# The Tip of the Iceberg: Modeling Trade Costs and Implications for Intra-Industry Reallocation*

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

*Job Market Paper*

January 2010

**Abstract**

When trade costs are of the iceberg type (Samuelson 1952) and markups are independent of trade costs, relative prices across markets are distorted, but relative prices within markets are not. When trade costs depart from the analytically convenient iceberg type, distortion will also occur within markets. In this paper we build a heterogeneous firm model of trade that allows for both iceberg and per-unit costs. An important theoretical finding is that these within-market distortions create an additional channel of gains from trade through within-industry reallocation. We fit the model to firm-level export data, by product and destination, using a novel minimum distance estimator and find that average per-unit costs, expressed relative to the consumer price, are $35 - 45\%$, depending on the elasticity of substitution. The pure iceberg model is therefore rejected. Finally, we calibrate the model and quantify the costs of protectionism. Simulations indicate that the welfare costs are roughly 50% higher when tariffs are per-unit compared to when they are iceberg.

*Acknowledgements : We would like to thank Gregory Corcos, Samuel Kortum, Ralph Ossa and Karen Helene Ulltveit-Moe for their helpful suggestions. We thank Statistics Norway for data preparation and clarifications. 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) of ISEG-Technical University of Lisbon, ldopromolla@bportugal.pt.

*JEL Classification*: F10

*Keywords*: Trade Costs, Intra-Industry Reallocation, Exports, Trade Liberalization

# 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 2001). Since Samuelson (1952), 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 distort the relative price of domestic to foreign goods and therefore distort 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 consisting of 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

[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.

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 Norway is 10.1, 13.2, and 55, respectively, in 2008.[3]

This modeling choice has important consequences when firms are heterogeneous, either in terms of efficiency or quality, as in Melitz (2003), Chaney (2008) or Eaton et al. (2008). When trade costs are incurred per-unit, trade costs not only distort relative prices across markets but also relative prices within markets.[4] Hence, we identify an additional channel of gains from trade through within-industry intensive margin reallocation. The intuition is that more efficient firms, obtaining lower unit 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.[5] As a consequence, per-unit costs tend to wash out the relationship between firm productivity (or quality) 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. In the special case where iceberg is the only type of trade 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. We show that the theoretical implication that relative prices (and therefore quantities) are distorted within markets allows us to retrieve the level of per-unit trade costs (measured as an ad valorem tax equivalent).

[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-favoured 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] 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.

[5] 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.

The third contribution of the paper is to use these estimates and quantify the costs of protectionism in our model compared to a model with iceberg costs exclusively.

Several strong results emerge from the analysis. First of all, per-unit costs are pervasive. The grand mean of 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 protectionism (compared to frictionless trade) are much higher in our model compared to the standard framework. Specifically, calibrating the model with plausible parameter values yields roughly 50% higher welfare costs compared to the iceberg model. The costs in terms of aggregate TFP loss are about three times as high. Therefore, we conclude that the somewhat technical issue of the form of trade costs is quantitatively important for our assessment of the effects of protectionism (and conversely trade liberalization). Furthermore, the benefit of the iceberg model, in terms of analytical tractability, is clearly not 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 well below the unitary elasticity implied by the iceberg assumption. Also, their estimates implied that doubling freight costs increases average 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.[6] 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. Our approach of using within-market dispersion in exports is conceptually different and provides an alternative approach to inferring trade barriers from 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 (although conditional on a value of the elasticity of substitution).

Furthermore, this paper relates to the extensive literature on gains from trade. Most recently, Arkolakis, Costinot and Rodríguez-Clare (2009) 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 due to the intensive margin channel of reallocation

[6] 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.

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 (2003), 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 theory, while Section 3 provides a snapshot of the data as well as lays out the econometric strategy. Section 4 evaluates the welfare effects in our model, while Section 5 concludes.

## 2 Theory

In this section we present a stylized theory of international trade that encompasses both iceberg costs and per-unit costs. This simple modification has important consequences when firms are heterogeneous, either in terms of efficiency or quality. When trade costs are incurred per-unit, they not only distort relative prices across markets but also relative prices within markets. We show that these distortions create an additional channel of gains from trade via within-industry intensive margin reallocation. In the special case where iceberg is the only type of variable trade cost, our model collapses to Chaney (2008).

## 2.1 The Basic Environment

We consider a world economy comprising $N$ potentially asymmetric countries; one factor of production, labor; 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.[7]

Preferences across varieties within a sector $k$ have the standard CES form with an elasticity of substitution $\sigma > 1$.[8] Each variety enters the utility function symmetrically. These preferences generate, in country $n$, for every variety within a sector $k$, a demand 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 $\mu_k$ 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.

## 2.2 Variable Trade Costs

Unlike much of the previous trade literature[9] (e.g. Melitz, 2003, Chaney, 2008, Eaton et al., 2008), 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 with one unit of labor producing $w_n$ units of the service in country $n$. The sector is perfectly competitive, and the price is normalized to one so that if country $n$ produces this service, the wage in the country is $w_n$. We only consider equilibria where every country produces some

[7] 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 a HS8 code.

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

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

of the transport services.

We assume that demand for the shipping service is proportional to the quantity produced (not proportional to value). Depending on shipping destination and product characteristics, $t_{in}^k$ units of labor are necessary for transferring one unit of the good from the firm's plant to its final destination.

Additionally, the economic environment consists of a standard iceberg cost $\tau_{in}^k$, so that $\tau_{in}^k$ 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.

### 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*. Profits are then[10]

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

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,[11]

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

Relative prices within markets are now distorted 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

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

[11] 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 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).

$p_{in}^k(z_1)/p_{in}^k(z_2) = \left(\tau_{in}^k/z_1 + t_{in}^k\right) / \left(\tau_{in}^k/z_2 + t_{in}^k\right)$. 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

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

However, while in previous models the sensitivity of quantity sold (and value of sales) to iceberg trade cost *only* depended on the elasticity of substitution $\sigma$, 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,[12]

$$\begin{aligned} \varepsilon_{t_{in}^k} &= -\sigma \left(\frac{\tau_{in}^k}{z t_{in}^k} + 1\right)^{-1} < 0 \text{ and} \\ \varepsilon_{\tau_{in}^k - 1} &= -\sigma \left(\frac{t_{in}^k z}{\tau_{in}^k} + 1\right)^{-1} \frac{\tau_{in}^k - 1}{\tau_{in}^k} < 0. \end{aligned}$$

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* $\sigma$.

***Proof.*** *See Appendix.* ■

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

The first statement in Proposition 1 emphasizes an asymmetry that affects most of the results in this paper. The first part of the statement says that: i) quantity sold is more sensitive to a change in the per-unit trade cost the higher is the efficiency of the firm while ii) quantity sold is less sensitive to a change in the ad valorem trade cost the higher is the efficiency of the firm. The second part of the first statement says that the effect of a reduction in per-unit trade costs is greater (in terms of quantity sold) than the effect of a reduction in ad valorem trade costs if per-unit costs are greater than iceberg costs ($\left(\tau_{in}^{k}-1\right)/z$ is the iceberg cost converted to labor units for a firm with efficiency $z$).[13] The second statement says that, *for a given firm*, the sensitivity of quantity sold with respect to per-unit trade costs is higher if the per-unit trade cost is initially high relative to the ad valorem trade cost. The opposite is true for changes in the ad valorem trade cost. 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) shows the qualitative relations between the elasticities, firm's efficiency and the variable trade costs. The figure also makes it clear why we expect intensive margin reallocation to occur in the model: The upward sloping curve of $\varepsilon_{t_{in}^{k}}$ means that a reduction in $t_{in}^{k}$ will benefit the high efficiency firms disproportionately more than the low productivity firms, in terms of increased sales. This occurs because lower $t_{in}^{k}$ has a stronger impact on consumer price for high efficiency firms than low efficiency firms, as the share of trade costs in the consumer price is greater for the more efficient firms. As a consequence, factors of production are reallocated from low to high efficiency firms.

[13] 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.

## 2.4 Entry and Cutoffs

We assume that the total mass of potential entrants in country $i$ is proportional to $w_i L_i$ 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. We assume that each consumer owns $w_i$ shares of a totally diversified global fund and that profits are redistributed to them in units of the numéraire good. The total income $Y_i$ spent by workers in country $i$ is the sum of their labor income $w_i L_i$ and of the dividends they get from their portfolio $w_i L_i \pi$, where $\pi$ is the dividend per share of the global mutual fund.

Firms will only enter market $n$ if they can earn positive profits there. Some low-productive 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 $\lambda_1^k$ a constant.[14]

## 2.5 Welfare and Trade Costs

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

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

In the appendix we prove the uniqueness of the price index. In the last two sections in the appendix we also work out the general equilibrium and show how we solve the model numerically.

