Thisse (2002). The assumed linear demand system implies that higher prices are associated with higher demand elasticities and therefore lower markups. Specifically, the price charged by an exporter with cost c from country h selling in market n is[78]

$$p^{lh}(c) = \frac{1}{2}\left(c_D^h + \tau^{lh} c\right)$$

where c_D is the domestic cost cutoff (see Appendix A.3 in Melitz and Ottaviano, 2008). Absolute markups are $p^{lh}(c) - \tau c = \frac{1}{2}\left(c_D^h - \tau^{lh} c\right)$, so that more efficient firms, facing lower demand elasticities, are charging higher markups. An increase in trade costs τ^{lh} will in this case lead to more dispersion in prices. To see this, let $c_1 < c_2$, so that $p^{lh}(c_1) < p^{lh}(c_2)$. Then $El_{\tau^{lh}} p^{lh} = c\tau / \left(c_D^h + \tau^{lh} c\right)$, so that prices will increase more, in percentage terms, among the low-efficiency (high cost) firms when τ^{lh} increases. As a consequence, $p^{lh}(c_2) / p^{lh}(c_1)$ rises. The intuition behind this result is that, as τ goes up, markups are reduced the most among high efficiency firms, since they are already charging high markups and face lower demand elasticities.

Naturally, when price dispersion increases, export (volume) dispersion increases as well. Using the expression for optimal exports in Melitz and Ottaviano (2008) we find that relative exports are

$$\frac{q^{lh}(c_1)}{q^{lh}(c_2)} = \frac{c_D^h - \tau^{lh} c_1}{c_D^h - \tau^{lh} c_2}$$

If $c_1 < c_2$, then this ratio increases, i.e. the more efficient firm increases its market share as trade costs rise.

All in all, this shows that introducing a standard model of endogenous markups (with only iceberg costs) will not generate the observed correlation between dispersion in exports and trade costs. However, the structural point estimate of trade costs would surely be affected, introducing endogenous markups. Specifically, since dispersion is increasing with trade costs in Melitz-Ottaviano, an extension of their model with per-unit costs would require higher per-unit costs (compared to what we estimate) in order to match the dispersion in the data. Therefore, we can interpret our estimate as a lower bound of trade costs if endogenous markups are believed to be important.

[78] In this subsection we use, for simplicity, Melitz and Ottaviano (2008) notation.

Table 1: Estimates of per-unit trade costs relative to consumer price

	$\sigma = 4$		$\sigma = 6$		$\sigma = 8$	
Trade costs, mean	.35	(.01)	.36	(.01)	.45	(.01)
Trade costs, median	.33	(.01)	.34	(.01)	.43	(.01)
Trade costs, st. dev.	.12		.13		.12	
γ	1.03	(.03)	1.31	(.03)	1.50	(.03)
Criterion f	558.43		539.22		531.37	
# of Countries (N)	21		21		21	
# of Products (K)	121		121		121	

Note: The mean, median, and standard deviation of trade costs estimates are computed considering all destination-product pairs used. Standard errors in parentheses.

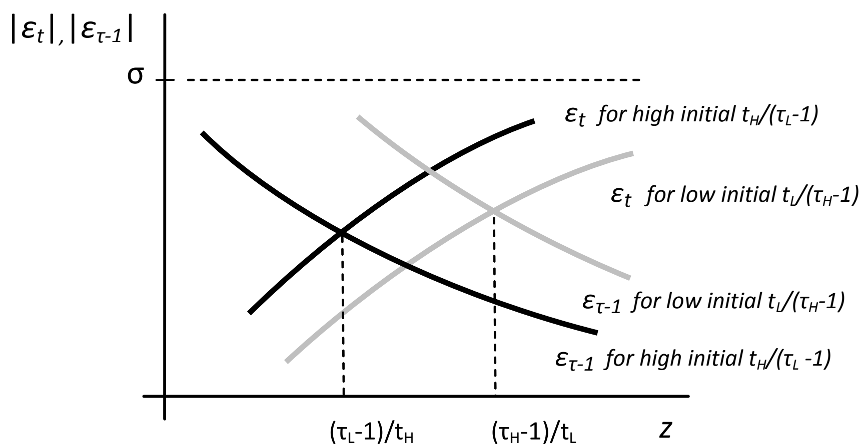

Figure 1: Elasticity of quantity sold to per-unit and ad-valorem trade cost as a function of t, $\tau - 1$ and z.

