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

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

July 2013

Abstract

Casual empiricism suggests that additive trade costs, such as quotas, per-unit tariffs, and, in part, transportation costs, are prevalent. In spite of this, we have no broad and systematic evidence of the magnitude of these costs. We develop a new empirical framework for estimating additive trade costs from standard firm-level trade data. Our results suggest that additive barriers are on average 14 percent, expressed relative to the median price. The point estimates are strongly correlated with common proxies for trade costs. Using our micro estimates, we show that a reduction in additive trade costs produces much higher welfare gains and growth in trade flows than a similar reduction in multiplicative trade costs.

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

*This is a substantially revised version of a paper previously circulated under the title "The Tip of the Iceberg: Modeling Trade Costs and Implications for Intra-Industry Reallocation". We would like to thank Costas Arkolakis, Gregory Corcos, Don Davis, Rob Johnson, Samuel Kortum, Ralph Ossa, Nina Pavcnik, Arvid Raknerud, Andrés Rodríguez-Clare, Alexandre Skiba, Karen Helene Ulltveit-Moe, and Kjetil Storesletten for their helpful suggestions, as well as seminar participants in various locations. We thank Statistics Norway for data preparation and clarifications and the project "European Firms in a Global Economy: Internal Policies for External Competitiveness" (EFIGE) for financial support. Alfonso Irarrazabal thanks the hospitality of the Chicago Booth School of Business where part of this research was conducted. Luca David Opromolla acknowledges financial support from national funds by FCT (Fundação para a Ciência e a Tecnologia). This article is part of the Strategic Project: PEst-OE/EGE/UI0436/2011. The analysis, opinions, and findings represent the views of the authors, they are not necessarily those of Banco de Portugal or Norges Bank.

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

[‡]Dartmouth College and NBER, andreas.moxnes@dartmouth.edu.

[§]Banco de Portugal, Research Department and Research Unit on Complexity and Economics (UECE), luca.opromolla@nyu.edu.

1 Introduction

The costs of international trade are the costs associated with the exchange of goods and services across borders. Trade costs impede international economic integration and may also explain a great number of empirical puzzles in international macroeconomics (Obstfeld and Rogoff, 2000). Since Samuelson (1954), economists usually model and estimate trade costs as multiplicative (iceberg) costs, implying that pricier goods are costlier to trade. Yet, casual empiricism suggests that additive trade costs are prevalent.[1] First, the pricing structure in shipping is often a fixed charge per unit (e.g. per pound or cubic meter), and a percentage charge for insurance. For example, according to UPS rates at the time of writing, a fee of $125 is charged for shipping a two pound package from Oslo to New York (UPS Worldwide Saver), while they charge an additional 0.85% of the declared value for full insurance.[2] Second, a number of trade policy instruments also act like additive trade costs. 19 percent of U.S. non-agricultural imports are subject to additive tariffs.[3] Quotas (through the imposition of a quota license price) also act like an additive tariff. In the U.S. and the European Union, 9.5 and 15.1 percent of the Harmonized System (HS) six-digit subheadings in the schedule of agricultural concessions are covered by tariff quotas. Third, distribution costs are also partly additive costs (e.g. Corsetti and Dedola, 2005).

Even though we can directly observe the magnitude of additive trade costs in some specific cases, e.g. for a freight company or in a country's tariff schedule, we have no broad and systematic evidence of the magnitude of additive trade costs in international trade. The first contribution of this paper is to fill this gap in the literature. We present a general framework to structurally estimate the magnitude of additive trade costs, using, now standard, firm-level trade data. Our methodology exploits a robust theoretical mechanism that shapes the association between producer prices and demand in the presence of additive costs. Specifically, as additive trade costs increase, the demand elasticity in a market becomes less negative and especially so among low price firms. This result holds for a wide range of utility functions. Our identification strategy resembles a triple differences approach: we compare the change in elasticities across

[1] Multiplicative costs are defined as a constant percentage of the producer price per unit traded, while additive costs are defined as a constant cost per unit traded (conditional on a product). We use the terminology additive costs throughout the paper. Per-unit or specific trade costs are also terms frequently used in the literature.

[2] The fee per pound varies according to origin and destination, while the insurance charge is independent of origin and destination.

[3] 2006 data from the WTO are presented in Table 8. We discuss the data in more detail in the appendix. Until the 1950's, two-thirds of dutiable U.S. imports were subject to additive tariffs. This proportion fell to less than 40 percent by the early 1970's (Irwin, 1998).

high and low price firms, as we move to different export destinations with different degrees of trade costs. Our methodology is robust to endogeneity concerns, as well as quality heterogeneity and non-constant demand elasticities within narrowly defined products (Section 4.3).

We apply our methodology to Norwegian firm-level trade data, and estimate trade costs for a large number of countries and products. Several strong results emerge from the empirical analysis. The unweighted mean of additive trade costs, expressed relative to the median price, is 14 percent. Our estimates are strongly positively correlated with observable proxies of trade costs, such as distance and product weight per value. We emphasize that our methodology can only identify additive trade costs *relative to* multiplicative costs, meaning that our estimates are a lower bound of the true value of additive costs.[4] Multiplicative costs are not separately identified by our framework, and are, as such, largely left unexplored.

The second contribution of this paper is to show that the presence of additive trade costs has important implications. First, gains from trade are potentially much larger than in standard models (e.g. Arkolakis, Costinot, and Rodríguez-Clare, 2012a). Using our micro estimates, we show that a reduction in additive trade costs would imply higher welfare gains than a similar reduction in multiplicative trade costs. The intuition is simple. If all firms within an industry charge the same price, then the difference between additive and multiplicative trade costs is trivial. *In the presence of heterogeneity in prices*, however, an additive trade barrier distorts the relative price of two varieties both within markets and across markets. As a consequence, and as shown by Alchian and Allen (1964), additive costs alter relative consumption patterns both within and across markets. Multiplicative barriers, on the other hand, only distort prices across markets.[5] This additional margin of distortion is the reason why welfare may be more adversely affected by additive than multiplicative barriers. Empirical findings of the gains from trade (e.g. Pavcnik, 2002 and Feyrer, 2009) are often large compared to the relatively modest gains predicted by the class of models considered by Arkolakis, Costinot, and Rodríguez-Clare (2012a). Hence, this paper contributes to reconciling the empirical and theoretical evidence. More generally, in our model, heterogeneity among producers has aggregate implications, in contrast to Arkolakis, Costinot, and Rodríguez-Clare (2012a).

[4]We estimate $\tilde{t} = t/\tau$, where $t \geq 0$ represents additive costs and $\tau \geq 1$ represents multiplicative (iceberg) costs. Hence, $t = \tilde{t}\tau \geq \tilde{t}$.

[5]E.g. the ratio of consumer prices for two varieties exported to the same market is $(\tau\tilde{p}(\omega_1) + t) / (\tau\tilde{p}(\omega_2) + t)$, where τ is the multiplicative barrier, t is the additive barrier, and $\tilde{p}(\omega_r)$ is the producer price of variety ω_r. The ratio equals relative producer prices $\tilde{p}(\omega_1)/\tilde{p}(\omega_2)$ if $t = 0$.

Second, again using our micro estimates, we show that removing additive barriers produces much more trade than reducing multiplicative barriers. This result suggests that inferring (iceberg) trade costs from trade flows using gravity models, as in Anderson and van Wincoop (2004), may overstate trade barriers, as additive costs dampen trade much more than multiplicative costs. It is well known that standard models have difficulties matching the growth in global trade over time (e.g., Yi, 2003); our results suggest that additive trade costs may play an important role.

Third, standard trade models (e.g. Melitz, 2003) can predict zero bilateral trade flows between any country pair only once fixed costs on the supply side are assumed in combination with a particular productivity cut-off, as in Helpman, Melitz, and Rubinstein (2008), or after assuming a finite integer number of firms, as in Eaton, Kortum, and Sotelo (2012). Our model offers a very natural way to reconcile empirics and theory, without assuming arbitrary productivity cut-offs or abandoning the continuum assumption. Since the presence of additive trade costs means that firm sales are always bounded, then even the most productive firm may not find it profitable to enter the export market.

In summary, we conclude that empirical and theoretical work should account for both (the tip of the) iceberg costs, as well as the part of trade costs that are largely hidden under the surface: additive costs.

More flexible modeling of trade costs is not new in international economics. Alchian and Allen (1964) pointed out that additive costs imply that the relative price of two varieties of some good will depend on the level of trade costs, and that relative demand for the high quality good increases with trade costs ("shipping the good apples out"). More recently, Hummels and Skiba (2004) found strong empirical support for the Alchian-Allen hypothesis. Specifically, the elasticity of freight rates with respect to price was estimated to be well below the unitary elasticity implied by the iceberg assumption. However, the authors could not identify the magnitude of additive costs, as we do here. Furthermore, our methodology identifies all kinds of trade costs, whereas their paper was concerned with shipping costs exclusively.

Khandelwal, Schott, and Wei (2011) examine Chinese exports and productivity growth before and after the elimination of externally imposed export quotas. Building on the model in this paper, they reinterpret the additive tariff as a (common) quota license fee which firms must pay in order to access restricted foreign markets. Martin (2012) shows that the presence of additive trade costs is necessary to reconcile the most commonly used theoretical framework with the empirical finding that individual firms set higher f.o.b. prices over long distances than over short ones, a sort of "reverse dumping". Sørensen (2012) investigates the welfare impact of additive versus multiplicative barriers in a symmetric two-country setting.

Our work also relates to a recent paper by Berman, Martin, and Mayer (2012). They introduce a model with heterogeneous firms and additive costs, but in their model the additive component is interpreted as local distribution costs that are independent of firm productivity. Their research question is very different, however, as their paper analyzes pricing to market and the reaction of exporters to exchange rate changes. They show that, in response to currency depreciation, high productivity firms optimally raise their markup rather than the volume, while low productivity firms choose the opposite strategy.

Finally, our work connects to the papers that quantify trade costs. Anderson and van Wincoop (2004) provides an overview of the literature, and recent contributions \include Anderson and van Wincoop (2003), Eaton and Kortum (2002), Head and Ries (2001), Helpman, Melitz, and Rubinstein (2008), 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 the within-market relationship between producer prices and exports is conceptually different and provides a complimentary approach to inferring trade barriers from data. This is possible thanks to the recent availability of detailed firm-level data. Furthermore, whereas the traditional approach can only identify iceberg trade costs relative to some benchmark, usually domestic trade costs, our method identifies the absolute level of (additive) trade costs.

The rest of the paper is organized as follows. Section 2 presents the general framework and summarizes its implications. Since the subsequent empirical framework is formulated conditional on a set of general equilibrium variables, we do not specify a full model here, but only the elements that are relevant to the empirical work. In Section 3, we describe the data and present some empirical patterns that are suggestive of the presence of additive trade costs. Section 4 lays out the econometric strategy, presents the baseline estimates, as well as validation exercises and robustness checks. In Section 5, we present a full general equilibrium model. Section 6 compares the welfare and trade flows impact of additive versus ad valorem tariffs. Finally, Section 7 concludes.

2 Basic Framework

In this section, we present a simple framework that features both iceberg and additive trade costs. This framework relies on a small number of assumptions that are sufficient to implement the empirical analysis of Section 4. We use the framework to derive two propositions (Section 2.3). The first proposition relates quantity demanded to the producer price and provides the backbone of the identification strategy of Section 4, i.e. it shows why we are able to estimate additive trade costs. The second proposition shows why additive trade costs can be more detrimental to welfare and trade flows than multiplicative costs. Later, we present the full model, solve for the general equilibrium, and quantify the impact of additive trade costs on welfare and trade flows.

2.1 Consumer Demand

We consider a world economy comprising N countries. Each country n is populated by a measure L_n of workers. The economy consists of a differentiated goods sector and a transport services sector. We describe the latter in detail in Section 5.1 and focus here on the differentiated goods sector.