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

It is not possible to find a closed-form solution for the price index when $t_{in}^k > 0$. However it is possible to study under which conditions the price index reacts more to a change in per-unit trade costs than to a change in ad valorem trade costs. When per-unit trade costs are initially high relative to ad valorem trade costs, the elasticity of the price index with respect to a change in per-unit trade costs is higher than with respect to a change in ad valorem trade costs. In the appendix we prove that a sufficient (but not necessary) condition is $t_{in}^k > \left(\tau_{in}^k - 1\right)/\bar{z}_{in}^k$. The interpretation of this condition is very intuitive: welfare is more sensitive to per-unit costs than to ad valorem costs when per-unit costs are initially higher than ad valorem costs (both expressed in terms of labor) for the least efficient exporter. In the appendix we also prove that the price index is *always* more sensitive to changes in per-unit costs compared to changes in iceberg costs when the initial equilibrium is frictionless. We summarize these finding in the following two propositions:

**Proposition 2** *Consider an initial equilibrium with $t_{in}^k > 0$ and $\tau_{in}^k > 1, \forall i, n, i \neq n$. The costs of trade protectionism in market $n$ are higher, in terms of reduced welfare, when raising $t_{in}^k$ by 1 percent compared to when raising $\tau_{in}^k$ by 1 percent, if $t_{in}^k > \left(\tau_{in}^k - 1\right)/\bar{z}_{in}^k$ (sufficient condition).*

***Proof.*** *See Appendix.* ■

**Proposition 3** *Consider an initial frictionless equilibrium with $t_{in}^k = 0$ and $\tau_{in}^k = 1, \forall i, n$. The costs of trade protectionism in market $n$ are always higher, in terms of reduced welfare, when raising $t_{in}^k$ compared to when raising $\tau_{in}^k$ by a marginal amount.*

***Proof.*** *See Appendix.* ■

## 2.6 The Export Volume Distribution

In this section we examine some properties of the distribution of exports in a model with per-unit costs. We will make extensive use of these properties later on when we estimate trade costs. We first derive the theoretical export 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 CDF $F\left(z|\bar{z}_n^k\right) = 1 - \left(z/\bar{z}_n^k\right)^{-\gamma}$. The Pareto shape coefficient $\gamma$ is assumed to be equal across products and destinations. Then the export volume CDF, conditional on $z > \bar{z}_n^k$, is[15]

$$
\begin{aligned}
Q\left(x|\bar{z}_n^k\right) &= \Pr\left[X < x|Z > \bar{z}_n^k\right] \\
&= 1 - \left[A_n^k x^{-1/\sigma} - B_n^k\right]^{\gamma}, \qquad (3)
\end{aligned}
$$

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

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

### 2.6.1 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 bigger 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 $\gamma/\sigma$. 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:

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

[15] 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.

Dispersion, as measured by the ratio between the $\phi_2^{th}$ and $\phi_1^{th}$ percentiles ($0 < \phi_1 < \phi_2 < 1$) is then

$$D\left(\phi_2, \phi_1; B_n^k, \gamma, \sigma\right) \equiv \frac{x_n^k(\phi_2)}{x_n^k(\phi_1)} = \left[\frac{(1-\phi_1)^{1/\gamma} + B_n^k}{(1-\phi_2)^{1/\gamma} + B_n^k}\right]^{\sigma}. \tag{4}$$

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, export volume becomes less dispersed with higher per-unit costs, controlling for the cutoff $\bar{z}_n^k$ and $\tau_n^k$. 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 4** *When per-unit costs are positive ($t_{in}^k > 0$), dispersion, as measured by the ratio between the $\phi_2^{th}$ and $\phi_1^{th}$ percentiles, is decreasing in $t_{in}^k$ and increasing in $\tau_{in}^k$. Moreover, when per-unit costs are zero ($t_{in}^k = 0$), then dispersion is invariant to a change in variable trade costs $\tau_{in}^k$.*

**Proof.** See appendix. ■

In the appendix we prove this proposition allowing for trade costs to alter the entry hurdles and the price index. The properties of the export volume distribution also survive, under some assumptions, in a framework where firms are heterogeneous both in terms of unit costs and quality.[16]In the appendix we also investigate whether departures from the CES framework can generate similar predictions as a model with per-unit costs. We show that for a popular class of linear demand systems (and

[16] More specifically, the result that dispersion decreases with per unit costs carries through if high price varieties sell less in terms of quantity than low price varieties. This occurs if unit costs are negatively correlated with quality or if they are positively correlated up to a limit. Derivations are available upon request. Johnson (2009) proposes a model where firms are heterogeneous both in terms of unit costs and quality.

with zero per-unit costs), dispersion in exports will increase in ad-valorem costs - the opposite of the case with per-unit costs.

## 3 Estimating the model

In this section we structurally estimate the magnitude of per-unit trade costs. We saw in the theory section that dispersion in export volume falls when per-unit trade costs increase. When per-unit trade costs are zero, dispersion in export 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.[17] The econometric strategy consists of using a minimum distance estimator that matches empirical dispersion in export volume (per product-destination) to simulated dispersion in export volume.[18]

Our approach of estimating trade costs from an economic model is very different from the previous literature.[19] First, most studies model trade costs as ad valorem exclusively, omitting the presence of per-unit costs. A notable exception is Hummels and Skiba (2004), who distinguish between both and find evidence for the presence of per-unit shipping costs.[20] 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 export volume to achieve identification of trade costs, whereas previous 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).

[17] In the data section below, we provide evidence that is consistent with the identifying assumption.

[18] 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).

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

[20] 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.

### 3.1 Data

The data consist of an exhaustive panel of Norwegian non-oil exports in the period 1996-2004.[21] We observe export quantity and export value.[22] Every export observation is associated with a firm, destination and product id. The product ids are based on the HS 8-digit nomenclature, and there are 5391 active HS8 products in the data. 203 unique destinations are recorded in the dataset. Since identification in the empirical model is solely based on cross-sectional variation, we choose to work on the 2004 cross-section, the most recent available to us.

In 2004, $17,480$ firms were exporting and total export value amounted to NOK 232 billion ($\approx$ USD 34.4 billion), or 48 percent of aggregate manufacturing revenue. On average, each firm exported 5.6 products to 3.4 destinations for NOK 13.3 million ($\approx$ USD 2.0 million). On average, there are 3.0 firms per product-destination (standard deviation 7.8). As we will see, we will 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.[23] In the robustness section, we evaluate the effect of this restriction by estimating the model on different sets of destination-product pairs. In what follows, extreme values of quantity sold, defined as values below the 1st 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.[24]

Before presenting the formal econometric model, we show some descriptives that suggest how dispersion is related to trade costs. In Figure 2, we first calculated

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

[22] The unit of measurement is kilos for 67.8% of the products, 27.5% are measured in quantities, while 4.7% are measured in other units ($m^3$, carat, etc.). The choice of unit depends on the product characteristics.

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

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

the ratio between the 90th and 10th percentile of export quantity for each product-destination. Second, we averaged the ratios across products for every destination, using export value for each product as weights.[25] Third, we plotted the mean ratio against distance, in logs. The relationship is clearly negative, indicating that trade costs tend to narrow the dispersion in export quantity. Regressions that include the usual gravity-type right hand side variables and product fixed effects will give the same result.[26] The relationship is also robust to other measures of dispersion, such as the Theil index or the coefficient of variation.

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 distribution charge lower prices than firms in the bottom of the distribution. This 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 $-.32$ (the average over all product-destinations). 84 percent of the correlations are negative.

### 3.2 Estimation

We use a minimum distance estimator that matches empirical dispersion in export volume (per product-destination) to simulated dispersion in export volume. Specifically, denote the empirical ratio between the $\phi_2^{th}$ and $\phi_1^{th}$ percentiles for product $k$ in destination $n$ as $\widetilde{D}_n^k(\phi_2, \phi_1)$ and stack a set of $(\phi_2, \phi_1)$ ratios in the $M \times 1$ column vector $\widetilde{\mathbf{D}}_n^k$. Denote its simulated counterpart $D\left(\phi_2, \phi_1; B_n^k, \gamma, \sigma\right)$, as defined in equation (4), and stack a set of $(\phi_2, \phi_1)$ 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 $\ln \widetilde{\mathbf{D}}_n^k$:

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

[25] 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.

[26] Specifically, we regress the 90/10 percentile ratio on a product fixed effect, distance, population and real GDP per capita (all in logs), as well as contiguity.

where $\mathbf{\Psi}$ is the vector of coefficients to be estimated, $N$ is the total number of destinations and $\Omega_n$ is the set of products sold in market $n$. We minimize $d(\mathbf{\Psi})$ with respect to $\mathbf{\Psi}$ and denote $\widehat{\mathbf{\Psi}}$ the equally weighted minimum distance estimator.[27]

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

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

and normalize $\beta_1 = 1$.[28] 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 $\beta_k$ is estimated relative to some normalization, the estimates of the $B$'s are invariant to the choice of normalization. Finally, we condition the criterion function on a guess of $\sigma$ (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). Hence, we have $M = 12$ moments per product-destination.[29]

As the covariance matrix of the vector of empirical percentile ratios ($\ln \widetilde{\mathbf{P}}_n^k$) 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 per-

[27] Theory suggests that for overidentified models it is best to use optimal GMM. In implementation, however, the optimal GMM estimator may suffer from finit-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 section 6.7).

[28] 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 only sold in a destination where no other products are sold, then one cannot separate the product from the destination effect.

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

forming 500 bootstrap replications, we form the standard errors by calculating the standard deviation for each coefficient in $\mathbf{\Psi}$.

### 3.3 Identification

In Figure 3 we plot the inverse of the theoretical export volume CDF ($1 - CDF$), on log scales. $1 - CDF$ is on the horizontal axis, while quantity exported is on the vertical axis. The solid line represents the case when $B_n^k = t_n^k / \left(\tau_n^k / \overline{z}_n^k\right) = 0$. 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, $1 - CDF$ becomes more and more concave. The set of percentile ratio moments enables us to trace out the curvature of the CDF, which will pin down $B_n^k$.[30] The Pareto shape parameter $\gamma$ is identified by the gradient of the CDF. Since $\gamma$ is independent of product-destination (in the baseline specification), $\gamma$ is identified by the slope of the CDF that is common to all markets, whereas $B_n^k$ is identified by the curvature that is product-destination specific.[31] The economic interpretation is that the higher the per-unit costs (embedded in $B_n^k$), the less dispersion in export volume (captured by more concavity in $1 - CDF$). As it is usual in trade models, the elasticity of substitution $\sigma$ is not identified. The criterion function $d(\mathbf{\Psi})$ is therefore conditional on a guess of $\sigma$. In the results section we report estimates based on different values of $\sigma$.