Table 2: Robustness: Alternative specifications

	(R1)	(R2)	(R3)	(R4)	(R5)	(R6)	(R7)
Trade costs, mean	.34	.36	.43	.44	.36	.34	.27
	(.03)	(.02)	-	(.02)	(.01)	(.02)	-
Trade costs, median	.31	.36	.42	.44	.33	.34	.17
	(.03)	(.02)	-	(.02)	(.01)	(.02)	-
Trade costs, st. dev.	.10	.10	.14	.16	.13	.09	.26
γ	1.16	1.28	1.36	1.26	1.40	.98	2.32
	(.19)	(.10)	-	(.09)	(.04)	(.05)	-
Criterion f	86.1	300.1	6067.2	109.7	493.8	131.1	456.0
# of Countries (N)	3	17	42	6	24	9	19
# of Products (K)	11	43	826	40	116	19	118

Notes: R1: For each product, the set of exporters is identical in every destination; R2: Only homogeneous goods, using Rauch (1999) classification; R3: Only product-destinations with >10 firms; R4: Only products with quantities measured in units; R5: 2003 cross-section; R6: Portuguese data; R7: Heterogeneity in σ and γ. σ=6 used in all specifications except in R7. Four moments are used in R1, R2, and R3, otherwise 12. In R7 the value for γ is an average across all products. Standard errors in parentheses.

Table 3: Robustness: Firm export prices and destination characteristics, 2004

	Norway		Portugal	
(log) Distance	.085^a	.087 ^{a}	.030^a	.033^a
	(.008)	(.008)	(.007)	(.007)
(log) GDP	-.020^a	-.021^a	.004	-.003
	(.004)	(.004)	(.003)	(.002)
(log) GDP per capita		.023		.068^a
		(.017)		(.010)
Firm-product FE	Yes	Yes	Yes	Yes
# of Observations	104,731	104,731	123,660	123,660
# of HS-8 clusters	3,663	3,663	5,730	5,730
# of Firms	15,684	15,684	16,132	16,132
# of Destinations	155	155	174	174

Notes: Dependent variable: f.o.b. unit value by firm, HS8 product and destination. Standard errors (in parentheses) are clustered by HS8 product.
^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

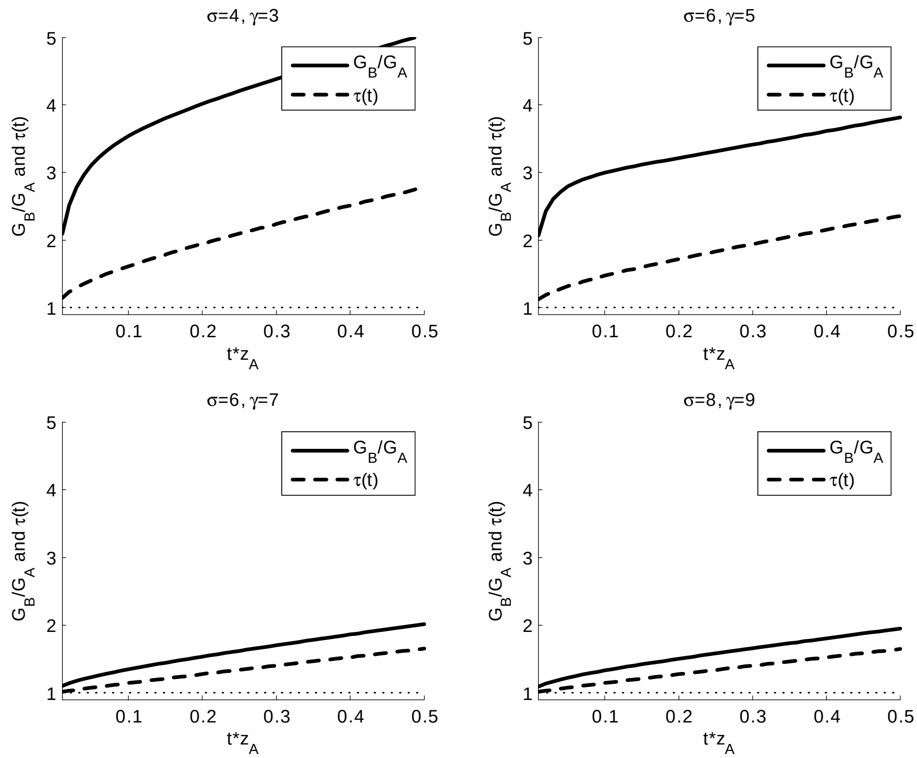

Figure 2: Relative tariff revenue G_B/G_A (A=t barriers, B=τ barriers).

