The Elasticity of Trade: Estimates and Evidence

Ina Simonovska

University of California, Davis and NBER

Michael E. Waugh

New York University

First Version: April 2009

This Version: November 2010

ABSTRACT

Quantitative results from a large class of structural gravity models of trade depend critically on the elasticity of trade with respect to trade frictions. Eaton and Kortum (2002) develop an innovative method to estimate this parameter from disaggregate price data. We prove that their estimator is biased in finite samples, but it is consistent. We quantitatively show that the bias is severe and that the data requirements necessary to eliminate it in practice are extreme. We then develop a simulated method of moments estimator that builds on their methodology and we demonstrate its effectiveness in small samples. We apply it to new disaggregate price and trade flow data for the year 2004 and estimate the elasticity of trade for 123 countries. Our estimate of the elasticity is roughly 4, nearly 50 percent lower than Eaton and Kortum's (2002) estimation strategy suggests. This difference doubles the welfare gains from international trade across various models.

JEL Classification: F10, F11, F14, F17

Keywords: elasticity of trade, bilateral, gravity, price dispersion, indirect inference

Email: inasimonovska@ucdavis.edu, mwaugh@stern.nyu.edu.

We are grateful to the World Bank for providing us with the price data from the 2005 ICP round. We thank George Alessandria, Alexander Aue, Robert Feenstra, Timothy Kehoe, Matthias Lux, B. Ravikumar, seminar participants at UC San Diego, Syracuse, ETH/KOF, Princeton, Uppsala, Oslo, San Francisco Fed, UC Berkeley, NYU and participants at the Conference on Microeconomic Sources of Real Exchange Rate Behavior at Vanderbilt, Midwest International Trade Meeting Fall 2010, SED 2010, WEAI 2010, Conference on Trade Costs and International Trade Integration: Past, Present and Future in Venice, International Comparisons Conference at Oxford, North American Winter Meeting of the Econometric Society 2010, AEA 2010 for their feedback.

1. Introduction

Quantitative results from a large class of structural gravity models of international trade depend critically on a single parameter that governs the elasticity of trade with respect to trade frictions.[1] To illustrate how important this parameter is, consider three examples: Anderson and van Wincoop (2003) find that the estimate of the tariff equivalent of the U.S.-Canada border varies between 48 and 19 percent depending upon the assumed elasticity of trade with respect to trade frictions. Yi (2003) points out that observed reductions in tariffs can explain almost all or none of the growth in world trade depending upon this elasticity. Arkolakis, Costinot, and Rodriguez-Clare (2009) argue that this parameter is one of only two parameters needed to measure the welfare cost of autarky in a large and important class of trade models. Therefore, this elasticity is key to understanding the size of the frictions to trade, the response of trade to changes in tariffs, and the welfare gains or losses from trade.

Estimating this parameter is difficult because quantitative trade models can rationalize small trade flows with either large trade frictions and small elasticities, or small trade frictions and large elasticities. Thus, one needs satisfactory measures of trade frictions *independent* of trade flows to estimate this elasticity. Eaton and Kortum (2002) provided an innovative and simple solution to this problem by arguing that, with product-level price data, one could use the maximum price difference across goods between countries as a proxy for bilateral trade frictions. The maximum price difference between two countries is meaningful because it is bounded by the trade friction between the two countries via simple no-arbitrage arguments.

We incorporate the intuition of Eaton and Kortum (2002) into a new simulated method of moments estimator for the elasticity of trade. The argument for a new estimator is that the method employed by Eaton and Kortum (2002) results in estimates that are biased upward by economically significant magnitudes. We demonstrate the result formally and provide quantitative measures of the bias via monte carlo simulations of the Eaton and Kortum (2002) model. Finally, we apply the new estimator on novel disaggregate price and trade flow data for the year 2004 spanning 123 countries that account for 98 percent of world output.[2] Our benchmark estimate for the elasticity of trade is 4.22, rather than 7-9 as the estimation strategy of Eaton and Kortum (2002) suggests. This difference doubles the measured welfare gains from international trade across various models.

Since the elasticity of trade plays a key role in quantifying the welfare gains from trade,

[1] These models include Krugman (1980), Anderson and van Wincoop (2003), Eaton and Kortum (2002), and Melitz (2003) as articulated in Chaney (2008), which all generate log-linear relationships between bilateral trade flows and trade frictions.