As already noted, $B_n^k = t_n^k / \left(\tau_n^k / \overline{z}_n^k\right)$ simply measures per-unit trade costs ($t_n^k$) relative to the unit costs of the least efficient firm, inclusive ad-valorem costs ($\tau_n^k / \overline{z}_n^k$). 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

[30] Note that with only one moment, $B_n^k$ and $\gamma$ are not separately identified, as one percentile ratio will only give information about the slope of the CDF. Also note that a linear CDF (in logs) will result in an estimate of zero per-unit trade costs.

[31] A model with product-specific $\gamma$'s is also identified. We estimate a model with heterogeneity in $\gamma$ and $\sigma$ in the robustness section.

firm-level consumer prices can as

$$p_n^k(\widetilde{z}) = \frac{\sigma w t_n^k}{\sigma - 1}\left(\frac{1}{\widetilde{z} B_n^k} + 1\right), \tag{5}$$

where $\widetilde{z}$ is productivity measured relative to the cutoff ($z = \widetilde{z}\bar{z}_n^k$).[32] Second, consider the average price of product $k$ in destination $n$:

$$\bar{p}_n^k = \int_1^{\infty} p_n^k(\widetilde{z})\, dF\left(\widetilde{z} | \bar{z}_n^k = 1\right).$$

Third, inserting equation (5) and solving for $wt_n^k/\bar{p}_n^k$ yields:

$$\frac{wt_n^k}{\bar{p}_n^k} = \left[\frac{\sigma}{\sigma - 1}\int_1^{\infty}\left(\frac{1}{\widetilde{z}B_n^k} + 1\right) dF\left(\widetilde{z}|\bar{z}_n^k = 1\right)\right]^{-1}.$$

The ratio $wt_n^k/\bar{p}_n^k$ measures (per-unit) trade costs relative to the average consumer price. Given our estimate of $B_n^k$ and $\gamma$, the expression on the right hand side can be computed. Note that integrating over productivities allows us to express trade costs only as a function of $B_n^k$, $\gamma$ and $\sigma$. This is due to the fact that a Pareto density is parameterized only by the cutoff ($\bar{z}_n^k$) and the shape parameter ($\gamma$). Our estimates of $B_n^k$ and $\gamma$ are therefore sufficient to get a meaningful measure of per-unit trade costs.[33]

### 3.4 Results

Table 1 summarizes the results.[34] 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 $\sigma = 6$,

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

[33] In the appendix, we consider an extension of our model, that departs from the standard CES framework, 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.

[34] The estimates of $\beta_k$ and $b_n$ are available upon request.

which we use as our baseline case.[35] Estimated trade costs drop to 0.35 for $\sigma = 4$ and rise to 0.45 for $\sigma = 8$. These estimates are similar to 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).[36]

Furthermore, 99 and 95 percent of the $\beta_k$ 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 bigger than zero only when per-unit costs $t_n^k > 0$, our findings suggest that the standard model with only iceberg costs is rejected.[37] The estimate of $\gamma$, the Pareto coefficient, is 1.31 (s.e. 0.03) in the baseline case.

Figure 4 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 (conditional on $\sigma = 6$). Trade costs are clearly increasing in distance. This mirrors the pattern we saw in Figure 2, that dispersion is decreasing 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 4, 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).[38] The distance coefficient is then 0.07 (s.e. 0.001), meaning that doubling distance yields a 7% increase in trade costs.

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

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

[37] We also test the hypothesis that all $t_n^k = 0$ formally. Let $n$ be is 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 $2n\left[d\left(\mathbf{\Psi}^{res};\sigma\right) - d\left(\mathbf{\Psi}^{unres};\sigma\right)\right]$, is $\chi^2(r)$ distributed under the null, where $r$ is the $K + P - 1$ restrictions. The null is rejected at any conventional p-values.

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

Figure 5 shows dispersion in trade costs from Norway to the U.S. across products (conditional on $\sigma = 6$). This figure essentially exploits the variability retrieved from the $\beta_k$ variables. As expected, per-unit trade costs are heterogeneous, with values ranging from roughly 10 to 70 percent of the product value.[39] Figure 6 shows the relationship between estimated trade costs and actual average weight/unit and weight/value in logs.[40] 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 0.55 (0.38).

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 $\beta_k$ and $b_n$ respectively. In the baseline case, the 90-10 percentile ratio of $\beta_k$ 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, 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

[39] Note that densities for other markets are simply shifted left or right compared to the density for 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.

[40] 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 taking the unweighted average of these ratios (in logs) over firms and destinations.

$\Omega_n$ 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.[41] 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 7 shows actual and simulated percentile ratios (95/05, 90/10, 75/25 and 60/40) (again conditional on $\sigma = 6$), 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.

### 3.5 Robustness

A concern in the econometric model 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 the 90/10 ratio is independent of truncation, implying the entry hurdle does not affect dispersion (when $t = 0$). However, under other distributions this is no longer 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 examine dispersion for a subsample of firms that exports a product to many destinations, so that extensive margin effects no longer operate. Specifically, we take the 3 most popular destinations, Sweden, Denmark and Germany, and extract the firm-product pairs that are present in all three markets. This ensures that, for a

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

given product, the same set of firms are present in all locations. We then estimate the function $p_{kn} = \alpha_k + \beta_1 d_n + \beta_2 y_n + \varepsilon_{kn}$, where $p_{kn}$ is the 90/10 percentile ratio, $\alpha_k$ is a product fixed effect, $d_n$ is distance and $y_n$ is GDP (all in logs). Results are shown in Table 2. Column (1) shows the coefficients when the number of firms present in a given product-destination is 2 or more, while columns (2) and (3) show the coefficients when the threshold is 5 and 10 respectively. Even though we lose many observations in this exercise, the results are reassuring. Dispersion is decreasing with distance even among this balanced group of firm-products. The distance coefficient is significant in cases (1) and (2), but not in (3), where the number of products is reduced to 13.

Next we present some re-estimations of the model that address several issues. The results are summarized in Table 3. First, a concern is that although the model presented in the theory section is about single-product firms, our econometric approach treats a multi-product firm as several firms producing different goods. We check the importance of this approach by re-estimating the model on single-product firms only. Specifically, whenever multiple products are exported within a given firm-destination pair, this firm-destination is deleted from the dataset. Naturally, this truncates the data quite substantially, and we are left with only 8 destinations and 6 products (when the product-destination cutoff is set to 40 firms, as before). Nevertheless, the results are reassuring. As shown in Table 3, column (R1), the grand mean of per-unit trade costs is in this case 0.51 (conditional on $\sigma = 6$).

Second, 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 between 30 and 40 firms present and re-estimate the model, resulting in 16 destinations and 149 products. Again, the estimate of trade costs does not change much. The grand mean is now 0.42, as shown in column (R2). Third, 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 6 markets. Again, the results do not change much, as shown in column (R3) in the table. Fourth, we re-estimate the model on the 2003 cross-section instead of the 2004 cross-section. The results in column (R4) show that the grand mean of trade costs is identical to the baseline result. Fifth, we estimate the model on a dataset of Portuguese exporters. The data has the same structure as the Norwegian one. The results in column (R5) show that mean per-unit trade costs for Portugal is 0.34, very close to the Norwegian estimates (for $\sigma = 6$).

We also check the sensitivity of the results to heterogeneity in the elasticity of substitution $\sigma$ and the Pareto coefficient $\gamma$. First, we take estimates of the $\sigma$ from Broda and Weinstein (2006), and take the unweighted average of their HS 10 digit estimates for every 4 digit product.[42] Second, we allow for product-specific $\gamma$'s, so that the theoretical percentile ratios become $D\left(\phi_2, \phi_1; B_n^k, \gamma_k, \sigma_k\right)$ 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 (R6). Again, per-unit costs are large and significant, although the point estimate falls somewhat compared to the baseline case.

Eaton, Kortum and Kramarz (2009) 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.[43] 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, unless one assumes that the variances and/or covariances of the shocks are correlated with distance. Second, our econometric model is expressed in closed-

[42] We average up to the 4 digit level because i) only the first 6 digits are internationally comparable and (ii) not all products are jointly present in the Norwegian and U.S. data.

[43] In a previous paper (Irarrazabal et al 2009) we estimated demand and fixed cost shocks in a model with heterogeneous firms, exports and horizontal FDI.

form, even though analytical expressions for many key relationships do not exist. This helps to keep the run-time of the estimation program down to an acceptable level.[44]

# 4 Simulation: The costs of protectionism

In this section we explore how protectionism, or conversely trade liberalization, will affect welfare and aggregate TFP in the model. As we have seen previously, raising per-unit costs will hit the most productive firms harder than less productive firms, and this will have adverse effects on the aggregate economy. The question here is how strong this effect is quantitatively. We solve three equilibria: (A) Frictionless trade, where both $\tau = 1$ and $t = 0$, (B) Government must raise tariff revenue relative to total import value $(G/I)^*$ through iceberg trade costs $\tau > 1$, and (C) Government must raise $(G/I)^*$ through per-unit costs. To simplify the analysis, we focus on symmetric two-country equilibria.[45] We also remove some heterogeneity by focusing on a single sector. Then, $t_{in}^k = t$ and $\tau_{in}^k = \tau$ for every $k, i, n$.