Preferences across varieties of the differentiated product have the standard CES form with an elasticity of substitution $\sigma > 1$. Each variety enters the utility function with its own exogenous country-specific weight η_n. These weights represent firm- and destination-specific demand shocks. These preferences generate a demand function $A_n \left(p_n/\eta_n\right)^{1-\sigma}$ in country n for a variety with price p_n and demand shock η_n. The demand level $A_n \equiv Y_n P_n^{\sigma-1}$ depends on total expenditure Y_n and the consumption-based price index P_n.[6]

2.2 Variable Trade Costs

As in e.g. Hummels and Skiba (2004), the consumer price of a good depends on its producer price, $\tilde{p}$, as well as on additive and multiplicative trade costs, $t \geq 0$ and $\tau \geq 1$ respectively,

$$p = \tau\tilde{p} + t. \tag{1}$$

Hence, total trade costs are partly proportional to the quantity shipped and partly proportional to the producer price. In the estimation of Section 4, we always condition on the observed producer (f.o.b.) price $\tilde{p}$ so that, at this stage, there is no need to make any assumption about market structure.

[6] From now until Section 5, where we introduce the rest of the model, we drop the country subindex.

2.3 Implications of Additive Trade Costs

We present two propositions that summarize a number of important properties of the theoretical framework. The first proposition characterizes the relationship between quantity demanded and producer price in the presence of additive trade costs, thereby providing the backbone of the identification strategy to quantify additive trade costs. The second proposition shows that additive trade costs act as a wedge both between the consumer prices of local and imported goods *and* between the consumer prices of different imported goods. Based on this result, we show in Section 6 that the gains in terms of welfare and trade flows from eliminating additive trade costs are much higher than those from eliminating multiplicative trade costs.[7]

2.3.1 Trade Costs and Demand Elasticities

The identification strategy we employ in Section 4 focuses on the elasticity of quantity demanded with respect to the producer price. The following proposition characterizes such an elasticity both in the absence and in the presence of additive trade costs.

Proposition 1. *When $t = 0$, the elasticity of quantity demanded with respect to the producer price (E), is equal to $-\sigma$. When $t > 0$, $E = -\sigma / \left[1 + t / \left(\tau \tilde{p}\right)\right]$, and the elasticity of E with respect to additive trade costs t is (i) negative and (ii) strictly increasing in the producer price $\tilde{p}$.*

The proposition shows that, in the presence of additive trade costs, the elasticity of quantity demanded with respect to the producer price is a function of the additive trade costs relative to the producer price and augmented by the iceberg cost. Moreover, Proposition 1 also shows that a higher additive trade cost "pushes" the elasticity E towards zero (part (i)), and particularly so when the producer price is low (part (ii)).[8] As we will explain in more detail in Section 4.2, variation in the elasticity of quantity demanded with respect to the producer price across product-market pairs can be exploited to identify $t / \left(\tau \tilde{p}\right)$.

We provide the proof of Proposition 1 in the Appendix and some intuition here. Using the chain rule, E is the product of the elasticity of quantity demanded with respect to the consumer price and the elasticity of the consumer price with respect to the producer price. In the (widely used) case of CES preferences, the first elasticity

[7] A reader that is not interested in the estimation of additive trade costs but is interested in the implications of such costs for welfare and for aggregate trade flows can read Proposition 2 and then go directly to Section 5.

[8] Recall that $E < 0$ so that $\left(\partial E / \partial t\right) / \left(t / E\right)$ carries the opposite sign of $\partial E / \partial t$.

does not depend on the level of the producer price and equals $-\sigma$, the (opposite of the) elasticity of substitution between any two varieties. Therefore,

$$E(\tilde{p}, t, \tau, \sigma) \quad = \quad -\sigma \frac{\partial \ln p}{\partial \ln \widetilde{p}}. \tag{2}$$

In the absence of additive trade costs, the second elasticity is equal to one (using equation (1)), so E is equal to $-\sigma$. In the presence of additive trade costs, the second elasticity, $\partial \ln p / \partial \ln \widetilde{p}$, is a function of additive trade costs relative to the producer price and augmented by the iceberg cost,

$$\frac{\partial \ln p}{\partial \ln \widetilde{p}} = \left(1 + \frac{t}{\tau \widetilde{p}}\right)^{-1} \geq 0. \tag{3}$$

The intution is as follows. Additive trade costs act as a wedge between the consumer price and the producer price. The size of the wedge depends on the magnitude of t relative to $\tau\tilde{p}$. The larger the wedge the weaker the relationship between the two prices (i.e. the lower $\partial \ln p / \partial \ln \widetilde{p}$) and the higher E.[9] The elasticity of quantity demanded to the producer price is an increasing function of additive trade costs. The last part of Proposition 1 states that the "dampening" effect of additive trade costs on the elasticity between quantity demanded and producer price is stronger for low producer prices. The intuition is straightforward: consider, for example, the limiting case in which the producer price is so high that additive trade costs becomes negligible: changes in the additive trade cost have zero impact on E. Hence, an increase in t always increases E but the more so the lower is the producer price. In the appendix, we discuss Proposition 1 under different demand systems, and show that the main content of the proposition continues to hold in a large class of demand systems.

We also refer the reader to Section 4.2, where we discuss identification more in detail and also provide a graphical representation of the content of Proposition 1.

2.3.2 Additive Trade Costs as Wedges Between Prices

The next proposition shows that additive trade costs impose more wedges between prices than multiplicative trade costs. As a consequence, a reduction in additive trade costs is associated with larger gains from trade (Section 6). Moreover, Proposition 2 shows that the source of additional gains from trade is the interaction between additive trade costs and producer price heterogeneity.

[9] Again, recall that $E < 0$.

Consider two varieties. One has producer price $\tilde{p}'$ and consumer price p'. The other variety is more expensive, with producer price $\tilde{p} = \nu\tilde{p}'$ ($\nu > 1$), and consumer price p. Both varieties are exported to the same market and are subject to the same additive trade cost t and multiplicative trade cost τ according to equation (1). Let $\chi = (p/p') \,/\, (\tilde{p}/\tilde{p}')$ be the wedge between the relative consumer (p/p') and producer price $(\tilde{p}/\tilde{p}')$ of the high-producer price variety.

Proposition 2. *When $t = 0$, $\chi = 1$. When $t > 0$, (i) $\chi < 1$; (ii) $\partial\chi/\partial\,(t/\tau) < 0$; and $\partial\left[\partial\chi/\partial\,(t/\tau)\right]/\partial\nu < 0$.*

Proposition 2 shows that, in the absence of additive trade costs, the relative consumer price of the two imported varieties equals their relative producer price. In other words, multiplicative trade costs do not affect the relative consumer price in the importing country. That's not the case when there are additive trade costs. In the presence of additive trade costs, the relative consumer price of the high-cost variety in the importing country is lower than the relative producer price,

$$\chi = \frac{\frac{p}{p'}}{\frac{\tilde{p}}{\tilde{p}'}} = \frac{1 + \frac{t}{\nu\tau\tilde{p}'}}{1 + \frac{t}{\tau\tilde{p}'}} < 1.$$

The second part of Proposition 2 shows how the wedge between the relative consumer and producer price depends on the magnitude of the additive trade cost. Holding producer prices constant,

$$\frac{\partial\chi}{\partial\,(t/\tau)} = \frac{1-\nu}{\nu\tilde{p}'\left(1 + \frac{t}{\tau\tilde{p}'}\right)^2} < 0,$$

since $\nu > 1$. Since the additive trade cost is the same for both varieties, an increase in t reduces the relative consumer price of the high price variety. Note that, in the absence of additive trade costs, an increase in multiplicative trade costs does not affect χ, which remains equal to one. The additional price wedge associated with additive trade costs is the reason why gains from trade may be quite different in a model with additive trade costs compared to a model with multiplicative trade costs (Section 5).

The last part of Proposition 2 shows that the impact of additive trade costs on χ is stronger the higher the degree of heterogeneity in producer prices, i.e. the higher is ν,

$$\frac{\partial}{\partial\nu}\left(\frac{\partial\chi}{\partial\,(t/\tau)}\right) = -\frac{1}{\nu^2\tilde{p}'\left(1 + \frac{t}{\tau\tilde{p}'}\right)^2} < 0.$$

We explore this intuition more formally in Section 6. Before turning to some suggestive evidence of the presence of additive trade costs, we make two final remarks. First, Proposition 2 is entriely independent from our assumption of CES preferences, and just relies on equation (1). Second, under some regularity conditions about demand (see e.g. Hummels and Skiba, 2004), an increase in t raises relative consumption of the high price variety relative to the low price variety. This is the well-known Alchian-Allen effect (Alchian and Allen, 1964).

3 Empirical Regularities

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

3.1 Data

The data cover all Norwegian non-oil exporters in 2004 and originate from customs declarations. Every export observation is associated with a firm r, a destination n and a product k, the quantity transacted x_{knr} and the total value.[10] We calculate f.o.b. prices $\widetilde{p}_{knr}$ by dividing total value by quantity. We define a product as a Harmonized System 8-digit (HS8) nomenclature category. The sample covers $17,480$ firms, exporting $5,391$ products to 203 destinations.

In 2004, total exports amount to NOK 232 billion ($\approx$ USD 34.4 billion), or 48 percent of the aggregate manufacturing revenue. On average, each firm exported 5.6 products to 3.4 destinations for NOK 13.3 million ($\approx$ USD 2.0 million). On average, there are 3.0 firms per product-destination pair, with a standard deviation of 7.8. As we will see in Section 4, our quantitative framework utilizes the relationship between f.o.b. price and export quantity *across firms within a product-destination pair*. In the formal econometric model, we therefore choose to restrict the sample to product-destinations where more than 40 firms are present.[11] In the robustness section, we evaluate the effect of this restriction by estimating the model on an expanded set of destination-product pairs. Extreme values of quantity sold, defined as values below the 1^{st} percentile or above the 99^{th} percentile for every product-destination are dropped

[10] The unit of measurement depends on the characteristics of the product. E.g. gloves and skis are measured in pairs, while mineral water is measured in liters. Firm-product-destination-year observations are recorded in the data as long as the f.o.b. exports value is NOK 1000 ($\approx$ USD 148) or higher.

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

from the sample. All in all, this brings down the total number of products to 121 and the number of destinations to 21. Exports to all possible combinations of these products and destinations amount to 26.2% of total exports. In the robustness section we consider an alternative sample that covers about 58.9% of total exports.

Several features of the Norwegian data are consistent with those from other countries. For example, Irarrazabal, Moxnes, and Opromolla (2013) report that firm-level facts for Norwegian exporters are consistent with those for French exporters shown in Eaton, Kortum, and Kramarz, 2011.

3.2 Suggestive Evidence

Proposition 1 shows that the elasticity of E with respect to additive trade costs is negative and increasing in the producer price. In the following exercise, we provide evidence that these implications hold in the data, using distance between Norway and the destination country as a proxy for additive trade costs. We regress export volume (x_{knr}) on a full set of interactions between f.o.b. price ($\widetilde{p}_{knr}$), distance ($Dist_n$) and a dummy equal to one if the price is above the product-destination median price, $M_{knr} \equiv 1\left[\widetilde{p}_{knr} > median_r\left(\widetilde{p}_{knr}\right)\right]$,

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

where $\times\times$ denotes the full set of interactions and β is the vector of coefficients. We also include product-destination fixed effects (α_{kn}) to exploit variation across firms within a product-destination cell, as suggested by the theory. The relationship between quantity exported and f.o.b. price is

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

which is allowed to vary between low- and high-price firms (β_3), and is allowed to depend on distance from Norway, with the slope being different for low-price (β_2) and high-price firms ($\beta_2 + \beta_4$). We expect to find a positive coefficient for β_2, showing that distance increases the negative elasticity of demand (i.e. that the elasticity approaches zero). We also expect to find a negative coefficient for β_4, showing that the "dampening" effect of additive trade costs is smaller for high-price firms.

Since the error ε_{knr} is presumably correlated with prices, the estimated coefficients will not reflect the true demand elasticity. In the formal econometric model in Section 4 we show that identification of additive trade costs does not rely on iden-

Table 1: Exports volumes and f.o.b. prices.