[2] We obtain price data from the International Comparison Programme (ICP) and bilateral trade flow data from UN Comtrade.

it is important to understand why our estimates of the parameter differ substantially from Eaton and Kortum's (2002). We show that the reason behind the difference is that Eaton and Kortum's (2002) estimator is biased in finite samples of price data. The bias arises because the model's equilibrium no-arbitrage conditions imply that the maximum operator over a finite sample of prices underestimates the trade cost with positive probability and overestimates the trade cost with zero probability. Consequently, the maximum price difference lies strictly below the true trade cost, in expectation. This implies that Eaton and Kortum's (2002) estimator delivers an elasticity of trade that lies strictly above the true parameter, in expectation. As the sample size grows to infinity, Eaton and Kortum's (2002) estimator can uncover the true elasticity of trade, which necessarily implies that the bias in the estimates of the parameter is eliminated.

Quantitatively the bias is substantial. To illustrate its severity, we discretize the model of Eaton and Kortum (2002), simulate trade flows and product-level prices under an assumed elasticity of trade, and apply their estimating approach on artificial data. Assuming a trade elasticity of 8.28, Eaton and Kortum's (2002) preferred estimate for nineteen OECD countries in 1990, the procedure suggests an estimate of 12.5, nearly 50 percent higher than originally postulated. Moreover, in practice, the true parameter can be recovered when 50,000 goods are sampled across the nineteen economies, which constitutes an extreme data requirement to produce unbiased estimates of the elasticity of trade.

Based on these arguments we propose an estimator that is applicable when the sample size of prices is small. Our approach exploits the ability to use observed bilateral trade flows to recover all sufficient parameters to simulate Eaton and Kortum's (2002) model and obtain trade flows and prices as functions of the parameter of interest. We then employ a simulated method of moments estimator that minimizes the distance between the parameters from the approach of Eaton and Kortum (2002) on real and artificial data. We explore the properties of this estimator numerically using simulated data and show that it can uncover the true elasticity of trade in contrast to model-independent bias-reduction methods.

Furthermore, since our simulation approach makes use of a structural model of international trade, we are able to address measurement error in price data by simulating prices with log-normal distributed errors. Aggregation bias, in turn, is an issue, because prices for our large sample of countries are reported at a so-called "basic-heading level". The price of a basic heading reflects an average price across a set of varieties of a particular good, thus potentially washing away extreme price differences across countries, which are necessary to obtain estimates of trade barriers. Consequently, we conduct robustness analysis by applying our approach to a more detailed cross-country product-level price database, provided by the Economist Intelligence Unit. The data, collected in high- and low-end stores across

77 countries, also allow us to check whether variable mark-ups affect the estimates of the elasticity.

Although we employ retail price data, we argue that the resulting elasticity of trade estimates are not tainted by the presence of country-specific sales taxes and distribution costs. The simple intuition behind this result is that, should relative retail prices reflect various mark-ups in a multiplicative fashion, these mark-ups are also reflected in the estimates of trade costs, which employ the price data, and thus they perfectly cancel out in all estimating equations.

The robustness exercises allow us to establish a range for the elasticity of trade between 2.55 and 4.49. These results favor well with estimates for the parameter obtained using different structural approaches. Our findings essentially bridge the gap between the estimates arising from the structured approach of Eaton and Kortum (2002) and the existing empirical literature. Thus, by resolving the bias present in Eaton and Kortum's (2002) estimate, we double the welfare gains from international trade across a large class of models.

2. Model

We outline the environment of the multi-country Ricardian model of trade introduced by Eaton and Kortum (2002). We consider a world with N countries, where each country has a tradable final goods sector. There is a continuum of tradable goods indexed by $j \in [0, 1]$.

Within each country i, there is a measure of consumers L_i. Each consumer has one unit of time supplied inelastically in the domestic labor market and enjoys the consumption of a CES bundle of final tradable goods with elasticity of substitution $\rho > 1$:

$$U_i = \left[\int_0^1 x_i(j)^{\frac{\rho-1}{\rho}} dj \right]^{\frac{\rho}{\rho-1}}.$$

To produce quantity $x_i(j)$ in country i, a firm employs labor using a linear production function with productivity $z_i(j)$. Country i's productivity is in turn the realization of a random variable (drawn independently for each j) from its country-specific Fréchet probability distribution:

$$F_i(z_i) = \exp(-T_i z_i^{-\theta}). \tag{1}$$

The country-specific parameter $T_i > 0$ governs the location of the distribution, higher values of it imply that a high productivity draw for any good j is more likely. The parameter $\theta > 1$

is common across countries and if higher, it generates less variability in productivity across goods.