First, we need to decide on the target tariff revenue we want to obtain. We saw in the previous section that the grand mean trade costs were 0.36 (conditional on $\sigma = 6$). Here we hypothesize that the government raises tariff revenues corresponding to this level of trade costs, i.e. that tariff revenue relative to total c.i.f. imports is $(G/I)^* = 0.36$.

Second, we need to find the value of $\tau^*$ (in case B) and $t^*$ (in case C) that achieves the target tariff revenue $(G/I)^*$. In case (B), the answer is simply $\tau^* = 1.36$.[46] In case (C) the problem is less trivial. Now tariff revenue per import observation is $tx(z)$.

[44] Run-time on a dual Intel Xeon L5520 is approximately 350 seconds.

[45] This implies that import tariffs are retaliated: The foreign country also imposes tariffs on home country exports.

[46] Tariff revenue in absolute terms is $G = \sum (\tau - 1) px = (\tau - 1) I$, where the summation is over every import observation, $x$ is the quantity imported (the quantity that arrives) and $p$ is the consumer price (c.i.f). We assume that the government can convert the melted iceberg into cash by selling it for the market price $p$.

Total tariff revenue is:

$$G^C = wL \int_{\overline{z}(t)} tx\,(t,z)\,dF\,(z)\,. \tag{6}$$

The problem then boils down to finding the $t^*$ that yields $\left(G^C/I\right)^* = 0.36$. The simulation consists of the following steps:

- First choose an initial value for $t^0$ (close to 0), holding $\tau = 1$. Solve the equilibrium and calculate tariff revenue $(G/I)^{C0}$ according to (6).
- If $\left|(G/I)^* - (G/I)^{C0}\right|$ is sufficiently small, the tariff rate $t^*\left((G/I)^*\right)$ that generates tariff revenue $(G/I)^*$ is found. Otherwise, choose a slightly higher $t^1$ and repeat the previous step.

After obtaining $t^*$ we compute the equilibrium in case C and compare welfare in all three cases, as measured by the inverse of the price index, and aggregate TFP, defined as average productivity $z$, weighted by employment for each firm.[47]

The other parameter values are summarized in Table 5 and are chosen as follows. In our first set of simulations, we set $\sigma = 6$ (elasticity of substitution), although we check the results for other values of $\sigma$ as well (see the second and third parameter sets in the table). The market size $Y$ is normalized to $1e+5$. Entry costs are chosen so that all potential entrants enter the domestic market in equilibrium B, which is equivalent to normalizing the home entry hurdle to 1 (for simplicity entry costs are assumed to be the same in the domestic and the foreign market). Finally, the number of potential entrants, which equals the number of random productivity draws, is set so that the accuracy of the numerical approximation of the equilibrium is reasonably high ($1e+5$).

There remains a numerical problem. The estimates of $\gamma$ (the Pareto shape parameter) we found earlier were between 1.0 and 1.5, depending on the choice of $\sigma$. However, as is standard in a Chaney (2008) model, the price index is only defined

[47] Note that this is different from measured TFP, where calculations are typically based on sales, instead of output (quantity).

when $\gamma > \sigma - 1$ (in our case this condition must hold when $t = 0$, i.e. in the frictionless equilibrium). Therefore, we choose to simulate the model using the lowest possible $\gamma$ that gives us a well-defined equilibrium, $\gamma = \sigma - .99$. We check the sensitivity of the choice of $\gamma$ as well (see the third parameter set in the table).

Table 4 presents the percentage change in welfare and TFP from A to B and A to C for all parameter sets. While raising revenue through $\tau$ lowers welfare and TFP by roughly $9 - 11\%$ and $1 - 3\%$ respectively, raising the same amount of revenue through $t$ lowers welfare and TFP by roughly $12 - 18\%$ and $4 - 8\%$ respectively. In other words, in terms of welfare, the costs of protectionism are roughly 50% higher when tariffs are per-unit compared to when tariffs are ad valorem (taking the average effect over all parameterizations). The adverse effect in terms of aggregate TFP is roughly three times higher when tariffs are per-unit compared to when tariffs are ad valorem (taking the average effect over all parameterizations). The flip side is of course that the gains from trade liberalization, either via reductions in tariffs or other trade costs, are much larger when lowering per-unit costs compared to when lowering ad valorem costs.

To better understand the mechanism behind the large differences in welfare and TFP, we plot productivity against domestic and export sales in all three cases in Figure 8. Panels (1) and (2) show domestic and export sales moving from a frictionless world to $\tau-$protectionism (case B), whereas panels (3) and (4) show domestic and export sales moving from a frictionless world to $t-$protectionism (case C). In both cases B and C protectionism translates into a higher price index, which pushes up domestic sales across all firms. Export sales fall in both cases, partly due to firm exit (extensive margin) and partly due to lower sales on the intensive margin. The panel makes clear that the effect on the intensive margin of exports is radically different in case C compared to case B (comparing figure 2 and 4). In case B the decrease is proportional across all exporters. In case C, however, the decrease is much larger, especially among the most productive firms. This anti-reallocation effect, operating on the intensive margin, is what generates the large TFP difference in our model compared to the standard ad valorem case.

## 5 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.

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 within product-destination dispersion in export quantity will fall when (per-unit) trade costs are high, 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%, depending on the elasticity of substitution. We therefore conclude that pure iceberg is rejected at the product level, and that empirical work 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 protectionism by calibrating the model with plausible parameter values. First, we ask what are the costs, in terms of welfare and aggregate TFP, by moving from a frictionless economy to an equilibrium with only iceberg tariffs. Second, we calculate the same costs when moving from a frictionless world to an equilibrium with only per-unit tariffs. It turns out that the intensive margin channel of reallocation induced by per-unit costs is not just a technical matter. The welfare costs are 50% higher when protecting with per-unit tariffs compared to protecting with iceberg tariffs. The costs in terms of aggregate TFP loss are about three times as high. All in all, this suggests that existing estimates of the potential for gains from trade may be too low, and furthermore that the potential for productivity growth induced by trade liberalization may be considerably larger than previously thought. A fairly robust policy implication of our work is therefore 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] John M. Abowd & Robert H. Creecy & Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Working paper

[2] Alchian, Armen A., and William R. Allen. 1964. University Economics. Belmont, Calif.: Wadsworth

[3] Altonji and Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, Vol. 14, No. 3, pp. 353-366

[4] Anderson, James E., van Wincoop, Eric. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, Volume 93, Issue 1, pp. 170-192

[5] Anderson, James E., van Wincoop, Eric. 2004. "Trade Costs," *Journal of Economic Literature*, Volume 42, Number 3, 1 September 2004 , pp. 691-751(61)

[6] Arkolakis, K., (2008) "Market Penetration Costs and the New Consumers Margin in International Trade", mimeo Yale University.

[7] Arkolakis, K., Costinot, A. and A. Rodríguez-Clare. 2009. "New Theories, Same Old Gains?," Working paper

[8] Berman, N., Martin, P. and T. Mayer. 2009 "How Do Different Exporters React to Exchange Rate Changes? Theory, Empirics and Aggregate Implications." Manuscript.

[9] Bernard, A., Jensen, B. and P. Schott. 2006. "Transfer Pricing by US-Based Multinational Firms." NBER Working Paper 12493

[10] C. Broda and D. Weinstein. 2006. "Globalization and the Gains from Variety," *Quarterly Journal of Economics,* Volume 121, Issue 2

[11] Cameron A.C. and P. K. Trivedi. 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

[12] Christian Broda & David Weinstein, 2006."Globalization and the gains from variety," *Quarterly Journal of Economics*, Vol. 121, No. 2, Pages 541-585

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

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

[15] Eaton, J., Kortum, S. and F. Kramarz. 2008. "An Anatomy of International Trade: Evidence from French Firms." NBER Working Paper No. 14610.

[16] Keith Head and John Ries. 2001. "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade," *The American Economic Review,* Vol. 91, No. 4, pp. 858-876

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

[18] Hummels, D. and A. Skiba (2004) "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture." *Journal of Political Economy*, 112(6) pp. 1384-1402.

[19] David Hummels, Peter J. Klenow, 2005. "The Variety and Quality of a Nation's Exports," *The American Economic Review*, Vol. 95, No. 3

[20] Irarrazabal, A., Moxnes A. and L.D. Opromolla. 2008. "The Margins of Multinational Production and the Role of Intra-firm Trade." Manuscript.

[21] Jacks, D. S., C. M. Meissner, and D. Novy, "Trade Costs, 1870–2000" American Economic Review: Papers & Proceedings 2008, 98:2, 529–534

[22] Johnson, R. C. 2009. "Trade and Prices with Heterogeneous Firms," mimeo.

[23] Timothy J. Kehoe & Kim J. Ruhl, 2006. "How Important is the New Goods Margin in International Trade?," 2006 Meeting Papers 733, Society for Economic Dynamics.

[24] Klenow, P. and Rodríguez-Clare, A., 1997. "Quantifying variety gains from trade liberalization", Working paper.

[25] Lugovskyy, V. and Skiba, A., 2009. "Quality Choice: Effects of Trade, Transportation Cost, and Relative Country Size," manuscript.