	OLS		IV	
$\ln \widetilde{p}_{knr} \times \ln Dist_n \times M_{knr}$ (β_4)	-0.04^a	(0.01)	-0.13^a	(0.04)
$\ln \widetilde{p}_{knr}$ (β_1)	-1.15^a	(0.08)	-2.98^a	(0.64)
$\ln \widetilde{p}_{knr} \times \ln Dist_n$ (β_2)	0.04^a	(0.01)	0.29^a	(0.09)
M_{knr}	-2.52^a	(0.62)	-3.88^b	(1.89)
$\ln \widetilde{p}_{knr} \times M_{knr}$ (β_3)	0.40^a	(0.10)	1.13^a	(0.29)
$\ln Dist_n \times M_{knr}$	0.22^b	(0.09)	0.28	(0.28)
Product-destination FE	Yes		Yes	
IV	No		Yes	
R^2	0.59		0.35	
Number of observations	66,403		33,445	

Notes: The dependent variable is log exports volume. Only product-destinations with more than 10 firms are included in the sample. Standard errors in parentheses, clustered by product-destination in column 1. Significance levels: a 1%; b 5%.

tifying the true demand elasticity. Nevertheless, we also create an instrument for prices. In the spirit of Hausman (1996), we use the average price in other markets, $z_{knr} = [1/(N_{kr} - 1)] \sum_{m \in S^-_{knr}} p_{kmr}$, where N_{kr} is the number of export destinations by firm-product kr and S^-_{knr} is the set of destinations served by firm-product kr, except destination n. We expect the price in other markets to be strongly correlated with the price charged in destination n, while not directly affecting demand in market n. In other words, we assume that prices are correlated due to cost shocks and not demand shocks. Since the interaction terms are presumably also endogenous, we construct instruments for them as well (e.g $\ln \widetilde{p}_{knr} \times \ln Dist_n$ is instrumented by $\ln z_{knr} \times \ln Dist_n$).

Results from the estimations are presented in Table 1. Column (1) is estimated with ordinary least squares, while in column (2) we instrument prices with prices in other markets.[12] Our results strongly support the theoretical implications of Proposition 1: the elasticity of E with respect to additive trade costs (proxied by distance) is negative and increasing in the producer price. Specifically,

$$\frac{\partial E_{knr}}{\partial Dist_n} \frac{Dist_n}{E_{knr}} = \begin{cases} \frac{1}{E_{knr}\big|_{M_{knr}=0}} \beta_2 < 0 \text{ if } M_{knr} = 0 \\ \frac{1}{E_{knr}\big|_{M_{knr}=1}} (\beta_2 + \beta_4) < 0 \text{ if } M_{knr} = 1 \end{cases}.$$

The magnitudes of the previous expressions can be easily computed, and their signs can be checked, using the coefficients in Table 1 and knowing that average (log) distance between Norway and foreign destinations is about 7.6 (i.e. about 2,000 km). Using OLS

[12] The first stage regressions are overall strong, with overall R-square values of 0.25, 0.71, 0.74, 0.08, 0.24 and 0.08 for the 6 first stage regressions.

estimates (Table 1, column 1), the elasticity of E with respect to additive trade costs is about -0.04 for low-price firms, and zero for high-price firms. Using IV estimates (Table 1, column 2), the corresponding values are even larger in absolute value, being -0.37 for low-price firms and -0.25 for high-price firms.

4 Estimating Trade Costs

In this section we structurally estimate the magnitude of trade costs, for every destination and every product in our sample. The intuition underlying our identification strategy is the following. If we had data on both producer and consumer prices for each export transaction, additive trade costs could be inferred by regressing consumer prices on producer prices (equation (1)). Unfortunately, that is not feasible since our data (as most of the transaction-level data sets) only include producer (f.o.b.) prices and quantities. However, since quantity demanded is a function of consumer prices, we can infer consumer prices from demand, and indirectly compare producer and consumer prices. Proposition 1 shows that the demand elasticity E contains information on additive trade costs. Moreover, Proposition 1 implies that the decrease in the absolute value of the demand elasticity E in response to an increase in additive trade costs is larger for low-price compared to high-price firms. It is this mechanism that provides identification and that allows us to recover estimates of trade costs consistent with our model. The methodology is reminiscent of a triple difference approach, where trade costs are identified by comparing the difference in the elasticity of the volume of exports to f.o.b. prices between low- and high-price firms, for a particular product, across destinations.[13]

The econometric strategy consists of finding expected exports volume conditional on the producer price charged by minimizing the sum of squared residuals using nonlinear least squares.[14] This strategy has, at least, three merits. First, we do not need to simulate a full general equilibrium to estimate trade costs. Second, we do not need assumptions about market structure or about the firm productivity distribution, as we

[13] In a previous version of this paper (Irarrazabal, Moxnes, and Opromolla, 2010), we identified additive trade costs from the exports volume distribution alone, without using information on prices. Even though the identifying assumption was very different from that used in this version of the paper, the previous methodology produced remarkably similar results.

[14] We choose to use data for exports volume (quantities) instead of export sales for the following reasons. First, using quantities instead of sales minimizes measurement error due to imperfect imputation of transport/insurance costs. Second, we avoid transfer pricing issues when trade is intra-firm (Bernard, Jensen and Schott 2006). Third, we do not get closed form expressions for the estimation equation when using sales value.

condition on observed f.o.b. prices. Third, since firm-level trade data by product and destination are now widely available, our methodology can be applied to a range of different countries and time periods.

4.1 Estimation

We employ a simple nonlinear least squares estimator where the objective is to minimize the squared difference between expected and actual log exports volume. The starting point is the quantity demanded equation $x_n = A_n p_n^{-\sigma} \eta_n^{\sigma-1}$ of Section 2.1. First, since we have data on quantities and prices for each firm-product-destination triplet, we add subscripts k (product) and r (firm) to the consumer price p and quantity x. Second, we allow the intercept term A to be product-destination-specific and the elasticity of substitution σ to be product-specific. We also allow the firm-product-destination-specific demand shock $\ln \eta_{knr}$ to be correlated with the corresponding consumer price $\ln p_{knr}$ (see Section 4.3). All in all, these changes imply the following equation,

$$\ln x_{knr} = a_{kn} - \sigma_k \ln p_{knr} + (\sigma_k - 1) \ln \eta_{knr}, \tag{4}$$

where the demand shifter $a_{kn} = \ln A_{kn} = \ln Y_{kn} + (\sigma_k - 1) \ln P_{kn}$ captures total expenditure and the price index of product k in market n.

The consumer price p_{knr} is unobserved, but the f.o.b. price $\widetilde{p}_{knr}$ is observable in our data. We substitute p_{knr} with $\widetilde{p}_{knr}$ using $p_{knr} = \tau_{kn} \widetilde{p}_{knr} + t_{kn}$. The resulting estimating equation is

$$\ln x_{knr} = \tilde{a}_{kn} - \sigma_k \ln \left(\tilde{p}_{knr} + \tilde{t}_{kn} \right) + \epsilon_{knr}, \tag{5}$$

where $\widetilde{t}_{kn} \equiv t_{kn}/\tau_{kn}$ is our coefficient of interest, $\epsilon \equiv (\sigma_k - 1) \ln \eta_{knr}$ and the intercept term $\tilde{a}_{kn} \equiv a_{kn} - \sigma_k \ln \tau_{kn}$.

Finally, we decompose $\widetilde{t}_{kn}$ into product- and destination-specific fixed effects, $\widetilde{t}_{kn} = \tilde{t}_k \tilde{t}_n$, and normalize $\tilde{t}_k = 1$ for $k = 1$. The normalization is similar to the one adopted in the estimation of two-way fixed effects in the employer-employee literature (Abowd, Creecy, and Kramarz, 2002). Even though $\tilde{t}_k$ is estimated relative to some normalization, the estimate of $\widetilde{t}_{kn}$ is invariant to the choice of normalization.[15] This decomposition enables us to identify trade costs that are due to product and market characteristics separately. We also decompose $\tilde{a}_{kn} = \tilde{a}_k \tilde{a}_n$, and normalize $\tilde{a}_k = 1$ for $k = 1$.

[15] We also need to ensure that all products and destinations belong to the same mobility group. The intuition is that if a market is only served by one product, then one cannot separate the product from the destination effect. In the robustness section we check whether our estimates are sensitive to the trade cost decomposition $\widetilde{t}_{kn} = t_k \widetilde{t}_n$ by estimating $\widetilde{t}_{kn}$ directly for all possible product-destination pairs.

This restriction helps us keeping down the number of coefficients to estimate. Finally, we minimize, with respect to the coefficient vector $\mathbf{\Psi} = \left(\tilde{t}_k, \tilde{t}_n, \sigma_k, \tilde{a}_k, \tilde{a}_n\right)_{k \in K, n \in N_k}$, the sum of squared residuals

$$O\left(\mathbf{\Psi}\right) = \sum_{k \in K} \sum_{n \in N_k} \sum_{r \in R_{kn}} \widehat{\epsilon}_{knr}^2,$$

where K is the set of products in the sample, N_k is the set of active destinations for product k, and R_{kn} is the set of firms exporting product k to destination n.[16]

4.2 Identification of trade costs

We discuss identification in the context of an example. Consider two products, feather (F) and stone (S) exported from Norway to Sweden (SE) and Japan (JP). Suppose the additive trade cost is larger for stone (than for feather) and for Japan (than for Sweden). That would likely be the case for transportation costs since stone is heavier than feather and Japan is more distant than Sweden from Norway. Figure 1 shows f.o.b. prices on the horizontal axis and quantity demanded for a simple numerical example ($\sigma_k = 1$, $\tilde{t}_{JP}/\tilde{t}_{SE} = 10$, $\tilde{t}_S/\tilde{t}_F = 5$). Additive trade costs are minimal in the case of feather shipped to Sweden: Figure 1 shows that the quantity demanded function in this case is almost linear, consistent with Proposition 1. As we move from Sweden to Japan, the quantity demanded function becomes more concave for low f.o.b. prices while it does not change much for high f.o.b. prices. This is, again, consistent with Proposition 1: the increase in the elasticity of quantity demanded with respect to the f.o.b. price, E, in response to an increase in additive trade costs (from Sweden to Japan), is larger for low-price compared to high-price firms.

The trade cost product and destination fixed effects, $\tilde{t}_k$ and $\tilde{t}_n$, are identified by comparing differences in the slopes of the quantity demanded function for low-price versus high-price firms across products for a given destination or across destinations for a given product. The methodology is therefore reminiscent of a triple difference approach, where trade costs are identified from the change in the difference in elasticities between low- and high price firms, as we compare different markets.

A potential concern is that the slope coefficients σ_k are not separately identified from the trade cost coefficients $\tilde{t}_{kn}$, since they are all identified from the slope and curvature of the demand function. This is where our use of both the product and destination

[16] In practice, we minimize $O(.)$ under a set of lower and upper bounds and linear inequalities, since this speeds up the search for the global minimum. The lower and upper bounds are $[-20, 20]$ for $\left(\ln \tilde{t}_k, \ln \tilde{t}_n, \sigma_k, \tilde{a}_k, \tilde{a}_n\right)$ $\forall k \in K, \forall n \in N_k$, while the linear inequalities are $\ln \tilde{t}_k + \ln \tilde{t}_n < 2 \ln \overline{\overline{p}}_{kn}$ $\forall k \in K, \forall n \in N_k$, where $\overline{\overline{p}}_{kn}$ is the median producer price in product-destination kn.