Having drawn a particular productivity level, a perfectly competitive firm from country i incurs a marginal cost to produce good j of $w_i/z_i(j)$, where w_i is the wage rate in the economy. Shipping the good to a destination n further requires a per-unit iceberg trade cost of $\tau_{ni} > 1$ for $n \neq i$, with $\tau_{ii} = 1$. We assume that cross-border arbitrage forces effective geographic barriers to obey the triangle inequality: For any three countries i, k, n, $\tau_{ni} \leq \tau_{nk}\tau_{ki}$.

Below we describe equilibrium prices, trade flows, and welfare.

Perfect competition forces the price of good j from country i to destination n to be equal to the marginal cost of production and delivery:

$$p_{ni}(j) = \frac{\tau_{ni} w_i}{z_i(j)}.$$

So consumers in destination n would pay $p_{ni}(j)$, should they decide to buy good j from country i.

Consumers purchase good j from the low-cost supplier, thus the actual price consumers in n pay for good j is the minimum price across all sources k:

$$p_n(j) = \min_{k=1,\ldots,N} \left\{ p_{nk}(j) \right\}. \tag{2}$$

The pricing rule and the productivity distribution allow us to obtain the following CES exact price index for each destination n:

$$P_n = \gamma \Phi_n^{-\frac{1}{\theta}} \quad \text{where} \quad \Phi_n = \left[\sum_{k=1}^{N} T_k (\tau_{nk} w_k)^{-\theta} \right]. \tag{3}$$

In the above equation, $\gamma = \left[\Gamma\left(\frac{\theta+1-\rho}{\theta} \right) \right]^{\frac{1}{1-\rho}}$ is the Gamma function and parameters are restricted such that $\theta > \rho - 1$.

To calculate trade flows between countries, let X_n be country n's expenditure on final goods, of which X_{ni} is spent on goods from country i. Since there is a continuum of goods, computing the fraction of income spent on imports from i, X_{ni}/X_n, can be shown to be equivalent to finding the probability that country i is the low-cost supplier to country n given the joint distribution of efficiency levels, prices, and trade costs for any good j. The expression for the share of expenditures that country n spends on goods from country i or, as we will call

it, the trade share is:

$$\frac{X_{ni}}{X_n} = \frac{T_i(\tau_{ni}w_i)^{-\theta}}{\sum_{k=1}^{N} T_k(\tau_{nk}w_k)^{-\theta}}. \tag{4}$$

Expressions (3) and (4) allow us to relate trade shares to trade costs and the price indices of each trading partner via the following equation:

$$\frac{X_{ni}/X_n}{X_{ii}/X_i} = \frac{\Phi_i}{\Phi_n}\tau_{ni}^{-\theta} = \left(\frac{P_i\tau_{ni}}{P_n}\right)^{-\theta}, \tag{5}$$

where $\frac{X_{ii}}{X_i}$ is country i's expenditure share on goods from country i, or its home trade share.

In this model, it is easy to show that the welfare gains from trade are essentially captured by changes in the CES price index a representative consumer faces. Because of the tight link between prices and trade shares, this model generates the following relationship between changes in price indices and changes in home trade shares as well as the elasticity parameter:

$$\frac{P_n'}{P_n} - 1 = 1 - \left(\frac{X_{nn}'/X_n'}{X_{nn}/X_n}\right)^{\frac{1}{\theta}}, \tag{6}$$

where the left-hand side can be interpreted as the percentage compensation a representative consumer in country n requires to move between two trading equilibria.

Expression (5) is not particular to the model of Eaton and Kortum (2002). Several popular models of international trade relate trade shares, prices and trade costs in the same exact manner. These models include the Armington framework of Anderson and van Wincoop (2003), as well as the monopolistic competition frameworks of firm heterogeneity by Melitz (2003) and Chaney (2008), and their homogeneous foundation introduced by Krugman (1980). Arkolakis, Costinot, and Rodriguez-Clare (2009) show how equation (6) aries within these same models.