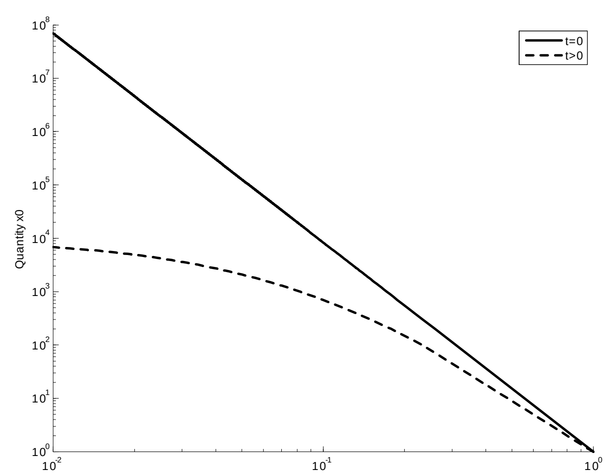

Figure 3: The export volume distribution.

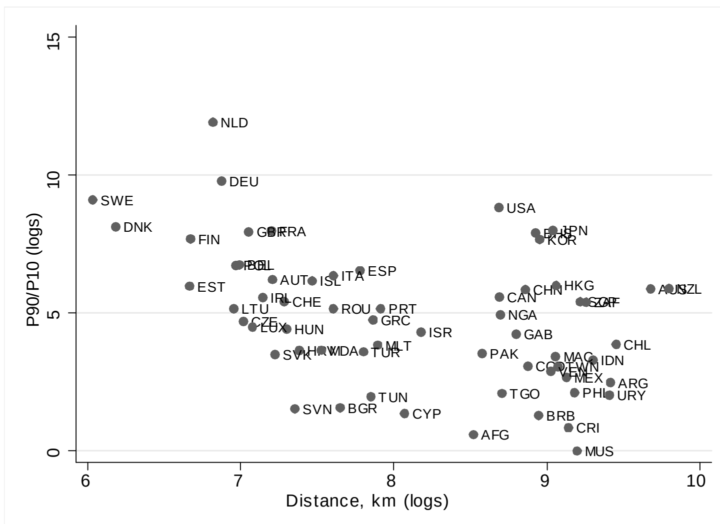

Figure 4: P90/P10 ratio of export quantity, weighted average across products.

Notes: This Figure shows the (log) weighted average (across products) ratio between the 90^{th} and the 10^{th} percentile of the exports volume distribution for several destinations vs. distance (in logs) between Norway and the destination. The weights used in computing each ratio are proportional to the product exports value. A regression that includes distance, population, real GDP per capita (all in logs), and an indicator function for contiguity, together with product fixed effects yields an estimate of the distance coefficient of -.667 (t-stat = -27.10), with standard errors clustered at the product level. Alternative regressions where the dependent variable is the coefficient of variation or the Theil index yield estimates of -.204 (t-stat = -23.99) and -.406 (t-stat = -23.34), respectively.

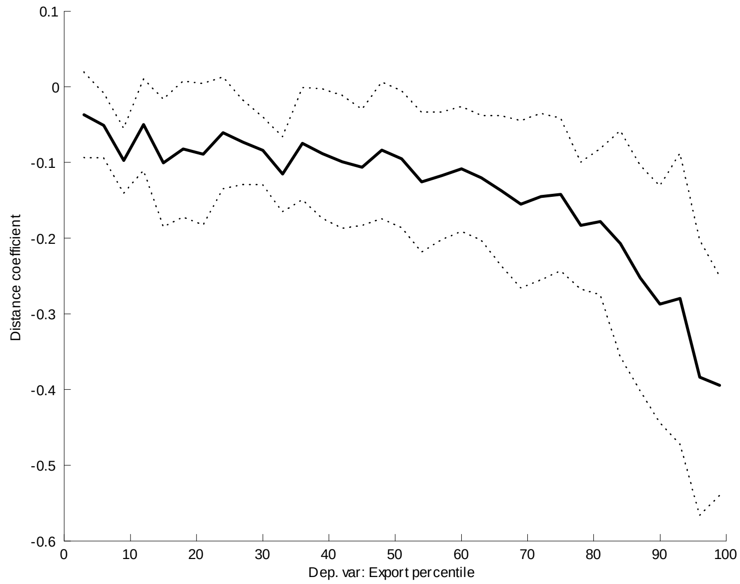

Figure 5: The impact of trade costs on the export distribution.