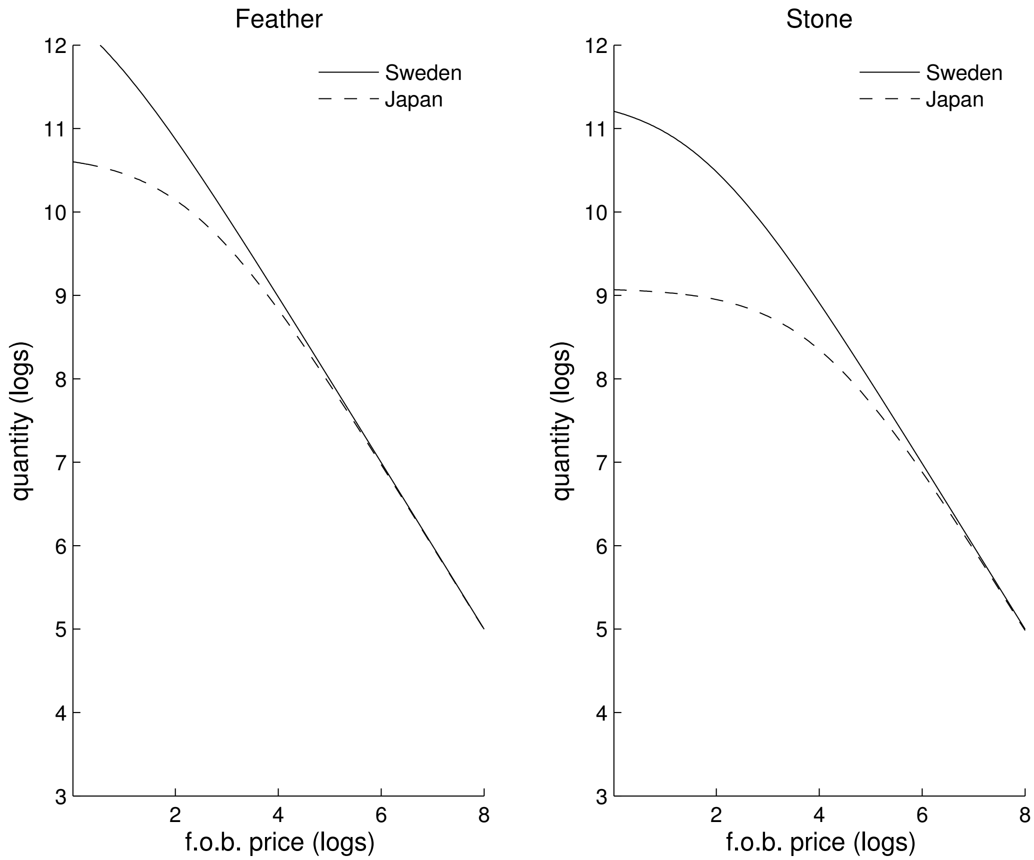

Figure 1: Identification.

dimensions of the data, as well as the corresponding two-way fixed effects, becomes important. Suppose we are interested in finding the trade cost parameter to Japan, $\tilde{t}_{JP}$. Given information that trade costs to Sweden are low, the demand coefficients σ_F and σ_S are identified, as trade costs have a negligible impact on curvature in Sweden. Given σ_F and σ_S, we can back out trade costs to Japan, $\tilde{t}_{JP}$, by examining the deviation from these slopes among low-price firms.

A different concern is that our model assumes that the demand elasticity with respect to the consumer price is constant, while this may not be true in the data. In terms of Figure 1, this means that the demand schedule for Sweden and feather might not be linear, even in the absence of additive trade costs. We discuss this case in the appendix, and show that Proposition 1 and the identification strategy would still hold in this case.

Finally, a comment about the interpretation of the results. Our methodology only allows identification of $\widetilde{t}_{kn} \equiv t_{kn}/\tau_{kn}$. When commenting on the magnitude of additive trade costs in Section 4.4, we divide the estimates of $\widetilde{t}_{kn}$ by the observed median f.o.b. price in product-destination kn, i.e. $TC_{kn} = (t_{kn}/\tau_{kn})/\overline{\overline{p}}_{kn}$. In other words, we measure additive trade costs relative to the f.o.b. price multiplied by the iceberg cost. As a consequence, our estimates of additive trade costs would be higher if we had information about τ_{kn} and were to report $t_{kn}/\overline{\overline{p}}_{kn}$.

4.3 Discussion

In this section, we address a number of potential concerns with our empirical framework.

Endogeneity. A potential issue is that prices and quantities are determined simultaneously, so that the error term is correlated with the explanatory variables. Our estimator for $\widetilde{t}_{kn}$ is, however, robust to supply side mechanisms that make $\widetilde{p}_{knr}$ endogenous. For example, assume that firms facing favorable demand shocks ϵ_{knr} also charge higher prices, e.g. $\epsilon_{knr} = \phi_k \ln p_{knr} + v_{knr}$ where v_{knr} is an i.i.d. error term. In that case, the estimating equation would be similar to equation (5), the only difference being the interpretation of the slope parameter, which would take the form $\sigma_k + \phi_k$. Therefore, even though the interpretation of the slope parameter would change, the estimate of $\widetilde{t}_{kn}$ would not. In general, the slope coefficient is a mixture of various structural supply and demand side parameters and any particular element is not separately identified (e.g. the demand elasticity σ_k). Identification of the trade cost coefficient is instead based on systematic nonlinear deviations from this equilibrium relationship between price and quantity.

Quality heterogeneity within and across markets. A related concern is that unobserved quality could be correlated with f.o.b. prices. As long as unobserved quality can be written as a linear function of the (log) price, we would get biased slope coefficients σ_k, whereas the estimates of trade costs would remain unchanged. Hence, our methodology is robust to unobserved quality heterogeneity *within* HS-8 product categories. Furthermore, a model with firms varying their level of quality across markets for a given product, perhaps due to country income differences such as in Verhoogen, 2008, would not affect the estimate of trade costs. In our framework, quality differences *across* markets would be captured by the constant term $\tilde{a}_{kn}$ in the demand equation (5).

Selection bias. Firms are not randomly entering into different product-destinations and this can create a correlation between prices and the error term. We hypothesize that the correlation is positive, since firms with both adverse demand shocks and high prices are less likely to be exporting. Analogous to the case with endogenous prices, such a selection effect would only affect the slope parameters, and not the estimates of trade costs. We refer the reader to appendix A.4 for further details.

A different selection issue is that product-destination pairs characterized by high additive trade costs might not be active at all and, therefore, not appear in our data. Indeed, one of the implications of additive trade costs, that we illustrate in Section 5, is that even the most productive firm receives finite revenues that may not be sufficient to cover the entry cost in a given export market. Hence, we are only able to identify trade costs of traded goods, in contrast to all potentially tradable goods. This is an inherent constraint with our methodology, and we emphasize that average trade costs for all goods, including non-traded goods, might be even higher than the ones we uncover here.

Interpretation. We emphasize that although $\widetilde{t}_{kn}$ is, by definition, constant across firms within an HS-8 product category (e.g. same \$20 trade cost for all pairs of shoes exported to the U.S.), our framework allows for *total* trade costs that vary across firms within a product-destination pair. Iceberg costs, τ_{kn}, are controlled for because they are subsumed into the intercept terms $\tilde{a}_{kn}$. Hence, any mechanism that would make $\widetilde{t}_{kn}$ vary systematically with product value would be subsumed into these terms. This shows that the $\widetilde{t}_{kn}$ that we identify is, by definition, the cost that is constant across all firms within a product-destination pair.

4.4 Results

Given the estimates of $\tilde{t}_n$ and $\tilde{t}_k$, we calculate trade costs relative to f.o.b. prices, $TC_{kn} = \widetilde{t}_{kn}/\overline{\overline{p}}_{kn}$, where $\overline{\overline{p}}_{kn}$ is the median f.o.b. price in product-destination pair kn. We report various moments of TC_{kn} in Table 2.[17] The unweighted mean of TC_{kn}, averaged over all products and destinations, is 0.14. The weighted mean and median are smaller, suggesting that product-destination pairs with low trade costs have higher export volumes. As expected, trade costs are heterogeneous: the standard deviation of TC_{kn} is 0.20, while the 75/25 percentile ratio is 9.59.

95 and 88 percent of the $\tilde{t}_n$ and $\tilde{t}_k$ coefficients (the destination and product fixed effects) are significantly different from zero at the 0.05 level.[18] This suggests that, for the large majority of product-destination pairs, the null hypothesis of zero additive

[17] A few point estimates are in the far left and right tail of the distribution, and they tend to disproportionally affect the means. We therefore truncate our point estimates to values within the 5th to 95th percentile of the distribution.

[18] As $\tilde{t}_n$ and $\tilde{t}_k$ are estimated in logs, the null hypotheses are $\ln \tilde{t}_k = \ln \varepsilon - \overline{\ln \tilde{t}_n}$ and $\ln \tilde{t}_n = \ln \varepsilon - \overline{\ln \tilde{t}_k}$, where $\varepsilon = 1$ is an arbitrary small amount of trade costs, in NOK, and $\overline{\ln \tilde{t}_n}$ and $\overline{\ln \tilde{t}_k}$ are the average of the log fixed effects. The alternative hypotheses are $\ln \tilde{t}_k > \ln \varepsilon - \overline{\ln \tilde{t}_n}$ and $\ln \tilde{t}_n > \ln \varepsilon - \overline{\ln \tilde{t}_k}$.

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

	Weighted mean	Unweighted mean	Median	Std. deviation
Trade costs TC_{kn}	0.06	0.14	0.06	0.20
σ_k	-0.73	-1.49	-1.10	1.85
Criterion function	48,410			
R^2	0.42			
# of Countries	21			
# of Products	121			

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

trade costs (i.e. a model with iceberg costs exclusively) is rejected.[19] Although the magnitude of the average TC_{kn} is relatively small, the economic consequences of those costs are potentially large. We investigate the welfare effects of our estimates in Section 6.

The top panel in Figure 2 shows the kernel density of the slope coefficients σ_k, while the bottom panel shows the kernel density of the intercept coefficients $\tilde{a}_{kn}$.[20] For most products, σ_k is negative, meaning that higher prices translate into lower sales volumes.

The overall fit of the model is adequate, with an R^2 of 0.44. We plot actual export volumes and prices ($\ln x_{knr}$ and $\widetilde{p}_{knr}$) as well as the conditional expectation of export volumes for a few product-destination pairs. In Figure 3, we have chosen all export destinations for product HS 73269000, one of the top products in terms of export value.[21] The solid markers represent the conditional expectation whereas 'x' markers represent the data. F.o.b. prices are on the horizontal axis and export volumes on the vertical axis (in logs). We observe that the model is able to capture a substantial share of the variation in the data.

In the next section we present the results of a number of exercises aimed at validating our estimates. We also provide a number of robustness checks.

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

[20] To improve readability, values below/above the 5th/95th percentile are dropped from the kernel densities.

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

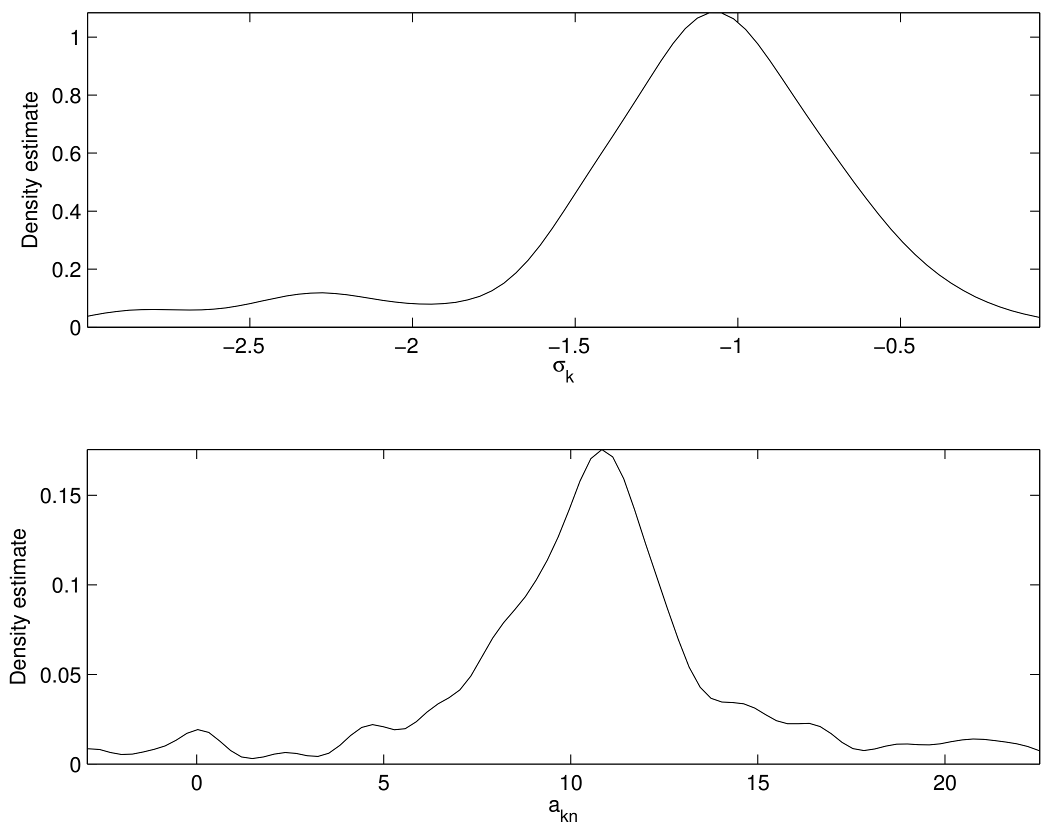

Figure 2: Estimates of σ_k and $\tilde{a}_{kn}$. Kernel densities.