2.1. The Elasticity of Trade

The key parameter determining trade flows (equation (5)) and welfare (equation (6)) is θ. To see the parameter's importance for trade flows, take logs of equation (5) yielding:

$$\log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right) = -\theta\log\left(\tau_{ni}\right) - \theta\log(P_i) + \theta\log(P_n). \tag{7}$$

As this expression makes clear, θ controls how a change in the bilateral trade costs, τ_{ni}, will change bilateral trade between two countries. This elasticity is important because if one wants to understand how a bilateral trade agreement will impact aggregate trade or simply understand the magnitude of the trade friction between two countries, then a stand on this elasticity is necessary. This is what we mean by the elasticity of trade.

To see the parameter's importance for welfare, it is fairly easy to demonstrate that (6) implies that θ represents the inverse of the elasticity of welfare with respect to domestic expenditure shares:

$$\log(P_n) = -\frac{1}{\theta}\log\left(\frac{X_{nn}}{X_n}\right) \tag{8}$$

Hence, decreasing the domestic expenditure share by 1% generates $(1/\theta)/100$ percent increase in consumer welfare. Thus, in order to measure the impact of trade policy on welfare, it is sufficient to obtain data on realized domestic expenditures and an estimate of the elasticity of trade.

Given θ's impact on trade flows and welfare, this elasticity is absolutely critical in any quantitative study of international trade.

3. Estimating θ: Eaton and Kortum's (2002) Approach

Equation (5) suggests that one could easily estimate θ if one had data on trade shares, aggregate prices, and trade costs. The key issue is that trade costs are not observed. In this section, we discuss how Eaton and Kortum (2002) approximate trade costs and estimate θ. Then we characterize the statistical properties of Eaton and Kortum's (2002) estimator. The key result is Proposition 1 that states their estimator is biased and overstates the elasticity of trade with a finite sample of prices. The second result is Proposition 2 which states that Eaton and Kortum's (2002) estimator is a consistent and an asymptotically unbiased estimator of the elasticity of trade.

3.1. Approximating Trade Costs

To estimate θ, the key problem is that one must disentangle trade costs from θ, when direct measures of trade costs are not observed. Eaton and Kortum (2002) propose to approximate trade costs in the following way. The idea is that by using *disaggregate* price information across countries, the maximum price difference between two countries bounds the trade cost and thus solves the indeterminacy issue.

To illustrate this argument, suppose that we observe the price of good ℓ across locations but we do not know its country of origin.[3] We know that the price of good ℓ in country n relative to country i must satisfy the following inequality:

$$\frac{p_n(\ell)}{p_i(\ell)} \leq \tau_{ni}. \tag{9}$$

That is, the relative price of good ℓ must be less than or equal to the trade friction. This inequality must hold because if it does not, then $p_n(\ell) > \tau_{ni} p_i(\ell)$ and an agent could import ℓ at a lower price. Thus, the inequality in (9) places a lower bound on the trade friction.

Improvements on this bound are possible if we observe a sample of L goods across locations. This follows by noting that the *maximum* relative price difference must satisfy the same inequality:

$$\max_{\ell \in L} \left\{ \frac{p_n(\ell)}{p_i(\ell)} \right\} \leq \tau_{ni}. \tag{10}$$

This suggests a way to exploit *disaggregate* price information across countries and arrive at an estimate of τ_{ni} by taking the maximum of relative prices over goods. Thus, Eaton and Kortum (2002) approximate τ_{ni}, in logs, by

$$\log \hat{\tau}_{ni} = \max_{\ell \in L} \left\{ \log \left(p_n(\ell) \right) - \log \left(p_i(\ell) \right) \right\}, \tag{11}$$

where the "hat" denotes the approximated value of τ_{ni}.