Notes: This Figure shows (solid line) the estimated distance coefficients (on the y-axis) from regressions of the (log of the) n^{th} percentile of the exports volume distribution against distance, GDP, and GDP per capita (all in logs), and product fixed effects, for $n = 3, 6, ..., 96, 99$. The dotted lines show the 95 percent confidence interval boundaries.

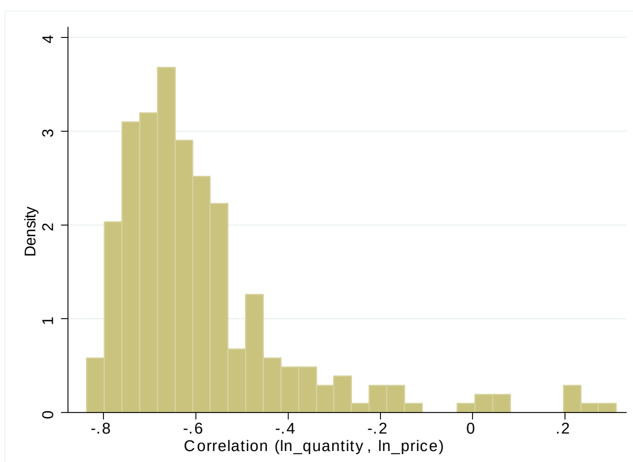

Figure 6: Histogram of corr(lnquantity,lnprice), based on the sample of product-destination pairs used in estimation.

Notes: This Figure shows the distribution of the correlation coefficients between unit value and (quantity) market share (all in logs). Each correlation coefficient corresponds to a specific product-destination pair.

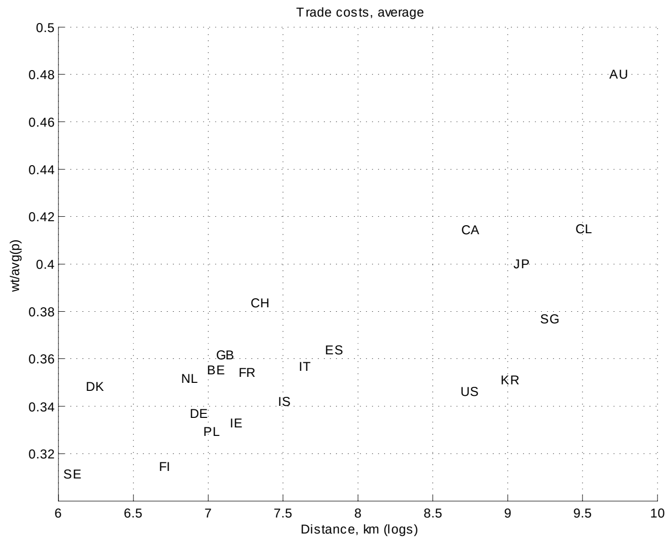

Figure 7: Per-unit trade costs relative to consumer price, averaged across products, conditional on $\sigma = 6$.

Notes: This Figure shows the estimates of $wt_n^k/\bar{p}_n^k$, averaged over products, on the vertical axis and distance (in logs) on the horizontal axis. Note that our two-way fixed effects approach enables us to construct $wt_n^k/\bar{p}_n^k$ even for product-destination pairs that are not present in the data. This implies that there is no selection bias in this Figure, since all products are included in every destination. A regression of (log) $wt_n^k/\bar{p}_n^k$ against distance, GDP, and GDP per capita (all in logs), contiguity, and product fixed effects yields an estimate of the distance coefficient equal to 0.07 (s.e. 0.001).