[26] Erzo G. J. Luttmer, 2007. "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, MIT Press, vol. 122(3), pages 1103-1144, 08.

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

[28] Maurice Obstfeld and Kenneth Rogoff. 2001. "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?," NBER Macroeconomics Annual, Vol. 15, (2000), pp. 339-390

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

[30] Samuelson, P. A., 1952. "The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent," *The Economic Journal*, Vol. 62, No. 246 (Jun., 1952), pp. 278-304

[31] Verhoogen, E. 2008. "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, MIT Press, vol. 123(2), pages 489-530

# A Appendix

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

**Proof.** $|\varepsilon_{t_{in}^k}|$ is increasing in $z$ while $|\varepsilon_{\tau_{in}^k-1}|$ is decreasing in $z$.

The relevant derivatives are

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

where $\sigma > 1$, $\tau_{in}^k \geq 1$ and $t \geq 0$. ■

**Proof.** $|\varepsilon_{t_{in}^k}|$ 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 relevant derivatives are

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

where $\sigma > 1$, $z \geq 1$, $t_{in}^k \geq 0$, $\partial\left(t_{in}^k/\tau_{in}^k\right)/\partial\left[t_{in}^k/\left(\tau_{in}^k-1\right)\right] = \left(\tau_{in}^k-1\right)^2/\left(\tau_{in}^k\right)^2$ and $\partial\left[\tau_{in}^k/\left(\tau_{in}^k-1\right)\right]/\partial\left[t_{in}^k/\left(\tau_{in}^k-1\right)\right] = 1/t_{in}^k$. ■

**Proof.** $|\varepsilon_{t_{in}^k}| > |\varepsilon_{\tau_{in}^k-1}|$ if if $z > \left(\tau_{in}^k-1\right)/t_{in}^k$.

The relevant inequalities are

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

∎

**Proof.** $|\varepsilon_{t_{in}^k}|$ and $|\varepsilon_{\tau_{in}^k-1}|$ have an upper bound equal to $\sigma$ and respectively.

$$\begin{aligned} Sup_{z,t,\tau}|\varepsilon_{t_{in}^k}| &= \lim_{\frac{\tau_{in}^k}{z t_{in}^k}\to 0} |\varepsilon_{t_{in}^k}| = \sigma \text{ and} \\ Sup_{z,t,\tau}|\varepsilon_{\tau_{in}^k-1}| &= \lim_{\frac{\tau_{in}^k}{\tau_{in}^k-1}+\frac{t_{in}^k z}{\tau_{in}^k-1}\to 0} |\varepsilon_{qty,\tau-1}| = \sigma. \end{aligned}$$

∎

## A.2 Uniqueness of the Equilibrium Price Index

The price index relative to sector $k$ in country $n$ is defined as

$$\begin{aligned} \left(P_n^k\right)^{1-\sigma} &= \int_{\bar{z}_{nn}^k}^{\infty} w_n L_n \left(\frac{\sigma}{\sigma-1}\frac{w_n}{z}\right)^{1-\sigma} \frac{\gamma}{z^{\gamma+1}} dz \\ &\quad + \sum_{i\neq n} \int_{\bar{z}_{in}^k}^{\infty} w_i L_i \left[\frac{\sigma}{\sigma-1} w_i \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)\right]^{1-\sigma} \frac{\gamma}{z^{\gamma+1}} dz. \end{aligned}$$

Rearranging terms,

$$\underbrace{\frac{\left(P_n^k\right)^{1-\sigma}}{\gamma\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}} - \int_{\bar{z}_{nn}^k}^{\infty} \frac{w_n L_n \left(w_n/z\right)^{1-\sigma}}{z^{\gamma+1}} dz}_{I_1} = \underbrace{\sum_{i\neq n} \int_{\bar{z}_{in}^k}^{\infty} \frac{w_i L_i \left(w_i \tau_{in}^k/z + w_i t_{in}^k\right)^{1-\sigma}}{z^{\gamma+1}} dz}_{I_2}.$$

Note that the left-hand-side is strictly decreasing in $P_n^k$ while the right-hand-side is strictly increasing in $P_n^k$ (recall that $\bar{z}_{nn}^k$ and $\bar{z}_{in}^k$ are both decreasing in $P_n^k$). Therefore there is a unique equilibrium $P_n^k$.

## A.3 Effect of $d\tau$ and $dt$ on the Price Index (Propositions 2 and 3)

Denote with $I_1$ the left-hand-side of the previous equation and with $I_2$ the right-hand-side. Note that $I_1$ does not depend on $t_{in}^k$ or $\tau_{in}^k$ while $I_2$ does. Using Leibnitz' formula we take the derivative of $I_2$ with respect to $t_{in}^k$ and $\tau_{in}^k$ (keeping $P_n^k$ constant). The first term of each derivative represents the intensive margin effect of a change in trade costs on the price index. The second term represents the extensive margin effect.

$$\frac{\partial I_2}{\partial t_{in}^k} = L_i w_i^{2-\sigma} \left[ \int_{\bar{z}_{in}^k}^{\infty} \frac{(1-\sigma)}{z^{\gamma+1}} \left( \frac{\tau_{in}^k}{z} + t_{in}^k \right)^{-\sigma} dz - \frac{\left( \frac{\tau_{in}^k}{\bar{z}_{in}^k} + t_{in}^k \right)^{1-\sigma}}{\left(\bar{z}_{in}^k\right)^{\gamma+1}} \frac{\partial \bar{z}_{in}^k}{\partial t_{in}^k} \right];$$

$$\frac{\partial I_2}{\partial \tau_{in}^k - 1} = L_i w_i^{2-\sigma} \left[ \int_{\bar{z}_{in}^k}^{\infty} \frac{(1-\sigma)}{z^{\gamma+2}} \left( \frac{\tau_{in}^k}{z} + t_{in}^k \right)^{-\sigma} dz - \frac{\left( \frac{\tau_{in}^k}{\bar{z}_{in}^k} + t_{in}^k \right)^{1-\sigma}}{\left(\bar{z}_{in}^k\right)^{\gamma+1}} \frac{\partial \bar{z}_{in}^k}{\partial \tau_{in}^k - 1} \right].$$

where $\partial \bar{z}_{vn}^k / \partial t_{in}^k = 0$ if $v \neq n$ since we are keeping $P_n^k$ constant. Note that the first term of $\partial I_2 / \partial t_{in}^k$ is greater than the first term of $\partial I_2 / \partial \left(\tau_{in}^k - 1\right)$ since both $\sigma$ and $\bar{z}_{in}^k$ are greater than one. In both cases, the first term is negative. To complete the analysis, we can, recalling the expression for $\bar{z}_{in}^k$,

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

compute the derivatives

$$\frac{\partial \bar{z}_{in}^k}{\partial t_{in}^k} = \frac{\left(\bar{z}_{in}^k\right)^2}{\tau_{in}^k} > 0 \text{ and}$$

$$\frac{\partial \bar{z}_{in}^k}{\partial \tau_{in} - 1} = \frac{\bar{z}_{in}^k}{\tau_{in}^k} > 0 \text{ if } \bar{z}_{in}^k > 0,$$

so that

$$\frac{\partial I_2}{\partial t_{in}^k} = L_i w_i^{2-\sigma} \left[ \int_{\bar{z}_{in}^k}^{\infty} \frac{(1-\sigma)}{z^{\gamma+1}} \left( \frac{\tau_{in}^k}{z} + t_{in}^k \right)^{-\sigma} dz - \frac{\left( \frac{\tau_{in}^k}{\bar{z}_{in}^k} + t_{in}^k \right)^{1-\sigma}}{\left(\bar{z}_{in}^k\right)^{\gamma+1}} \frac{\left(\bar{z}_{in}^k\right)^2}{\tau_{in}^k} \right] < 0,$$

$$\frac{\partial I_2}{\partial \tau_{in}^k - 1} = L_i w_i^{2-\sigma} \left[ \int_{\bar{z}_{in}^k}^{\infty} \frac{(1-\sigma)}{z^{\gamma+2}} \left( \frac{\tau_{in}^k}{z} + t_{in}^k \right)^{-\sigma} dz - \frac{\left( \frac{\tau_{in}^k}{\bar{z}_{in}^k} + t_{in}^k \right)^{1-\sigma}}{\left(\bar{z}_{in}^k\right)^{\gamma+1}} \frac{\bar{z}_{in}^k}{\tau_{in}^k} \right] < 0.$$

Since $\bar{z}_{in}^k > 1$ the second term in $\partial I_2 / \partial t_{in}^k$ is greater than the second term in $\partial I_2 / \left(\partial \tau_{in}^k - 1\right)$ so that we can conclude that

$$\frac{\partial I_2}{\partial t_{in}^k} < \frac{\partial I_2}{\partial \tau_{in}^k - 1} < 0,$$

and establish that a reduction in $t_{in}^k$ increases $I_2$ more and therefore lowers $P_n^k$ more than a reduction in $\tau_{in}^k - 1$.