3.2. Estimating the Elasticity

Given the approximation of trade costs, Eaton and Kortum (2002) simply take equation (7) and run a regression. Thus, for a sample of L goods, Eaton and Kortum (2002) essentially estimate a parameter, β, using the following equations:

[3] This is the most common case, though Donaldson (2009) exploits a case where he knows the place of origin for one particular good, salt. He argues convincingly that in India salt was produced in only a few locations and exported everywhere, thus the relative price of salt across locations identifies the trade friction.

$$\log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right) = -\beta\left(\log\hat{\tau}_{ni} + \log\hat{P}_i - \log\hat{P}_n\right), \tag{12}$$

$$\text{where} \quad \log\hat{\tau}_{ni} = \max_{\ell\in L}\left\{\log p_n(\ell) - \log p_i(\ell)\right\},$$

$$\text{and} \quad \log\hat{P}_i = \frac{1}{L}\sum_{\ell=1}^{L}\log(p_i(\ell)).$$

As discussed, the second line of expression (12) approximates the trade cost. The third line approximates the aggregate price indices. The top line relates these observables to the regression they run.

To recover β, Eaton and Kortum (2002) employ a method of moments estimator by taking the average of the left-hand side of (12) divided by the average of the right-hand side of (12), with the averages across all country pairs. Mathematically their estimator is:

$$\hat{\beta} = -\frac{\sum_n\sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}{\sum_n\sum_i\left(\log\hat{\tau}_{ni} + \log\hat{P}_i - \log\hat{P}_n\right)}. \tag{13}$$

The value of β is Eaton and Kortum's (2002) estimate of the elasticity θ.[4] Throughout, we will denote by $\hat{\beta}$ the estimator defined in equation (13) to distinguish it from the value θ.

3.3. Properties of Eaton and Kortum's (2002) Estimator

Before proving properties of the estimator $\hat{\beta}$, we want to be clear about the sources of randomness in equation (12). We view the trade data on the left-hand side of (12) as being fixed. The variables on the right-hand side of (12) are random variables. That is we are treating the micro-level prices as being randomly sampled from the equilibrium distribution of prices. The parameterization of productivity draws (1), marginal cost pricing, and equation (2) are sufficient to characterize these distributions. This interpretation is consistent with the interpretation of Eaton and Kortum (2002).

Given that random variation in the sampled prices is the source of randomness, we define the following objects.

[4] To alleviate measurement error, they resort to using the second-order statistic over price differences rather than the first-order statistic. Our estimation approach is robust to the consideration of using the first- or second-order statistic.

Definition 1 *Define the following objects*

1. *Let* $\epsilon_{ni} = \theta[\log p_n - \log p_i]$ *be the log price difference of a good between country* n *and country* i*, multiplied by* θ*.*
2. *Let the vector* $\boldsymbol{S} = \{\log(T_1 w_1^{-\theta}), \ldots, \log(T_N w_N^{-\theta})\}$*.*
3. *Let the vector* $\tilde{\boldsymbol{\tau}}_{\boldsymbol{i}} = \{\theta \log(\tau_{i1}), \ldots, \theta \log(\tau_{iN})\}$*.*
4. *Let* $g(p_i; \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}})$ *be the pdf of prices of individual goods in country* i*,* $p_i \in (0, \infty)$*.*
5. *Let* $f_{\max}(\epsilon_{ni}; L, \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}})$ *be the pdf of* $\max(\epsilon_{ni})$*, given prices of a sample* $L \geq 1$ *of goods.*

The first item is simply the scaled log price difference. As we show in Appendix B, this happens to be convenient to work with as the second line in (12) can be restated in terms of scaled log price differences across locations. The second item is a vector in which each element is a function of a country's technology parameter and wage rate. The third item is a vector of trade costs country i's trading partners incur. The fourth item specifies the probability distribution of prices in each country. The fifth item specifies the probability distribution over the maximum scaled log price difference and its dependence on the sample size of prices of L goods. We derive this distribution in Appendix B.

3.4. $\hat{\beta}$ is a Biased Estimator of θ

Given these definitions, we establish two intermediate results and then state Proposition 1, which characterizes the expectation of $\hat{\beta}$, shows that the estimator is biased and discusses the reason why the bias arises.

The first intermediate result is the following:

Lemma 1 *Consider an economy of N countries with a sample of L goods' prices observed. The expected value of the maximal difference of logged prices,* $\log \hat{\tau}_{ni}$*, is strictly less than the true trade cost,*

$$\Psi_{ni}(L; \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}) \equiv \frac{1}{\theta} \int_{-\theta \log(\tau_{in})}^{\theta \log(\tau_{ni})} \epsilon_{ni} f_{\max}(\epsilon_{ni}; L, \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}) d\epsilon_{ni} \; < \; \log(\tau_{ni}). \tag{14}$$

The difference in the expected values of logged prices, $(\log \hat{P}_n - \log \hat{P}_i)$*, equals the difference in the*

price parameters, Φ, of the two countries,

$$\Omega_{ni}(\boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}) \equiv \int_0^\infty \log(p_n) g(p_n; \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}) dp_n - \int_0^\infty \log(p_i) g(p_i; \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}) dp_i = \frac{1}{\theta} \left(\log \Phi_i - \log \Phi_n \right), \tag{15}$$

with Φ_n defined in equation (3).