4.5 Validation and Robustness

4.5.1 Validation

In this section, we perform a first validation of our empirical results by correlating the destination component of our trade cost estimates, $\tilde{t}_n$, with distance between Norway and the destination countries; we also correlate our overall trade cost measure, TC_{kn}, with the actual product weight per unit of value. We expect, in both cases, a positive relationship: transportation costs are increasing both in distance and weight (Hummels and Skiba, 2004), and transportation costs are largely additive.[22]

Figure 4 shows our estimates of $\tilde{t}_n$, for every destination, on the vertical axis against distance on the horizontal axis. Both variables are expressed in logarithmic terms. Estimated trade costs are increasing in actual trade costs, as proxied by distance. Note that our two-way fixed effects approach implies that $\tilde{t}_n$ does not depend on the set of products actually exported to n. This implies that there is no selection bias in Figure 4, e.g. that low $\widetilde{t_{kn}}$ products are sold in one destination and high $\widetilde{t_{kn}}$ products in

[22]Referring back to the UPS example in the introduction (see footnote 2), t/τ is increasing in distance and weight, while τ is independent of shipping distance and weight.

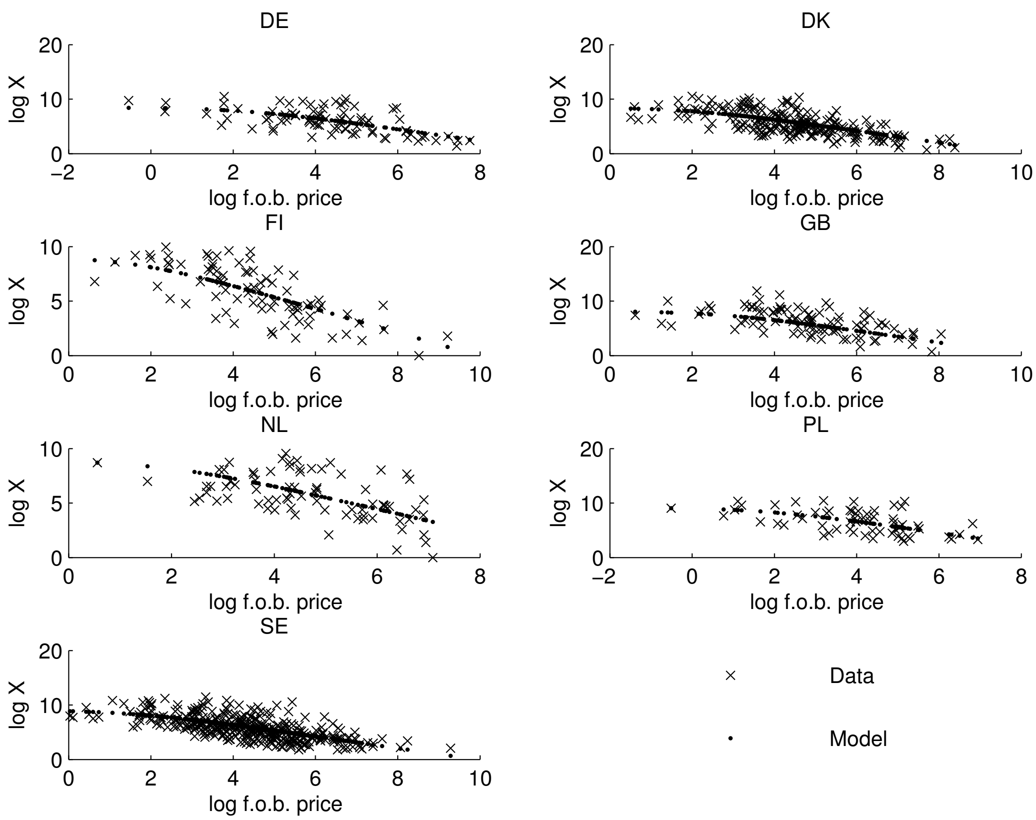

Figure 3: Predicted and actual exports volumes (normalized). HS73269000: articles of iron or steel.

another destination. According to our estimates, trade costs to the U.S. are about 90 percent higher than trade costs to the Netherlands. The robust relationship between distance and trade costs also emerges when regressing estimated trade costs $\tilde{t}_n$ on a set of gravity variables (distance, GDP, and GDP per capita, all in logs). The distance elasticity is then 0.23 (s.e. 0.10).[23]

Figure 5 shows the relationship between TC_{kn} and actual average weight/value across products. Both variables are expressed in logarithmic terms.[24] Since heavier and bulkier goods are more expensive to ship, we expect a positive relationship between weight/value and estimated trade costs. Indeed, the scatter plot shows an upward sloping relationship, with a correlation of 0.32 (p-value 0.002). Most of the estimates in the product dimension also make intuitive sense. For example, aluminum profiles (HS 76042900) are among the products with estimated TC_{kn} above the 95^{th} percentile.

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

[24] Average weight/value is obtained by dividing total weight (summed over firms) over total value (summed over firms) in Sweden. We condition on Sweden to minimize selection effects and to maximize the number of products with non-missing values. Estimated trade costs per product are simply TC_{kSE}.

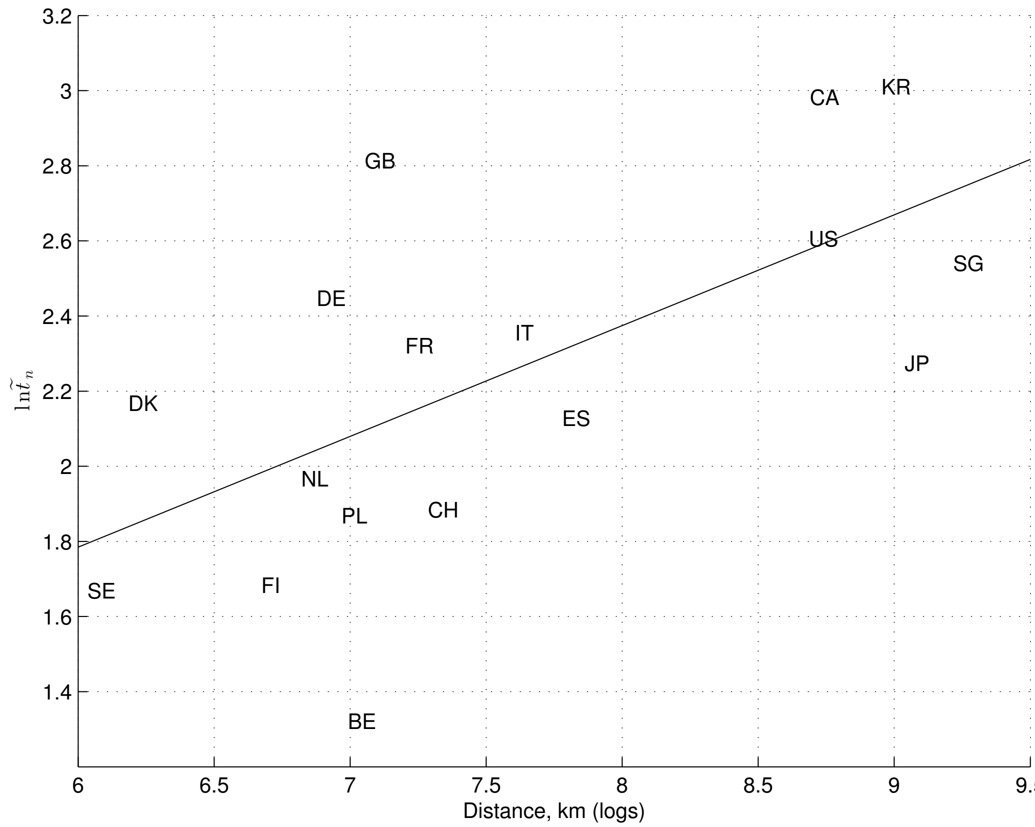

Figure 4: Estimates of $\tilde{t}_n$ and distance (logs).

Lightweight computer equipment (HS 84713000) is among the products with estimated TC_{kn} below the 5^{th} percentile.[25]

A Monte Carlo Experiment. We evaluate the precision of our estimates using a Monte Carlo simulation. We simulate the full general equilibrium model, presented in Section 5, to generate 200 datasets of f.o.b. prices and quantities for a few destinations and products, and estimate additive trade costs using the methodology from Section 4.1. Our methodology recovers the true value of additive trade costs with high precision.

We start by drawing 200 i.i.d. lognormally distributed demand shocks ϵ_{knr} for each firm-destination-product combination. For each realization of ϵ_{knr}, we solve the model for three export destinations and two products according to the steps shown in appendix A.3. The full set of parameters used in the simulation is shown in Table 3. The chosen values of $\ln \tilde{t}_n$ and $\ln \tilde{t}_p$ correspond to trade costs t_{kn}/τ_{kn} relative to the median export price (TC_{kn}) between 0.09 and 0.35. The model generates 200 simulated

[25] HS 76042900 = "Bars, rods and solid profiles, of aluminium alloys"; HS 84713000 = "Data-processing machines, automatic, digital, portable, weighing <= 10 kg, consisting of at least a central processing unit, a keyboard and a display".

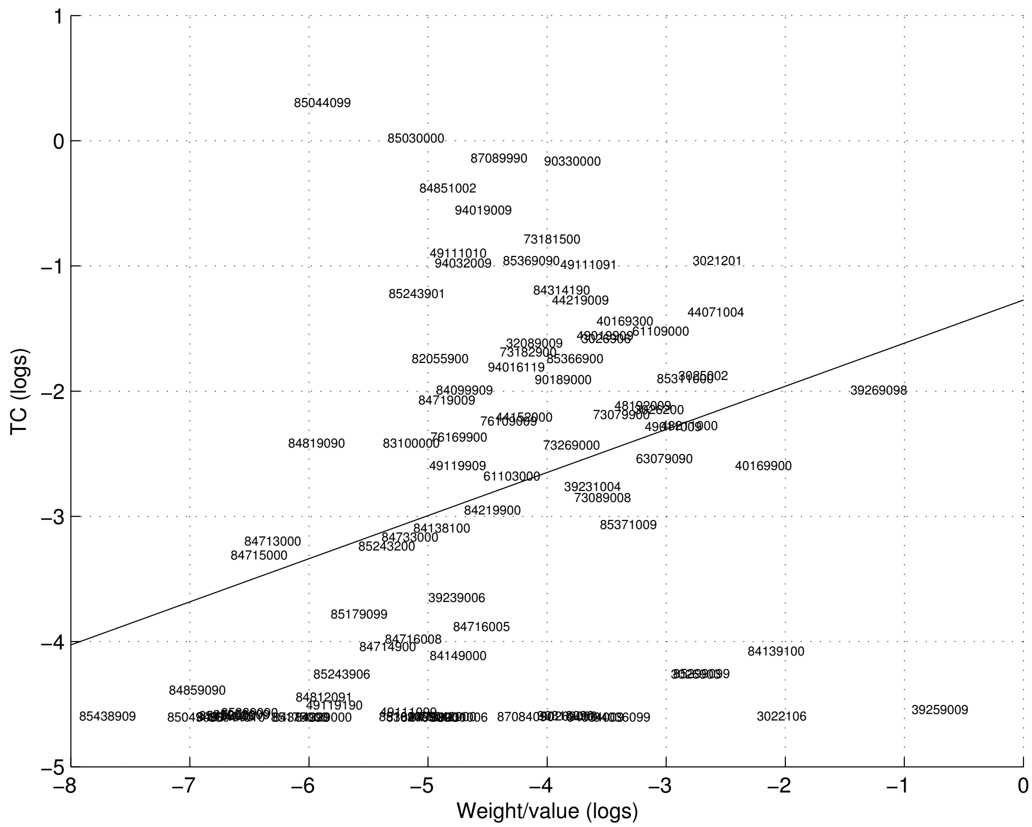

Figure 5: Estimates of TC_{kn} and weight/value (in logs).

datasets consisting of f.o.b. prices $\tilde{p}_{knr}$ and quantities exported x_{knr}. For each simulated dataset, we estimate the reduced form model of equation (5). The results are shown in the lower part of Table 3. The average of the estimated values of $\ln \tilde{t}_n$ and $\ln \tilde{t}_k$, across all samples, is very close to the true value, both across destinations and products. The values in parentheses are the standard deviations of the estimates, showing that the standard errors are acceptable, even in this artificial sample with relatively few observations.[26]

4.5.2 Robustness

In Table 4 we present some re-estimations of the model that address several issues. First, we check whether our estimates are sensitive to the trade cost decomposition $\widetilde{t}_{kn} = \tilde{t}_k \tilde{t}_n$ by estimating $\widetilde{t}_{kn}$ directly for all possible product-destination pairs. As there are no longer any interlinkages between different products, we minimize the objective

[26] The artificial sample consists, on average, of 546 exporters per product-destination, for 2 products and 3 destinations.