We can re-write the expressions in terms of elasticities

$$\varepsilon_{I_2,t_{in}^k} = \frac{L_i w_i^{2-\sigma} t_{in}^k}{I_2}\left[\int_{\bar{z}_{in}^k}^{\infty} \frac{(1-\sigma)}{z^{\gamma+1}}\left(\frac{\tau_{in}^k}{z}+t_{in}^k\right)^{-\sigma} dz - \frac{\left(\frac{\tau_{in}^k}{\bar{z}_{in}^k}+t_{in}^k\right)^{1-\sigma}}{\tau_{in}^k\left(\bar{z}_{in}^k\right)^{\gamma-1}}\right] < 0,$$

$$\varepsilon_{I_2,\tau_{in}^k-1} = \frac{L_i w_i^{2-\sigma}\left(\tau_{in}^k-1\right)}{I_2}\left[\int_{\bar{z}_{in}^k}^{\infty} \frac{(1-\sigma)}{z^{\gamma+2}}\left(\frac{\tau_{in}^k}{z}+t_{in}^k\right)^{-\sigma} dz - \frac{\left(\frac{\tau_{in}^k}{\bar{z}_{in}^k}+t_{in}^k\right)^{1-\sigma}}{\left(z_{in}^k\right)^{\gamma}\tau_{in}^k}\right] < 0.$$

From here we can see that a sufficient condition (but not necessary) for $\varepsilon_{I_2,t_{in}^k} < \varepsilon_{I_2,\tau_{in}^k-1} < 0$ is $t_{in}^k > \left(\tau_{in}^k - 1\right)\bar{z}_{in}^k$, i.e. that per-unit costs are higher than iceberg costs for every exporter ($\left(\tau_{in}^k-1\right)/\bar{z}_{in}^k$ is the iceberg cost in terms of labor usage for the exporter with the lowest efficiency).

Finally we show that protectionism is always more costly, in terms of reduced welfare, when raising $t_{in}^k$ compared to raising $\tau_{in}^k$ when the initial equilibrium is frictionless. Consider the change in $I_2$ when the initial equilibrium is characterized by $\tau_{in}^k = 1$ and $t_{in}^k = 0$,

$$\frac{\partial I_2}{\partial t_{in}^k}|_{\tau_{in}^k=1,t_{in}^k=0} = (1-\sigma)\, L_i w_i^{2-\sigma}\int_{\bar{z}_{in}^k}^{\infty} z^{\sigma} dF(z) - \left(\bar{z}_{in}^k\right)^{\sigma-1}\left(\bar{z}_{in}^k\right)^{-\gamma-1}\frac{z_{in}^2}{\tau_{in}} < 0,$$

$$\frac{\partial I_2}{\partial \tau_{in}^k - 1}|_{\tau_{in}^k=1,t_{in}^k=0} = (1-\sigma)\, L_i w_i^{2-\sigma}\int_{\bar{z}_{in}^k}^{\infty} z^{\sigma-1} dF(z) - \left(\bar{z}_{in}^k\right)^{\sigma-1}\left(\bar{z}_{in}^k\right)^{-\gamma-1}\frac{\bar{z}_{in}^k}{\tau_{in}^k} < 0.$$

Note that we choose to express changes as derivatives instead of elasticities since percentage changes are meaningless when the initial $t_{in}^k$ is zero. Clearly, $\partial I_2/\partial t_{in}^k < \partial I_2/\partial\left(\tau_{in}^k - 1\right)$ is satisfied as long as the marginal change in $t_{in}^k$ and $\tau_{in}^k$ is comparable.[48] Hence the costs of protectionism are *always* higher when increasing $t_{in}^k$ compared to when increasing $\tau_{in}^k$.

[48] Recall that $\sigma > 1$ and $\bar{z}_{in}^k > 1\ \forall i, n, k$.

## A.4 The export volume distribution (Proposition 4)

**Proof.** Recall that the percentile ratio is

$$D_{\phi_2/\phi_1} = \frac{x_n^k(\phi_2)}{x_n^k(\phi_1)} = \left[\frac{(1-\phi_1)^{1/\gamma} + B_n^k}{(1-\phi_2)^{1/\gamma} + B_n^k}\right]^\sigma,$$

where $B_n^k = \bar{z}_n^k t_n^k / \tau_n^k$. Consider the impact on $D_{\phi_2/\phi_1}$ of a small change in $t_n^k$,

$$\begin{aligned} \frac{dD_{\phi_2/\phi_1}}{D} &\equiv \widehat{D}_{\phi_2/\phi_1} \\ &= \sigma D^{-1/\sigma} \frac{(1-\phi_2)^{1/\gamma} - (1-\phi_1)^{1/\gamma}}{\left[(1-\phi_2)^{1/\gamma} + B_n^k\right]^2} \widehat{B}_n^k B_n^k. \end{aligned}$$

The fraction is negative since $0 < \phi_1 < \phi_2 < 1$. It remains to evaluate $\widehat{B}_n^k$:

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

The change in the cutoff is

$$d\bar{z}_n^k = -\left(\bar{z}_n^k\right)^2 \left[\left(\left(\bar{z}_{in}^k\right)^{-1} + \frac{t_{in}^k}{\tau_{in}^k}\right)\widehat{P}_n^k - \frac{t_{in}^k}{\tau_{in}^k}\widehat{t}_{in}^k\right],$$

or equivalently

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

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

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

Since $\widehat{t}_n^k > \widehat{P}_n^k$ (see proof below), $\widehat{B}_n^k > 0$ and $\widehat{D}_{\phi_2/\phi_1} < 0$. Therefore, dispersion measured by the $D_{\phi_2/\phi_1}$ percentile ratio is declining as per-unit trade costs rise. ■

**Proof.** Consider the impact on $D_{\phi_2/\phi_1}$ on a small change in $\tau_n^k$. The expression for $\widehat{D}_{\phi_2/\phi_1}$ remains the same as above, but $\widehat{B}_n^k$ now becomes

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

and the change in cutoff becomes

$$d\bar{z}_n^k = -\left(\bar{z}_n^k\right)^2 \left[\left(\left(\bar{z}_{in}^k\right)^{-1} + \frac{t_{in}^k}{\tau_{in}^k}\right)\left(\widehat{P}_n^k - \widehat{\tau}_{in}^k\right) + \frac{t_{in}^k}{\tau_{in}^k}\frac{\tau_n^k - 1}{\tau_n^k}\widehat{\tau}_{in}^k\right],$$

or equivalently

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

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

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

Note that $\widehat{B}_n^k < 0$ since $\widehat{P}_n^k > 0$. Therefore $\widehat{D}_{\phi_2/\phi_1} > 0$ and dispersion rises when iceberg costs increase. ■

**Proof.** Consider the impact on $D_{\phi_2/\phi_1}$ on a small change in $\tau_n^k$ when $t_{in}^k = 0$. The percentile ratio then collapses to

$$D_{\phi_2/\phi_1} = \left[\frac{(1-\phi_1)^{1/\gamma}}{(1-\phi_2)^{1/\gamma}}\right]^{\sigma},$$

and $dD_{\phi_2/\phi_1} = 0$, showing that dispersion is independent of variable trade costs in the Chaney (2008) model. ■

### A.5 The change in $P_n$ relative to $t_{in}$

**Proof.** Above we stated that $\widehat{t}_n^k > \widehat{P}_n^k$. Price index is defined as

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

where $F_1(z)$ is the Pareto CDF with support $z \in [1, +\infty)$ and $f_1(z)$ is the PDF. For now, we will consider only sector $k$. The other sectors are analogous. For notational clarity, we drop the $k$ subscript and all sectoral variables will refer to sector $k$ when there is no ambiguity.

Aggregate sales (c.i.f.) from $i$ to $n$ is

$$\begin{aligned} S_{in} &= w_i L_i \int_{\bar{z}_{in}} p_{in} x_{in} dF_1(z) \\ &= w_i L_i \int_{\bar{z}_{in}} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} w_i^{1-\sigma}\left(\tau_{in}/z + t_{in}\right)^{1-\sigma} P_n^{\sigma-1}\mu Y_n dF_1(z), \end{aligned}$$

where $x_{in}$ is quantity demanded and $p_{in}$ is consumer price $p_{in} = \sigma / (\sigma - 1) w_i (\tau_{in}/z + t_{in})$. Define $S_n \equiv \sum_i S_{in}$. The share of country $n$'s total expenditure that is devoted to goods from country $i$ (the import share) is then

$$\begin{aligned} \lambda_{in} &= S_{in}/S_n \\ &= \frac{w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}} (\tau_{in}/z + t_{in})^{1-\sigma} dF_1(z)}{\sum_j w_j L_j w_j^{1-\sigma} \int_{\bar{z}_{jn}} (\tau_{jn}/z + t_{jn})^{1-\sigma} dF_1(z)}. \end{aligned}$$

Consider a percentage change in the price index $\widehat{P}_n \equiv dP_n/P_n$ due a marginal change in $t_{in}$, $\forall i$,

$$\begin{aligned} \widehat{P}_n &= \frac{1}{1-\sigma} \frac{\sum_i w_i L_i w_i^{1-\sigma} dI_{in}}{\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}} \left(\frac{\tau_{in}^k}{z} + t_{in}^k\right)^{1-\sigma} dF_1(z)} \\ &= \frac{\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}} \left(\frac{\tau_{in}}{z} + t_{in}\right)^{1-\sigma} dF_1(z) \left[\widehat{I}_{in}/(1-\sigma)\right]}{\sum_i w_i L_i w_i^{1-\sigma} \int_{\bar{z}_{in}} \left(\frac{\tau_{in}}{z} + t_{in}\right)^{1-\sigma} dF_1(z)} \\ &= \sum_i \lambda_{in} \frac{\widehat{I}_{in}}{1-\sigma}, \end{aligned} \tag{7}$$

where we defined

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

and $dI_{in}$ is the change in $I_{in}$ due to a marginal change in $t$.