The key result in Lemma 1 is the strict inequality in (14). It says that Ψ_{ni}, the expected maximal logged price difference, is less than the true trade cost. Two forces drive this result. First, with a finite sample L of prices, there is positive probability that the maximal logged price difference will be less than the true trade cost. In other words, there is always a chance that the weak inequality in (10) does not bind. Second, there is zero probability that the maximal logged price difference can be larger than the true trade cost. This comes from optimality and the definition of equilibrium. These two forces imply that the expected maximal logged price difference lies strictly below the true trade cost.

The second result in Lemma 1 is that the difference in the expected log prices in expression (15) equals the difference in the aggregate price parameters defined in equation (3). This result is important because it implies that any source of bias in the estimator $\hat{\beta}$ does not arise because of systematic errors in approximating the price parameter Φ_n.

The next intermediate step computes the expected value of $1/\hat{\beta}$. This step is convenient because the inverse of $\hat{\beta}$ is linear in the random variables Lemma 1 characterizes.

Lemma 2 *Consider an economy of N countries with a sample of L goods' prices observed. The expected value of $1/\hat{\beta}$ equals:*

$$E\left(\frac{1}{\hat{\beta}}\right) = \frac{1}{\theta} \left\{ - \frac{\sum_n \sum_i (\theta \Psi_{ni}(L) - (\log \Phi_i - \log \Phi_n))}{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)} \right\} < \frac{1}{\theta}, \tag{16}$$

with

$$1 > \left\{ - \frac{\sum_n \sum_i (\theta \Psi_{ni}(L) - (\log \Phi_i - \log \Phi_n))}{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)} \right\} > 0. \tag{17}$$

This results says that the expected value of the inverse of $\hat{\beta}$ equals the inverse of the elasticity multiplied by the bracketed term of (17). The bracketed term is the expected maximal logged price difference minus the difference in expected logged prices, both scaled by theta, and divided by trade data. This term is strictly less than one. The reason is because Ψ_{ni} does

not equal the trade cost established in Lemma 1. If Ψ_{ni} did equal the trade cost, then the bracketed term would equal one and the expected value of the inverse of $\hat{\beta}$ would be equal to the inverse of θ. This can be seen by examining the relation between Φ's and aggregate prices P's in (3), equation (7), and how they relate to (17).

Inverting (16) and then applying Jensen's inequality establishes the main result: that Eaton and Kortum's (2002) estimator is biased above the true value of θ.

Proposition 1 *Consider an economy of N countries with a sample of L goods' prices observed. The expected value of $\hat{\beta}$ is*

$$E\left(\hat{\beta}\right) \;\geq\; \theta \times \left\{ -\frac{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}{\sum_n \sum_i (\theta \Psi_{ni}(L) - (\log \Phi_i - \log \Phi_n))} \right\} \;>\; \theta. \tag{18}$$

The proposition establishes that the estimator $\hat{\beta}$ provides estimates that exceed the true value of the elasticity θ. The weak inequality in (18) comes from applying Jensen's inequality to the strictly convex function of $\hat{\beta}$, $1/\hat{\beta}$. The strict inequality follows from Lemma 1 which argued that the expected maximal logged price difference is strictly less than the true trade cost. Thus, the bracketed term in expression (18) is always greater than one and the elasticity of trade is always overestimated.

3.5. Consistency and Asymptotic Bias

While the estimator $\hat{\beta}$ is biased in a finite sample, the asymptotic properties of Eaton and Kortum's (2002) estimator are worth understanding. Proposition 2 summarizes the result.