Table 3: Monte Carlo Simulation.

	True		
$\ln \tilde{t}_n$	-2.30	-1.61	-1.20
$\ln \tilde{t}_k$	-0.22		
	Estimated		
$\ln \tilde{t}_n$	-2.32	-1.62	-1.20
	(0.31)	(0.19)	(0.14)
$\ln \tilde{t}_k$	-0.22		
	(0.01)		
Avg. obs per prod.-dest.	546		
Avg. criterion f	0.34		

Note: The estimated values of $\ln \tilde{t}_n$ and $\ln \tilde{t}_k$ are average values across the 200 simulated samples. The values in parentheses are standard deviations of the estimates. Estimates of σ_k, $\tilde{a}_k$ and $\tilde{a}_n$ are omitted from the table. The following parameters are used in the simulated model, see Section 5: $\sigma = 4$, $R = 2000$, $w_i L_i = 1$, $\tau_{in} = 1$, $f_{in} = 0.6$, $\kappa = 0.12$. Productivity z is distributed Pareto with shape parameter $\gamma = 4$. Demand shocks ϵ_{knr} are i.i.d. lognormal , $\ln \mathcal{N}(0, .2)$, which implies that the standard deviation of $\ln \epsilon$ is approximately 11 percent of the mean of log of home sales.

Table 4: Robustness: Alternative specifications

	Separate estimations for each product (R1)	Product-destinations with $\geq$ 20 firms (R2)
Trade costs,		
weighted mean	0.04	0.04
unweighted mean	0.10	0.10
median	0.08	0.09
std. deviation	0.08	
# product-destinations	270	917
# of countries (N)	21	33
# of products (K)	212	378

function product by product.[27] As shown in column (R1), the results are very close to the baseline case.

We also investigate whether the choice of truncating the data set to product-destinations with more than 40 firms affects the results. We choose product-destinations with more than 20 firms, resulting in 33 destinations and 378 products. Exports to all possible combinations of these products and destinations amount to 58.9 percent of total export value. The increase in product-destination pairs makes joint estimation computationally infeasible, so we proceed by estimating product by product, as above. The estimate of average trade costs is again very similar to the baseline case, as shown

[27]The main disadvantage of this approach is that estimates of the trade costs per country are partly driven by selection of different products to different markets.

in column (R2).

5 Back to Theory: General Equilibrium

So far, we have estimated trade costs from a parsimonious framework that highlights the relationship between sales and producer prices, with sparse assumptions about market structure, technologies, and without needing to solve for the general equilibrium. This section completes the presentation of the model. This is a necessary tool to evaluate gains from trade in a world with both additive and multiplicative trade costs, and to quantify the importance of additive trade costs in shaping aggregate trade flows (Section 6).

Our model builds on Melitz (2003), Chaney (2008) and Eaton, Kortum, and Kramarz (2011), but we depart from the previous literature in two important respects. First, we introduce additive trade costs. Second, we allow for two, potentially correlated, layers of heterogeneity: productivity heterogeneity at the firm-level and demand shocks at the firm-destination level.[28] Higher values of the demand shock, resulting in higher demand for a given price, can be interpreted as being associated with higher quality (Khandelwal, 2010, Sutton, 1991). By allowing for a (positive) correlation between demand shocks and prices, we can account for the possibility that the largest exporters are not necessarily the lowest-cost firms.

An implication of the model is that, in the presence of finite additive trade costs, firm revenues are always finite. Hence, even for the most productive firm, latent export revenues may not be sufficient to cover the export market entry cost. This opens up the possibility of zero trade flows between country-pairs, without the need to impose any bound on the support of the productivity distribution (as in Helpman, Melitz, and Rubinstein, 2008) and without treating the set of firms as finite (as in Eaton, Kortum, and Sotelo, 2012).

5.1 Additive Trade Costs

To introduce additive trade costs in the simplest way we assume that the economic environment includes a transport sector, whose services are used as an intermediate input in final goods production. Before going into the specifics, two remarks are necessary. First, as mentioned above, our intepretation of additive trade costs is much

[28]Eaton, Kortum and Kramarz (2011) and Irarrazabal, Moxnes and Opromolla (2013) also allow for firm-destination-specific demand shocks but they are uncorrelated with the productivity draws.

broader than transport costs but we adopt this modelling terminology for simplicity. Second, the way we introduce additive costs in the model is, in practical terms, very similar to the widely used assumption of a frictionless homogeneous good sector.

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

5.2 Prices

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

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

where $x_{in} = A_n \eta_n^{\sigma-1} p_{in}^{-\sigma}$ is the *quantity* demanded.[29] Given a monopolistically competitive market structure and preferences, a firm with efficiency z maximizes profits by setting its consumer price as a constant markup over total marginal production cost,

$$p_{in} = \frac{\sigma}{\sigma-1}\left(\frac{w_i \tau_{in}}{z} + t_{in}\right). \tag{6}$$

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

$$p_{in} = \tau_{in}\tilde{p}_{in} + t_{in}, \tag{7}$$

the producer price can be written as

$$\tilde{p}_{in} = \frac{\sigma}{\sigma-1}\left(\frac{w_i}{z} + \frac{t_{in}}{\sigma \tau_{in}}\right). \tag{8}$$

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

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

5.3 Entry and Cutoffs

As in Chaney (2008), the total mass of potential entrants in country i is $\kappa w_i L_i$, where $\kappa > 0$ is a proportionality constant, so that larger and wealthier countries have more entrants. 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 earn from their portfolio $w_i L_i \pi$, where π is the dividend per share of the global mutual fund.

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

$$\bar{z}_{in}(\eta_n) = \begin{cases} w_i \tau_{in} \left[\lambda_1 \left(\frac{f_{in}}{\eta_n Y_n} \right)^{1/(1-\sigma)} P_n - t_{in} \right]^{-1} & \text{if } t_{in} < \bar{t}_{in}, \\ \infty & \text{if } t_{in} \geqslant \bar{t}_{in}, \end{cases} \tag{9}$$

where $\bar{t}_{in} = \lambda_1 \left[f_{in} / \left(\eta_n Y_n \right) \right]^{1/(1-\sigma)} P_n$, and $\lambda_1 = (\sigma/\mu)^{1/(1-\sigma)} \left(\sigma - 1 \right) / \sigma$ is a constant. In the presence of finite additive trade costs, even the most productive firm receives finite revenues that may not be sufficient to cover the entry cost in market n. Therefore, the entry hurdle can be infinite, opening up the possibility of zero trade flows between country-pairs. Note that, unlike in Helpman, Melitz, and Rubinstein (2008), zero trade flows will emerge without imposing an upper bound on productivity levels. Also unlike in Eaton, Kortum, and Sotelo (2012), zero trade flows will emerge without assuming a finite integer number of firms.

5.4 Price Levels

Productivity and demand shocks in market n are drawn from a joint distribution with density $f(z, \eta_n)$. The price index is then

$$P_n^{1-\sigma} = \sum_i \kappa w_i L_i \int \int_{\bar{z}_{in}(\eta_n)}^{\infty} \left(p_{in}(z)/\eta_n\right)^{1-\sigma} f(z, \eta_n)\, dz d\eta_n. \tag{10}$$

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

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

The first equation states that the price index is a function of itself (since $\bar{z}_{in}(\eta_n)$ is a function of P_n) and the dividend share π (since $\bar{z}_{in}(\eta_n)$ is a function of Y_n which is a function of π). The second equation states that the dividend share is a function of itself and all price indices. In appendix Section A.3, we explain this further and show how to numerically solve for the general equilibrium.

6 Welfare and Trade Costs

6.1 Welfare

In this section we show that the gains from trade, defined as the increase in real wages from going to frictionless trade, are higher in the case of additive barriers compared to multiplicative barriers, and that the negative impact of additive barriers is exacerbated in the presence of within-industry price heterogeneity.

To fix ideas, we can consider the case of government-imposed barriers like tariffs and quotas. Section A.5 in the appendix shows that a significant share of duties are non-ad valorem (NAVs). In the presence of heterogeneous prices, an additive barrier acts as a wedge between the prices of any two imported goods (Proposition 2). As a consequence, an additive barrier affects the relative consumption of different imported goods. An ad valorem tariff introduces a deadweight loss due to a distortion in the prices of imported goods relative to the prices of domestic goods. An additive tariff also features a distortion in consumption *among* imported goods. This implies that removing trade barriers is associated with a larger increase in real wages in the additive case compared to the multiplicative case.

To quantitatively investigate the gains from trade, we simulate a symmetric two-country version of the model presented above. First, we choose values commonly used in the literature for key parameters of the model, such as the elasticity of substitution and the Pareto slope parameter (Table 6). Second, we use the estimated value of average trade costs, $\bar{TC}_{kn}$ from Table 2. To isolate the welfare consequences of additive trade costs, we focus on the extreme case of only additive costs and set multiplicative costs to zero ($\tau_{in} = 1$).[30] Third, given these assumptions, we (i) simulate the baseline model, (ii) simulate a free trade counterfactual, and (iii) calculate the rise in real wages from (i) to (ii).

The next step is to evaluate gains from trade in the model under the assumption that trade costs are purely multiplicative, and they are similar in magnitude to the additive trade costs used above. However, we face the problem that additive and multiplicative costs are not directly comparable. To see this, note that the t-equivalent iceberg cost is $\tau(t) = p/(p-t)$, so that $\tau(t)$ depends on the price p.[31] Our solution is to construct three different baseline τ's, corresponding to the $\tau(t)$ for the firm in the 25^{th}, 50^{th} and 75^{th} percentile of the export price distribution, respectively. This produces $\tau^{P25} = 1.17$, $\tau^{P50} = 1.14$ and $\tau^{P75} = 1.13$. Hence, e.g. in the $P25$ case, for 75 percent of the firms, $\tau^{P25} > \tau(t)$. This is a very conservative scenario, as three quarters of the firms face higher trade costs in the multiplicative case than in the additive case. We then proceed as above, and (i) simulate the three baseline models, (ii) simulate a free trade counterfactual, and (iii) calculate the rise in real wages from (i) to (ii).

The main results are shown in Table 5. Column (1) shows the % change in real wages under the assumption that the Pareto shape coefficient is $\gamma = 4$ (Simonovska and Waugh, 2011). In the t-only scenario, real wages increase by 11.2 percent. In the τ-only scenarios, the rise is between 5.5 and 6.8 percent. In other words, the gains from trade are between 65 and 104 percent higher in the additive case versus the multiplicative case. Column (ii) shows the gains from trade when firm heterogeneity is higher (γ lower). As expected, the difference between the t and τ cases is now exacerbated, as across-firm distortions are now greater. Overall, our results suggest that additive barriers are much more harmful than multiplicative ones, and that eliminating additive frictions enhances welfare more than eliminating multiplicative ones.