The change in $I_{in}$ is

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

where

$$\begin{aligned} \chi_{in} &= \frac{\int_{\bar{z}_{in}} t_{in} \left(\frac{\tau_{in}}{z} + t_{in}\right)^{-\sigma} dF_1(z)}{\int_{\bar{z}_{in}} \left(\frac{\tau_{in}}{z} + t_{in}\right)^{1-\sigma} dF_1(z)} \\ \bar{\chi}_{in} &= \frac{\left(\frac{\tau_{in}}{\bar{z}_{in}} + t_{in}\right)^{1-\sigma} f_1(\bar{z}_{in})}{\int_{\bar{z}_{in}} \left(\frac{\tau_{in}}{z} + t_{in}\right)^{1-\sigma} dF_1(z)}. \end{aligned}$$

Note that $\chi_{in}$ is always less than one.

Consider a change in per-unit costs from $k$ to $n$, $dt_{kn}$. Disregard any possible second-order effects so that $d\bar{z}_{in} = 0$ for $i \neq k$. Then

$$\begin{aligned}
\widehat{P}_n &= \lambda_{kn} \frac{\widehat{I}_{kn}}{1-\sigma} \\
&= \lambda_{kn} \left( \chi_{kn} \widehat{t}_{kn} + \frac{\bar{\chi}_{kn}}{\sigma - 1} \bar{z}_{kn} \widehat{\bar{z}}_{kn} \right).
\end{aligned}$$

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. Next we use the expression for $\widehat{\bar{z}}_{kn}$, insert into $\widehat{P}_n$ and solve for $\widehat{P}_n$. This yields

$$\widehat{P}_n = \frac{\lambda_{kn}\chi_{kn} + \lambda_{kn} \frac{\bar{\chi}_{kn}}{\sigma-1} \bar{z}_{in} \left( \bar{z}_n \frac{t_{in}}{\tau_{in}} \right)}{1 + \lambda_{kn} \frac{\bar{\chi}_{kn}}{\sigma-1} \bar{z}_{in} \left( 1 + \bar{z}_n \frac{t_{in}}{\tau_{in}} \right)} \widehat{t}_{in}$$

Since $\lambda_{kn}\chi_{kn} < 1$, the fraction is less than 1 and therefore $\widehat{P}_n < \widehat{t}_{in}$. ■

### A.6 Simulating the model

The numerical approximation of the equilibrium consists of the following steps. For simplicity we restrict the number of products to one.

1. Choose a starting value of the price index $P_n^0$. Superscripts denote the round of iteration.

2. Solve the equilibrium cutoffs and global profits simultaneously, conditional on $P_n^0$. The cutoffs and global profits are

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

   where only the endogenous arguments in functions $f$ and $g$ are explicitly shown. The expression for $\pi$ 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 $\pi$, solving $\bar{z}_{in}$ using $f$, inserting the solution back into $g$, etc., until the system converges.

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

4. Iterate over 2 and 3. When $\left|P_n^r - P_n^{r-1}\right|$ is sufficiently small, the equilibrium $\{P_n, z_{in}, \pi\}$ 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 $g\left(z\right)$. An integral of the form

$$P = \int_{Z}^{\infty} p\left(z\right)^{1-\sigma} g\left(z\right) dz$$

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

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

### A.6.1 Global profits

Following Chaney (2008), we assume that each worker owns $w_n$ shares of a global fund. The fund collects global profits $\Pi$ 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 $Y_n = w_n L_n + w_n L_n \pi = w_n L_n \left(1 + \pi\right)$. 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 $\pi$ 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 $n_{in} = w_i L_i \int_{\bar{z}_{in}} dF_1(z) = w_i L_i \bar{z}_{in}^{-\gamma}$, $\pi$ 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.7 Extension: 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:

$$\max_{n_{in}(z), p_{in}(z)} x_{in}(z)\left[p_{in}(z) - w_i\left(\frac{\tau_{in}}{z} + t_{in}\right)\right] - w_n^{\theta} w_i^{1-\theta} \frac{L_n^{\alpha}}{\psi} \frac{1 - [1 - m_{in}(z)]^{1-\beta}}{1-\beta},$$

$$\text{s.t. } \; m_{in}(z) \in [0,1].$$

where demand is

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

$y_n$ is per-capita spending in country $n$, $L_n$ is population of country $n$ and $m_{in}(z)$ 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 volume of goods exported to country $n$ to trade costs include instead an additional term that reflects how the "new consumer margin" reacts to changes in trade costs,

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

It turns out that $\varepsilon_{n_{ij}(z),t_{ij}} < 0$ and

$$\frac{\partial \varepsilon_{n_{ij}(z),t_{ij}}}{\partial z} > 0 \quad iff \quad m_{ij}(z) < \frac{\sigma - 1}{\beta} \equiv \bar{m}$$

so that, within the set of firms that reach a fraction of consumers lower than $\bar{m}$, 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 export volume depends on how strong is the "marketing effect" compared to the "per-unit trade costs" effect.

### A.8 Extension: Endogenous markups

The CES assumption in the main text ensures that markups 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 $l$ is

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

where $c_D$ is the domestic cost cutoff (see Melitz and Ottaviano 2008 appendix A.3). 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

$\tau^{lh}$ 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 then $\tau^{lh}$ increases. As a consequence, $p^{lh}(c_2) / p^{lh}(c_1)$ rises. The intuition behind this result is that, as $\tau$ 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 appendix A.3 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.

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, stdev | .12 | | .13 | | .12 | |
| $\gamma$ | 1.03 | (.03) | 1.31 | (.03) | 1.50 | (.03) |
| Criterion $f$ | 558.43 | | 539.22 | | 531.37 | |
| $N$ markets | 21 | | | | | |
| $K$ products | 121 | | | | | |

Note: Standard errors in parentheses.

Table 2: Robustness: Controlling for entry

| | (1) | | (2) | | (3) | |
|---|---|---|---|---|---|---|
| Distance | −.92 | (.46) | −3.11 | (1.03) | −2.64 | (1.62) |
| GDP | .24 | (.14) | .90 | (.33) | .80 | (.52) |
| Product FEs | Yes | | Yes | | Yes | |
| $N$ | 3 | | 3 | | 3 | |
| $K$ | 321 | | 60 | | 13 | |

Note: Standard errors in parentheses.

(1): Firms per product-destination $\geq 2$

(2): Firms per product-destination $\geq 5$

(3): Firms per product-destination $\geq 10$

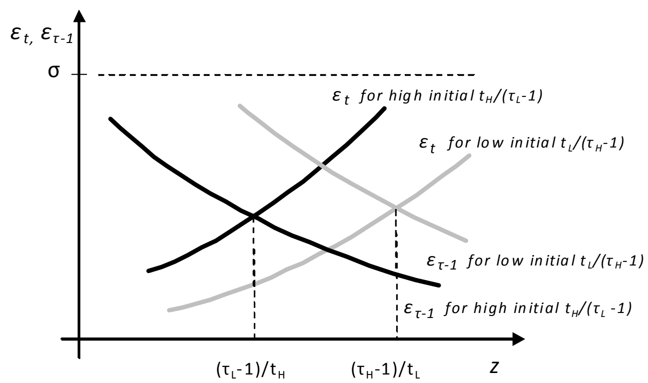


Figure 1: Elasticity of quantity sold to per-unit and ad-valorem trade cost as a function of $t$, $\tau - 1$ and $z$.

Table 3: Robustness: Alternative specifications

| | (R1) | (R2) | (R3) | (R4) | (R5) | (R6) |
|---|---|---|---|---|---|---|
| Trade costs, mean | .51 | .42 | .44 | .36 | .34 | .27 |
| | (.02) | (.01) | (.02) | (.01) | (.02) | (.02) |
| Trade costs, median | .42 | .41 | .44 | .33 | .34 | .17 |
| | (.03) | (.01) | (.02) | (.01) | (.02) | (.02) |
| Trade costs, stdev | .20 | .12 | .16 | .13 | .09 | .26 |
| $\gamma$ | 1.19 | 1.12 | 1.26 | 1.40 | .98 | 2.32[1] |
| | (.13) | (.04) | (.09) | (.04) | (.05) | (1.40[1]) |
| Criterion $f$ | 22.9 | 447.9 | 109.7 | 493.8 | 131.1 | 456.0 |
| $N$ markets | 8 | 16 | 6 | 24 | 9 | 118 |
| $K$ products | 6 | 149 | 40 | 116 | 19 | 19 |

Notes: Standard errors in parentheses. $\sigma$=6 used in all specifications except R6.

R1: Only single-product firms used in estimation.

R2: Only product-destinations with >29 and <41 firms used in estimation.

R3: Only products with quantities measured in units used in estimation.

R4: 2003 cross-section instead of 2004.

R5: Portuguese exports.

R6: Heterogeneity in $\sigma$ and $\gamma$. [1]: Average

Table 4: The costs of protectionism

| | | $\tau \uparrow$ | $t \uparrow$ |
|---|---|---|---|
| Welfare | $\{\sigma = 6, \gamma = \sigma - 1\}$ | $-9.1\%$ | $-11.6\%$ |
| | $\{\sigma = 4, \gamma = \sigma - 1\}$ | $-11.0\%$ | $-18.0\%$ |
| | $\{\sigma = 4, \gamma = \sigma + 1\}$ | $-9.2\%$ | $-13.3\%$ |
| TFP | $\{\sigma = 6, \gamma = \sigma - 1\}$ | $-1.6\%$ | $-4.4\%$ |
| | $\{\sigma = 4, \gamma = \sigma - 1\}$ | $-0.9\%$ | $-7.4\%$ |
| | $\{\sigma = 4, \gamma = \sigma + 1\}$ | $-3.2\%$ | $-7.7\%$ |

Note: Numbers represent the percentage change in welfare and TFP in case (B) and (C) compared to the baseline (A).