Proposition 2 *Consider an economy of N countries. The maximal log price difference is a consistent estimator of the trade cost,*

$$\operatorname*{plim}_{L\to\infty} \; \Psi_{ni}(L; \boldsymbol{S}, \tilde{\boldsymbol{\tau_n}}, \tilde{\boldsymbol{\tau_i}}) \;=\; \log(\tau_{ni}). \tag{19}$$

The estimator $\hat{\beta}$ is a consistent estimator of θ,

$$\operatorname*{plim}_{L\to\infty} \; \hat{\beta}\left(L; \boldsymbol{S}, \tilde{\boldsymbol{\tau_n}}, \tilde{\boldsymbol{\tau_i}}, \sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)\right) = \theta, \tag{20}$$

and the asymptotic bias of $\hat{\beta}$ is zero,

$$\lim_{L\to\infty} E\left[\hat{\beta}\left(L; \boldsymbol{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)\right)\right] - \theta = 0. \tag{21}$$

There are three elements to Proposition 2 with each building on the previous one. The first statement says that the probability limit of the maximal log price difference equals the true trade cost between two countries. Intuitively, this says that as the sample size becomes large the probability that the weak inequality in (10) does not bind becomes vanishingly small.

The second statement says that the estimator $\hat{\beta}$ converges in probability to the elasticity of trade, i.e. $\hat{\beta}$ is a consistent estimator of θ. The reasons are the following. Because the maximal price difference converges in probability to the true trade cost, and the difference in averages of log prices converges in probability to the difference in price parameters, the probability limit of $1/\hat{\beta}$ converges to $1/\theta$. Since $1/\hat{\beta}$ is a continuous function of $\hat{\beta}$ (with $\hat{\beta} > 0$), the probability limit of $\hat{\beta}$ must converge to θ because of the preservation of convergence for continuous functions (see Hayashi (2000)).

The third statement says that in the limit the bias is eliminated. This follows immediately from the previous argument that the estimator $\hat{\beta}$ is a consistent estimator of θ (again see Hayashi (2000)).

The results in Proposition 2 are important for two reasons. First, they suggest that with enough data Eaton and Kortum's (2002) estimator provides informative estimates of the elasticity of trade. However, as we will show in the next section, monte-carlo exercises suggest the data requirements are extreme. Second, because Eaton and Kortum's (2002) estimator has desirable asymptotic properties, it will provide the basis for our simulation based estimator that we develop in Section 5.

4. How Large is the Bias? How Much Data is Needed?

Proposition 1 shows Eaton and Kortum's (2002) estimator is biased in a finite sample. Many estimators have this property, which raises the question: How large is the bias? Furthermore, even if the magnitude of the bias is large, perhaps only moderate increases in the sample size are sufficient to eliminate the bias (in practical terms). The natural question is then: How much data is needed to achieve that?

To answer these questions, we perform monte carlo experiments where we simulate trade flows and samples of micro-level prices under a known θ. Then we apply Eaton and Ko-

rtum's (2002) estimator to the artificial data. We employ the same simulation procedure described in Steps 1-3 in Section 5.2, and all parameters (except for θ) are estimated using the trade data from Eaton and Kortum (2002). We set the true value of θ equal to 8.28 which is Eaton and Kortum's (2002) estimate when employing the approach described above. The sample size of prices is set so $L = 50$ which is the number of prices Eaton and Kortum (2002) had access to in their data set.

Table 1 presents the findings. The columns of Table 1 present the mean and median estimates of β over 100 simulations. The rows present different estimation approaches such as method of moments and least squares with the constant suppressed. Also reported is the true average trade costs and the estimated average trade cost using maximal log price differences.

Table 1: Monte Carlo Results, True θ = 8.28

Approach	Mean Estimate of θ (S.E.)	Median Estimate of θ
EK(2002) Estimator	12.5 (0.06)	12.5
Least Squares	11.8 (0.06)	11.8
True Mean τ = 1.79	Estimated Mean τ = 1.48	

Note: S.E. is the standard error of the mean. In each simulation there are 19 countries and 500,000 goods. Only 50 realized prices are randomly sampled and used to estimate θ. 100 simulations performed.

The first row in Table 1 shows that the estimates using Eaton and Kortum's (2002) approach are larger than the true θ of 8.28 which is consistent with Proposition 1. The key source of bias in Proposition 1 was that the estimates of the trade costs were biased downward as Lemma 1 argued. The final row in Table 1 illustrates that the estimated trade costs are below the true trade costs, where the latter correspond to an economy characterized by a true elasticity of trade among 19 OECD countries of 8.28.