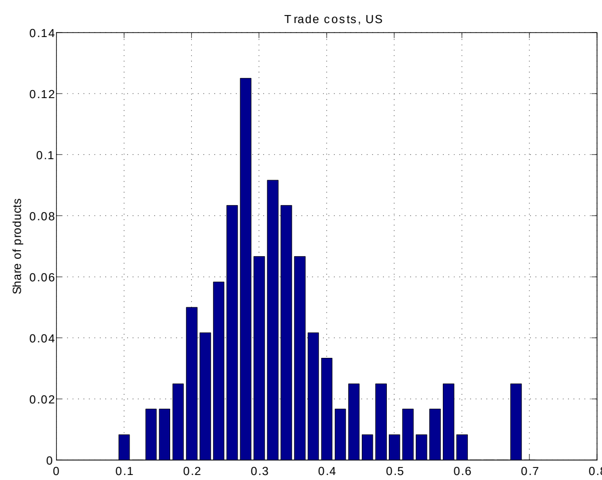

Figure 8: The density of trade costs, conditional on $\sigma = 6$. Norway to the U.S.

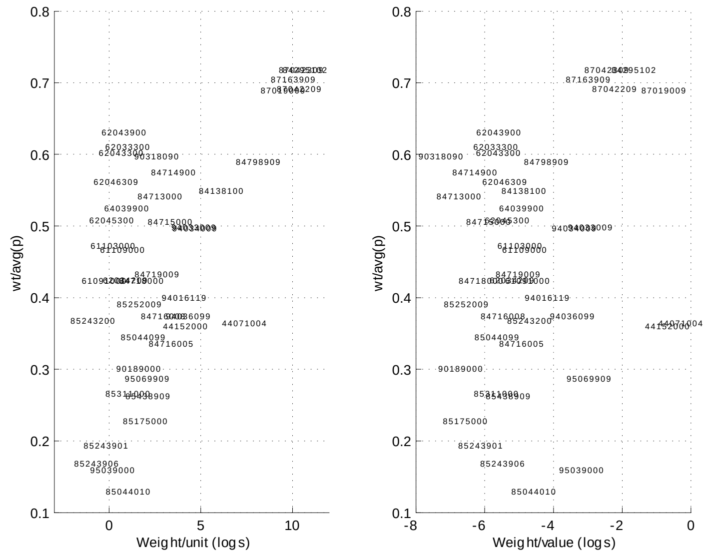

Figure 9: Relationship between estimated trade costs and actual weight/unit or weight/value (in logs).

Notes: This Figure plots the average (over destinations) $wt_n^k/\bar{p}_n^k$ on the vertical axis against the average (log) weight per unit (left panel) and (log) weight per value (right panel). Average weight/unit and weight/value are obtained by dividing total weight (summed over firms and destinations) over total units, or value (summed over firms and destinations).

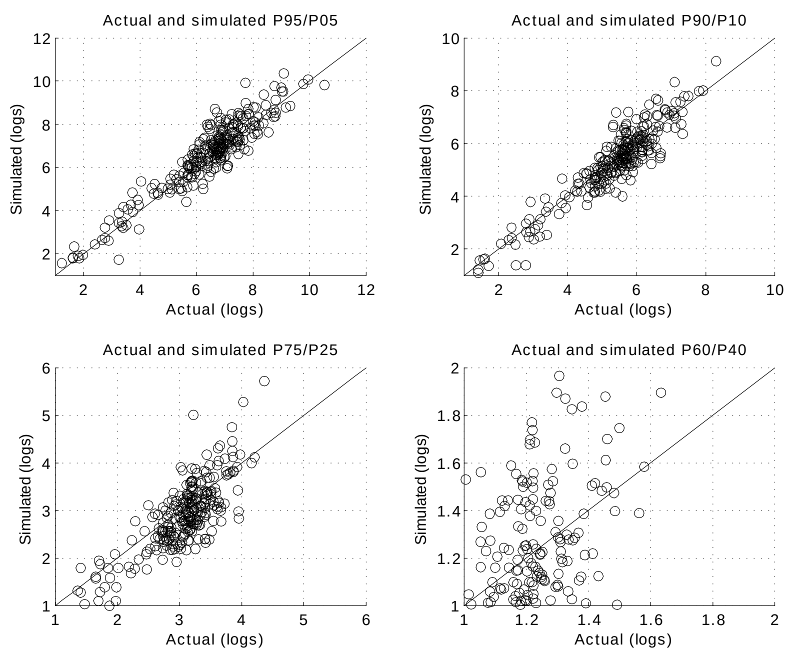

Figure 10: Model evaluation. Empirical and simulated percentile ratios.