[30] We use the unweighted mean of $TC_{kn} = 0.14$ and normalize domestic trade costs to zero. Note that as $TC_{kn} = (t_{kn}/\tau_{kn})/\bar{\bar{p}}_{kn}$, the input into the model is $t = \bar{TC}\bar{\bar{p}}$, i.e. average trade cost relative to the median f.o.b. price multiplied by the median f.o.b. price in the model.

[31] $\tau\tilde{p} = \tilde{p} + t \Longleftrightarrow \tau(t) = p/(p-t)$.

Table 5: Gains from Trade.

t	τ	w/p, % change		Exports, % change	
		(1) $\gamma = 4$	(2) $\gamma = 3$	(3) $\gamma = 4$	(4) $\gamma = 3$
Yes	No	11.2	19.2	106.3	451.3
No	τ^{25}	6.8	6.7	42.0	33.7
No	τ^{50}	5.8	5.7	33.2	26.9
No	τ^{75}	5.5	5.4	30.5	24.7

Note: Columns show % change in real wages and exports from baseline equilibrium to frictionless trade. $\tau^{P25} = 1.17$, $\tau^{P50} = 1.14$ and $\tau^{P75} = 1.13$.

Table 6: Parameter values and data used in the simulation.

Notation	Description	Value	Source
γ	Pareto shape parameter	4	Simonovska and Waugh (2011)
f	Fixed costs	0.7m USD 2004	Das, Roberts and Tybout (2008)[1]
wL	GDP Norway 2004	260,029m USD	World Bank, current USD.
σ	Elasticity of substitution	4.0	Broda and Weinstein (2006), mean SITC 3-digit.

Notes: 10,000 draws used in simulation. [1]In their paper, the average cost of foreign market entry is estimated to be 0.4m in 1986-USD which is approximately 0.7m 2004-USD.

6.2 Trade Flows

Our simulation also allows us to examine the growth in trade as trade barriers are removed. Intuitively, the additional distortions associated with additive barriers should also dampen trade, so that removing trade barriers will result in higher trade growth in the additive versus multiplicative scenario. Columns (3) and (4) in Table 5 indeed show that this is the case. Compared to the conservative τ^{P25} baseline, removing additive barriers generates more than twice the growth in exports. In other words, even if the majority of firms face higher trade barriers in the multiplicative case then the additive case, removing the additive barriers produces much more trade. This result suggests that inferring (iceberg) trade costs from trade flows using gravity models, as in Anderson and van Wincoop (2004), may overstate trade barriers, as additive costs dampen trade much more than multiplicative costs. It is well known that standard models have difficulties matching the growth in global trade over time (e.g., Yi, 2003); our results suggest that additive trade costs may play an important role. We leave these important questions for future research.

7 Conclusions

In this paper, we develop a new methodology for estimating trade costs. An important property of our framework, which provides the basis for identification, is that an increase in additive trade costs has a systematically different impact on the demand elasticity among low price versus high price firms. It is the marriage of additive costs and price heterogeneity, within narrowly defined industries, that drives the theoretical and empirical results in this paper.

We structurally estimate the magnitude of additive trade costs, for every product and destination in our dataset. Estimated additive trade costs are on average 14 percent, expressed relative to the median price. They are quite heterogeneous across product-destination pairs, and as expected, are highly correlated with distance and product weight/value ratios. We reject the null hypothesis of zero additive trade costs (i.e. a model with iceberg costs exclusively) for the large majority of product-destination pairs.

Using our micro estimates, we show that both the welfare and the trade flows impact of additive trade costs is much higher than the impact of multiplicative trade costs. A reduction in additive trade costs would imply higher welfare gains and larger increases in trade flows than an equivalent reduction in multiplicative trade costs. We show that, especially in industries where price heterogeneity is high, the gains from trade can be much higher if markets are protected by additive instead of multiplicative barriers. Hence, firm level heterogeneity matters for aggregate outcomes, such as gains from trade. Furthermore, additive trade costs can help us understand the prevalence of zeros in bilateral trade flows. We conclude that empirical and theoretical work should account for both (the tip of the) iceberg costs, as well as the part of trade costs that are largely hidden under the surface: additive costs.

Our quantitative framework is potentially useful in a number of applications. Trade costs can easily be estimated for other countries and years, as firm-level trade data are now available for a number of developing and developed countries. For future work, our analysis points to the need for further research in understanding the time-series and geographic response of aggregate trade flows to additive trade costs.

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," *Longitudinal* *employer-Household Dynamics Technical Papers*, 2002-06.

Alchian, A. A., and W. R. Allen (1964): *University Economics*. Wadsworth Publishing Company, Belmont, CA.

Anderson, J. E., and E. van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 1, 170–192.

——— (2004): "Trade Costs," *Journal of Economic Literature*, 42(3), 691–751.

Arkolakis, K., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2012b): "The Elusive Pro-Competitive Effects of Trade.," *Working Paper*.

Arkolakis, K., A. Costinot, and A. Rodríguez-Clare (2012a): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.

Behrens, K., and Y. Murata (2007): "General Equilibrium Models of Monopolistic Competition: A New Approach," *Journal of Economic Theory*, 136, 776–787.

Berman, N., P. Martin, and T. Mayer (2012): "How do Different Exporters React to Exchange Rate Changes? Theory, Empirics and Aggregate Implications," *Quarterly Journal of Economics*, 127(1), 437–492.

Chaney, T. (2008): "Distorted gravity: the Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707–1721.

Corsetti, G., and L. Dedola (2005): "A macroecnomic model of international price discrimination," *Journal of International Economics*, 67(1), 129–155.

Eaton, J., and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

Eaton, J., S. S. Kortum, and F. Kramarz (2011): "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 79(5), 1453–1498.

Eaton, J., S. S. Kortum, and S. Sotelo (2012): "International Trade: Linking Micro and Macro," *NBER Working Paper No. 17864*.

Feenstra, R. C. (2003): "A Homothetic Utility Function for Monopolistic Competition Models, Without Constant Price Elasticity," *Economics Letters*, 78, 79–86.

Feyrer, J. (2009): "Distance, Trade, and Income - The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," *NBER Working Paper No. 15557*.

HAUSMAN, J. A. (1996): “Valuation of New Goods under Perfect and Impefect Competition,” *The Economics of New Goods*, 58.

HEAD, K., AND J. RIES (2001): “Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade,” *American Economic Review*, 4, 858–876.

HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” *The Quarterly Journal of Economics*, 123(2), 441–487.

HUMMELS, D. (2007): “Transportation Costs and International Trade in the Second Era of Globalization,” *Journal of Economic Perspectives*, 21(3), 131–154.

HUMMELS, D., AND A. SKIBA (2004): “Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture,” *Journal of Political Economy*, 6, 1384–1402.

IRARRAZABAL, A. A., A. MOXNES, AND L. D. OPROMOLLA (2013): “The Margins of Multinational Production and the Role of Intrafirm Trade,” *Journal of Political Economy*, 121(1), 74–126.

JACKS, D. S., C. M. MEISSNER, AND D. NOVY (2008): “Trade Costs, 1870–2000,” *American Economic Review Papers and Proceedings*, 2, 529–534.

KHANDELWAL, A. K. (2010): “The Long and Short (of) Quality Ladders,” *Review of Economic Studies*, 77, 1450–1476.

KHANDELWAL, A. K., P. K. SCHOTT, AND S.-J. WEI (2011): “Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters,” *NBER Working Paper No. 17524*.

MARTIN, J. (2012): “Markups, Quality, and Transport Costs,” *European Economic Review*, 56, 777–791.

MELITZ, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71(6), 1695–1725.

OBSTFELD, M., AND K. ROGOFF (2000): “The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?,” *NBER Macroeconomics Annual*, 15, 339–390.

Ottaviano, G., T. Tabuchi, and J.-F. Thisse (2002): "Agglomeration and Trade Revisted," *International Economic Review*, 43, 409–435.

Pavcnik, N. (2002): "Liberalization, Exit, and Productivity Improvements: Evidence From Chilean Plants," *The Review of Economic Studies*, 69, 245–276.

Samuelson, P. A. (1954): "The Transfer Problem and the Transport Costs II: Analysis of Effects of Trade Impediments," *The Economic Journal*, 64, 264–289.

Simonovska, I., and M. E. Waugh (2011): "The Elasticity of Trade: Estimates and Evidence," *NBER Working Paper No. 16796.*

Sørensen, A. (2012): "Additive versus Multiplicative Trade Costs and the Gains from Trade," *Economics Working Papers, University of Aarhus: 2012-07.*

Sutton, J. (1991): *Price Competiton, Advertising, and the Evolution of Concentration.* The MIT Press, Cambridge, MA.

Verhoogen, E. (2008): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 123(2), 489–530.

Yi, K.-M. (2003): "Can Vertical Specialization Explain the Growth of World Trade?," *Journal of Political Economy*, 111(1), 52–102.

A Appendix

A.1 Proof of Proposition 1

Proof. Consider the quantity demanded $x = YP^{\sigma-1}p^{-\sigma}\eta^{\sigma-1}$ for a variety with consumer price p and demand shock η. Let, as in equation (1), $p = \tau\tilde{p} + t$. Using the chain rule, E is the product of the elasticity of quantity demanded with respect to the consumer price and the elasticity of the consumer price with respect to the f.o.b. price ($\tilde{p}$),

$$E\left(\tilde{p}, t, \tau, \sigma\right) = \frac{\partial \ln x}{\partial \ln \widetilde{p}} = \frac{\partial \ln x}{\partial \ln p}\frac{\partial \ln p}{\partial \ln \widetilde{p}} = -\sigma\left(1 - \frac{t}{p}\right) < 0,$$

since

$$\begin{gathered} \tfrac{\partial \ln x}{\partial \ln p} = -\sigma < 0, \\ \tfrac{\partial \ln p}{\partial \ln \widetilde{p}} = \tfrac{\tau\tilde{p}}{\tau\tilde{p}+t} = \left(1 - \tfrac{t}{p}\right) \geq 0. \end{gathered}$$

since $\sigma > 1$ and $p \geq t$.

In the absence of additive trade costs ($t = 0$), $\partial \ln p / \partial \ln \widetilde{p} = 1$ so that the elasticity of quantity demanded with respect to the f.o.b. price is equal to the elasticity of substitution across varieties. In the presence of additive trade costs ($t > 0$), the following results apply:

$$\frac{\partial E}{\partial t}\frac{t}{E} = -\frac{\sigma}{E}\frac{\partial\left(1 - \frac{t}{p}\right)}{\partial \ln t} = \frac{\sigma t}{Ep^2}\left(p - t\right) = \frac{\sigma t \tau \tilde{p}}{Ep^2} \leq 0,$$

since $\sigma > 1$, $t \geq 0$, $\tau \geq 1$, $\tilde{p} > 0$, and $E < 0$. This shows that, in the presence of additive trade costs, the elasticity of E to additive trade costs is negative. Finally,

$$\begin{aligned}\frac{\partial}{\partial \tilde{p}}\left(\frac{\partial E}{\partial t}\frac{t}{E}\right) &= \sigma t \tau \frac{\partial}{\partial \tilde{p}}\left(\frac{\tilde{p}}{Ep^2}\right) = \sigma t \tau \left(\frac{1}{Ep^2} - \frac{\tilde{p}}{E^2p^4}\left(p^2\frac{\partial E}{\partial \tilde{p}} + E\frac{\partial p^2}{\partial \tilde{p}}\right)\right) \\ &= \sigma t \tau \left(\frac{1}{Ep^2} - \frac{\tilde{p}}{E^2p^4}\left(-\sigma t \tau + 2p\tau E\right)\right) = \frac{\sigma t \tau}{Ep^2}\left(1 + \frac{\sigma \tau \tilde{p}}{Ep^2}\left(2\tau\tilde{p} + t\right)\right) \\ &= \frac{\sigma t \tau}{Ep^2}\left(1 - \frac{2\tau\tilde{p} + t}{p}\right) = -\frac{\sigma t \tilde{p} \tau^2}{Ep^3} > 0\end{aligned}$$

since

$$\frac{\partial E}{\partial \tilde{p}} = -\frac{\sigma t \tau}{p^2}$$

and

$$\frac{\partial p^2}{\partial \tilde{p}} = 2p\tau.$$

This shows that the elasticity of E with respect to additive trade costs t is strictly increasing in the producer price $\tilde{p}$.