Table 5: Parameters used in simulation

| |
|---|
| First parameterization $\{\sigma = 6, \gamma = \sigma - 1\}$ |
| Second parameterization $\{\sigma = 4, \gamma = \sigma - 1\}$ |
| Third parameterization $\{\sigma = 4, \gamma = \sigma + 1\}$ |
| Other parameters common to all simulations: |
| $G^*/I = 0.42$, $Y = 1e + 5$ |
| # draws = # potential firms = $1e + 5$ |

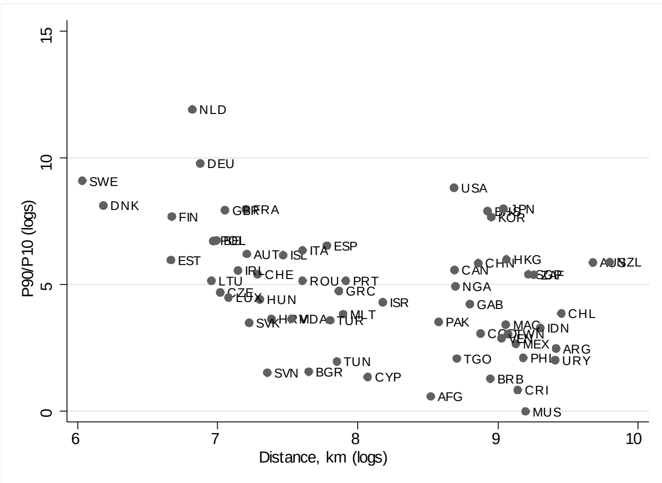


Figure 2: P90/P10 ratio of export quantity, weighted average across products.

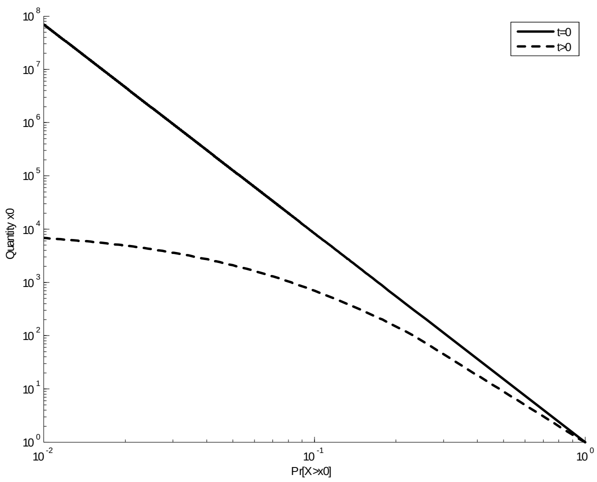


Figure 3: The export volume distribution.

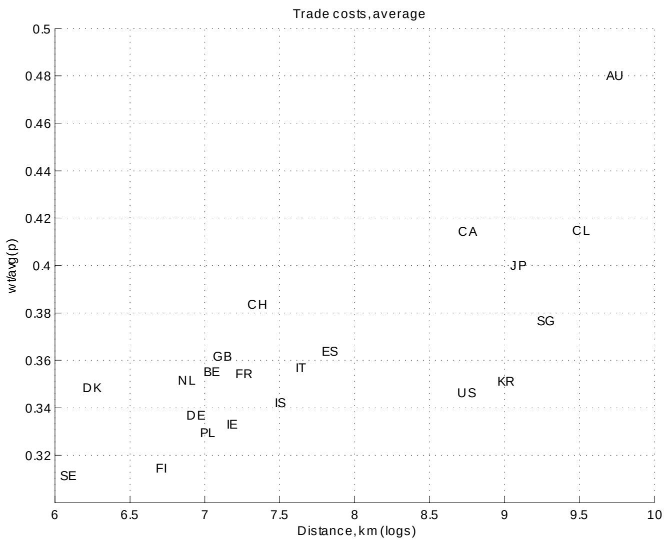


Figure 4: Per-unit trade costs relative to consumer price, averaged across products, conditional on $\sigma = 6$.

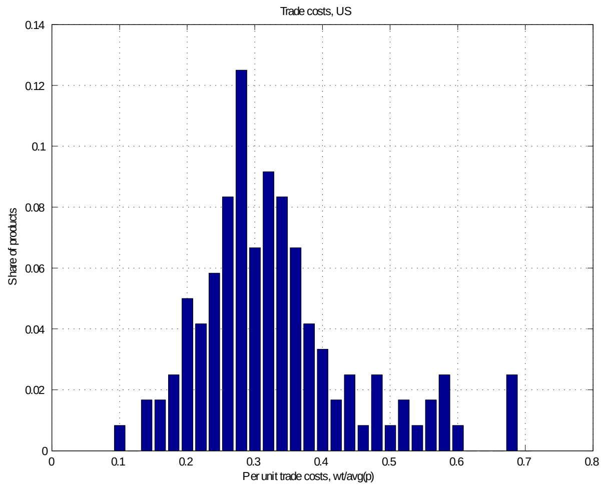


Figure 5: The density of trade costs, conditional on $\sigma = 6$. Norway to the U.S.

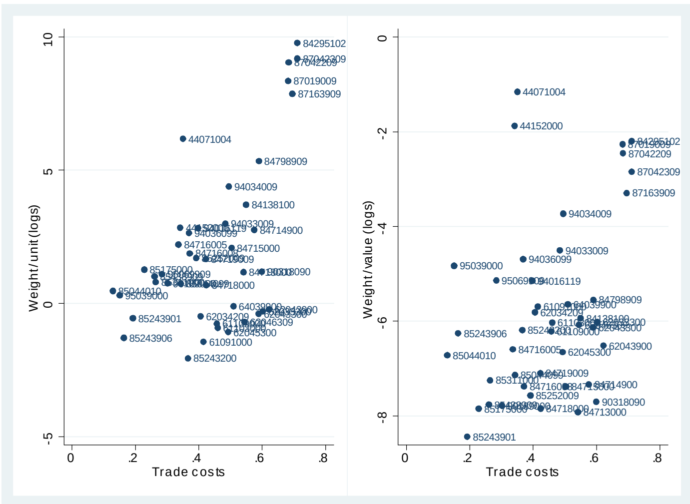


Figure 6: Relationship between estimated trade costs and actual weight/unit / weight/value.

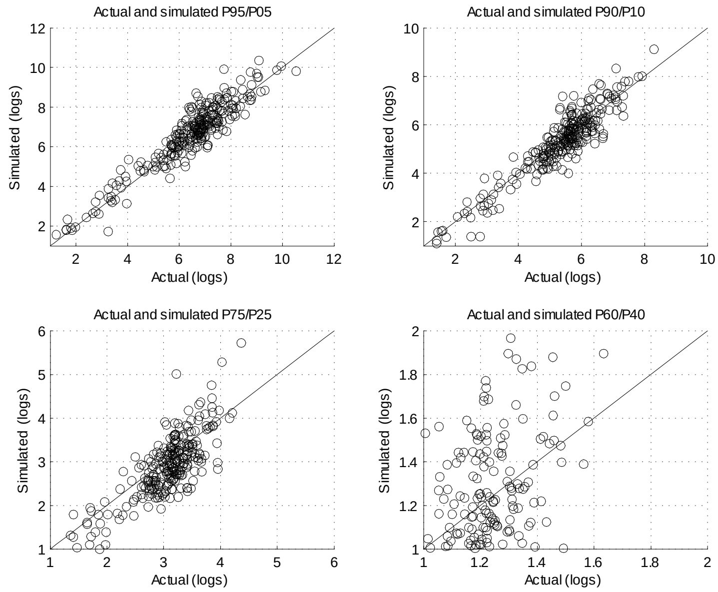


Figure 7: Model evaluation. Empirical and simulated percentile ratios.

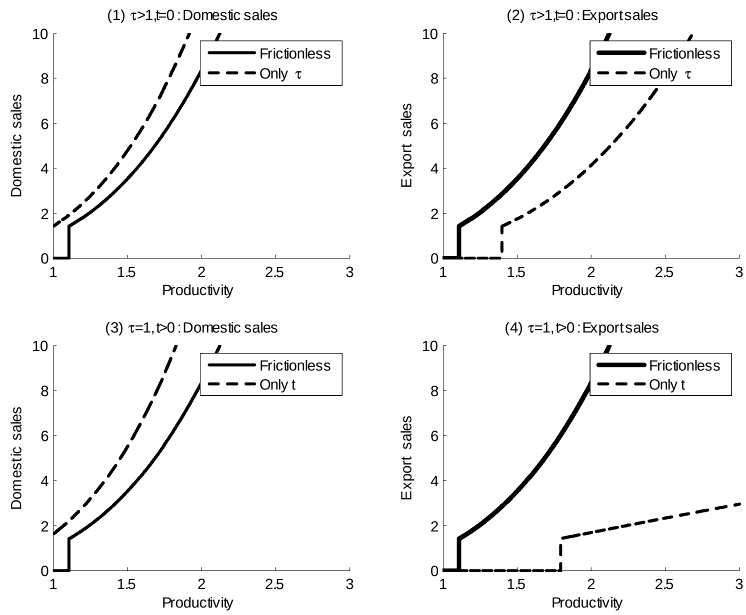


Figure 8: Simulation: The effect on domestic and export sales of protectionism.