$\square$

A.2 A Broader Class of Demand Systems

In this section we show how Proposition 1 and the identification strategy used in the estimation of trade costs extend to a general class of demand systems.

A.2.1 General Demand System

Similarly to Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2012b), we consider a general demand system for differentiated goods. All consumers have the same preferences. If a consumer with income w faces a vector of prices $\boldsymbol{p}$, her Marshallian demand for any differentiated good takes the form

$$\ln x\left(p, p^*, w\right) = -\beta \ln p + \gamma \ln w + f\left(\ln p - \ln p^*\right), \tag{11}$$

where x and p are quantity and consumer price respectively, and w is income. $p^*\left(\boldsymbol{p}\right)$ is a price index which is symmetric in all prices $\boldsymbol{p}$. Hence, other prices only affect demand through their effect on the aggregator p^*. Depending on the assumptions on β, γ, $f(.)$, and p^*, the above framework encompasses, among others, four different utility functions that have been widely used in the literature: (a) CES utility; (b) quadratic, but non-separable utility, as in Ottaviano, Tabuchi, and Thisse (2002); (c) a translog expenditure function, as in Feenstra (2003), and (d) additively quasi-separable utility, as considered by Behrens and Murata (2007).[32]

Table 7 summarizes the results regarding Proposition 1. In the absence of additive trade costs, the elasticity of the quantity demanded to the producer price is $(-\beta + f')$. In the presence of additive trade costs, the elasticity becomes

$$E = \left(-\beta + f'\right)\left(1 - \frac{t}{p}\right).$$

The elasticity of E with respect to additive trade costs is

$$\begin{aligned}
\frac{\partial E}{\partial t}\frac{t}{E} &= \frac{t}{E}\frac{f''}{p}\left(1 - t/p\right) - \frac{t}{E}\left(-\beta + f'\right)\frac{p-t}{p^2} \\
&= \frac{t}{p}\left(\frac{f''}{f' - \beta} - \frac{1}{1 - t/p}\frac{1 - t/p}{1}\right) \\
&= \frac{t}{p}\left(\frac{f''}{f' - \beta} - 1\right).
\end{aligned}$$

Note that the sign of the previous elasticity varies with the type of preferences considered. In the case of quadratic, but non-separable utility function, as in Ottaviano, Tabuchi, and Thisse (2002),

$$\frac{f''}{f' - \beta} - 1 = \frac{1}{e^{-s} - 1} > 0,$$

[32]Specifically, $\gamma = 1$, $f(s) = (1 - \sigma)\, s$, $\beta = 1$ in case (a); $\beta = -1$, $\gamma = 0$, and $f(s) = -\ln \kappa_2 + \ln\left(e^{-s} - 1\right)$ in case (b); $\beta = \gamma = 1$ and $f\left(s\right) = ln\xi + ln\left(-s\right)$ in case (c); and $\beta = \gamma = 0$ and $f(s) = \ln \zeta + \ln\left(-s\right)$ in case (d).

since $s<0$. In the case of translog preferences, as in Feenstra (2003),

$$\frac{f''}{f'-\beta}-1=-\frac{1+s\left(1-s\right)}{s\left(1-s\right)},$$

where

$$1+s\left(1-s\right)>0,$$

so that $\frac{\partial E}{\partial t}\frac{t}{E}<0$ for $\frac{-1+\sqrt{5}}{-2}<s<0$. In the case of Behrens and Murata (2007) preferences,

$$\frac{f''}{f'-\beta}-1=\frac{-1-s}{s}<0,$$

for $-1<s<0$.

Finally, we show under what conditions the elasticity of E with respect to additive trade costs is increasing in $\tilde{p}$.

$$\begin{aligned}\frac{\partial}{\partial \ln p}\left(\frac{\partial E}{\partial t}\frac{t}{E}\right) &= -te^{-lnp}\left(\frac{f''}{f'-\beta}-1\right)+te^{-lnp}\frac{f'''\left(f'-\beta\right)-f''^{2}}{\left(f'-\beta\right)^{2}}\\ &= -\frac{t}{p}\left(\frac{f''}{f'-\beta}-1\right)+\frac{t}{p}\frac{f'''-f''^{2}/\left(f'-\beta\right)}{f'-\beta}\\ &= \frac{t}{p\left(f'-\beta\right)^{2}}\left[\left(f'-\beta\right)^{2}+\left(f'''-f''\right)\left(f'-\beta\right)-f''^{2}\right].\end{aligned}$$

Therefore, as long as

$$\left(f'-\beta\right)^{2}+\left(f'''-f''\right)\left(f'-\beta\right)-f''^{2}>0, \tag{12}$$

the elasticity of E with respect to additive trade costs is increasing in $\ln p$. Given that $\partial \ln p/\partial\widetilde{p}=\tau/p>0$, the same condition guarantees that the elasticity of E with respect to additive trade costs is increasing in $\tilde{p}$.

In the case of quadratic, but non-separable utility function, as in Ottaviano, Tabuchi, and Thisse (2002), condition (12) becomes

$$\begin{aligned}&\left(-\frac{e^{-x}}{\left(e^{-x}-1\right)}-\beta\right)^{2}+\left(-\frac{\left(e^{-x}+e^{-2x}\right)}{\left(e^{-x}-1\right)^{3}}+\frac{e^{-x}}{\left(e^{-x}-1\right)^{2}}\right)\left(-\frac{e^{-x}}{\left(e^{-x}-1\right)}-\beta\right)-\frac{e^{-2x}}{\left(e^{-x}-1\right)^{4}}\\ =&\frac{1}{\left(e^{-x}-1\right)^{2}}+\frac{2e^{-x}}{\left(e^{-x}-1\right)^{3}}\frac{1}{\left(e^{-x}-1\right)}-\frac{e^{-2x}}{\left(e^{-x}-1\right)^{4}}\\ =&\frac{2e^{-x}-e^{-2x}+\left(e^{-x}-1\right)^{2}}{\left(e^{-x}-1\right)^{4}}=1>0,\end{aligned}$$

	CES	Quadratic	Translog	Behrens & Murata (2007)
$\frac{\partial \ln x}{\partial \ln \tilde{p}} \mid_{t=0}$	$-\sigma < 0$	$-\frac{1}{e^{-s}-1} < 0$	$\frac{1-s}{s} < 0$	$\frac{1}{s} < 0$
$E = \frac{\partial \ln x}{\partial \ln \tilde{p}} \mid_{t>0}$	$-\sigma \left(1 - \frac{t}{p}\right) < 0$	$-\frac{1}{e^{-s}-1} \left(1 - \frac{t}{p}\right) < 0$	$\frac{1-s}{s} \left(1 - \frac{t}{p}\right) < 0$	$\frac{1}{s} \left(1 - \frac{t}{p}\right) < 0$
$\frac{\partial E}{\partial t} \frac{t}{E}$	< 0	> 0	$< 0 \;\; if \;\; \frac{1-\sqrt{5}}{2} < s < 0$ $> 0 \;\; if \;\; s < \frac{1-\sqrt{5}}{2}$	$< 0 \;\; if \;\; -1 < s < 0$ $> 0 \;\; if \;\; s < -1$
$\frac{\partial}{\partial \ln \tilde{p}} \left(\frac{\partial E}{\partial t} \frac{t}{E}\right)$	> 0	> 0	> 0	> 0

Table 7: Demand systems and additive trade costs.

which is clearly satisfied.

In the case of translog preferences, as in Feenstra (2003), condition (12) becomes

$$1 - 2x^{-1} + x^{-4} - x^{-3} > 0,$$

which is clearly satisfied since $x < 0$.

In the case of Behrens and Murata (2007) preferences, condition (12) becomes

$$\begin{aligned} & \left(x^{-1}\right)^2 + \left(2x^{-3} + x^{-2}\right)\left(x^{-1}\right) - x^{-4} \\ = \;\; & x^{-4}\left(1 + x^2 + x\right) > 0, \end{aligned}$$

which is clearly satisfied since $x < 0$.

Table 7 summarizes the signs of E, $(\partial E/\partial t)/(E/t)$ and $\partial\left[(\partial E/\partial t)/(E/t)\right]/\partial \ln \tilde{p}$ under the different demand systems (recall that $s = \ln p - \ln p^* \leq 0$).

In sum, under all demand systems considered here, the finding in Proposition 1, that the elasticity of E with respect to additive trade costs t is strictly increasing in the producer price $\tilde{p}$, continues to hold. The reduced form evidence in Table 1 (comparing $(\partial E/\partial Dist)/(E/Dist)$ for high and low price firms) therefore still holds in the more general case. The structural estimation in Section 4 can also be extended so that the demand elasticity with respect to the c.i.f. price is allowed to vary over the price distribution. Specifically, instead of the linear specification in equation (4), a non-linear specification with higher-order polynomial terms can be used.

A.3 Simulating the Model

A.3.1 Numerical approximation

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

1. Choose a starting value of the the dividend share and the price indices, π^0 and

P_n^0, where superscripts denote the round of iteration.

2. Solve the the system of $N+1$ equations and $N+1$ unknowns characterized by

$$P_n^d = g\left(P_n^{d-1}, \pi^{d-1}\right) \ \forall \ n,$$
$$\pi^d = h\left(\pi^{d-1}, P_1^{d-1}, .., P_N^{d-1}\right),$$

as shown in the main text. This involves solving the equilibrium cutoffs (9) and the expression for the price index (10).

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

Since the price index has no closed-form solution, we approximate it with Monte Carlo methods. Specifically, we take $R = 1e+5$ random draws z^r and η^r from the density $f(z, \eta_n)$. An integral of the form

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

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

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

A.3.2 Global Profits

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

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

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

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

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

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

Solving for π yields

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

Note that since $n_{in} = \kappa w_i L_i \int \int_{\bar{z}_{in}(\eta_n)} f(z, \eta_n) \, dz d\eta_n$, π is only a function of the entry hurdle function $\bar{z}_{in}(\eta_n)$. Replacing $\bar{z}_{in}(\eta_n)$ with the entry hurdle expression (9), π becomes a function of itself and the price indices, $\pi = h(\pi, P_1, .., P_N)$ (suppressing all exogenous variables).

A.4 Selection bias

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

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

$$\ln p_{knr} - \ln \xi_{kn} - \ln \eta_{knr} > 0.$$

Export volume is, from equation (4),

$$\ln x_{knr} = a_{kn} + \sigma_k \ln p_{knr} + (\sigma_k - 1) \ln \eta_{knr}.$$

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

$$E\left[\ln \eta_{knr} \middle| \ln p_{knr} - \ln \xi_{kn} - \ln \eta_{knr} > 0\right] = \lambda \left[\ln p_{knr} - \ln \xi_{kn}\right]$$

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

$$\ln x_{knr} = a_{kn} - \sigma_k \ln p_{knr} + \lambda \left[\ln p_{knr} - \ln \xi_{kn}\right] + v_{knr}.$$

Approximating the Mills ratio with the polynomial $\lambda() = \check{a}_{kn} - \check{\sigma}_k \ln p_{knr}$, we get

$$\ln x_{knr} = a_{kn} + \check{a}_{kn} - (\sigma_k + \check{\sigma}_k) \ln p_{knr} + v_{knr}.$$

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

A.5 The prevalence of non-ad valorem duties (NAVs)

A significant share of duties are non-ad valorem (NAVs). According to the WTO World Tariff Profiles (2006), "NAVs are applied by 68 out of the 151 countries shown in this publication including several LDCs..." Table 8 reports, for a set of countries, the share of Harmonized System six-digit subheadings (both for agricultural and non-agricultural products) subject to non-ad valorem duties. The share of products subject to NAVs is usually higher in the case of agricultural products but is also important for non-agricultural products. For example, in the United States, the 3.4% of non-agricultural products that are subject to NAVs account for 18.9% of imports. Still according to the WTO World Tariff Profiles (2006) "One of the peculiarities of NAVs resides in the fact that even if they are applied to a limited number of tariff lines, the products concerned are often classified as sensitive, either because governments collect significant tariff revenues, e.g. cigarettes and alcoholic drinks, or for protecting domestic products against lower priced imports. These highlight the importance of analyzing NAVs."

Table 8: Non-ad Valorem Tariffs and Tariff Quotas

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

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