The second row in Table 1 reports results using a least squares estimator with the constant suppressed rather than the method of moments estimator.[5] Similar to the method of moments estimates, the least squares estimates are substantially larger than the true value of θ. This is important because it suggests that the key problem with Eaton and Kortum's (2002)

[5]Including a constant in least squares results in slope coefficient that underestimate the true elasticity. This result is symptomatic of an "errors in variables" problem.

approach is not the method of moment estimator per se, but instead the poor approximation of the trade cots.

The final point to notice is that the magnitude of the bias is substantial. The underlying θ was set equal to 8.28 and the estimates in the simulation are between 11.8 and 12.5. Equation (8) can be used to formulate the welfare cost of the bias. It suggests that the welfare gains from trade will be underestimated by 50 percent as a result of the bias.

How much data is needed to eliminate the bias? Table 2 provides a quantitative answer. It preforms the same monte carlo experiments described above, as the sample size of micro-level prices varies.

Table 2: Increasing the Sample of Prices Reduces the Bias, True θ = 8.28

Sample Size of Prices	Mean θ (S.E.)	Median θ	Mean τ
50	12.51 (0.06)	12.50	1.48
500	9.34 (0.02)	9.32	1.68
5,000	8.43 (0.01)	8.43	1.77
50,000	8.30 (0.002)	8.30	1.78

Note: S.E. is the standard error of the mean. In each simulation there are 19 countries and 500,000 goods. The results reported use least squares with the constant suppressed. 100 simulations performed. True Mean τ = 1.79

As the sample size becomes larger, Table 2 shows that the estimate of θ becomes less biased and begins to approach the true value of θ. The final column shows how the reduction in the bias coincides with the estimates of the trade costs becoming less biased. This is consistent with the arguments of Proposition 2 which describe the asymptotic properties of this estimator.

We should note that the rate of convergence is extremely slow; even with a sample size of 5,000 the estimate of β is meaningfully larger than the value generating the data. Only when 50,000 prices are sampled does the estimate approach the true value. This exercise allows us to conclude that the data requirements to minimize the bias in estimates of the elasticity of trade (in practice) are extreme. This motivates our alternative estimation strategy in the next section.

5. A New Approach To Estimating θ

In this section we suggest a new approach to estimating θ and discuss its performance on simulated data.

5.1. The Idea

Our idea is to exploit the structure of the model in the following way. First, in Section 5.2 we show how to recover all parameters necessary to simulate the model up to the unknown scalar θ from trade data only. These parameters are the vector $\mathbf{S}$ and scaled trade costs, $\theta \log(\tau_{ni})$. Given these values we are able to simulate moments from the model as a function of θ.

Second, Lemma 1 and Lemma 2 actually suggest which moments are informative. Inspection of the integral (14) and the density f_{max} in (b.28) shows that the expected maximal log price difference monotonically varies with θ and linearly with $1/\theta$. This follows because of the previous point—the vector $\mathbf{S}$ and scaled log trade costs $\tilde{\boldsymbol{\tau}}$ are pinned down by trade data and these values completely determine all parameters in the integral (14) except the value $1/\theta$ lying outside the integral. Similarly, the integral (15) is completely determined by these values and scaled in the same way by $1/\theta$ as (14) is.

These observations have the following implication. While the maximum log price difference is biased below the true trade cost, if θ is large, then the value of the maximum log price difference will be small. Similarly, if θ is small, then the value of the maximum log price difference will be large. A large or small maximum log price difference will result in a small or large estimate of $\hat{\beta}$. This suggests that the estimator $\hat{\beta}$ will vary monotonically with the true value of θ. Furthermore, this suggests that $\hat{\beta}$ is an informative moment with regard to θ.[6]

Figure 1 quantitatively illustrates this intuition by plotting $\beta(\theta)$ from simulations as we varied θ. It is clear that β is a biased estimator because these values do not lie on the 45^o line. However, β varies near linearly with θ. These observations suggest an estimation procedure that matches the data moment β to the moment $\beta(\theta)$ implied by the simulated model under a known θ.[7] Because of the monotonicity implied from our arguments, the known θ must be the unique value that satisfies the moment condition specified.

[6] While we have not shown this formally, it can be shown that the expected value of $1/\hat{\beta}$ is proportional to $1/\theta$. Modulo effects from Jensen's inequality, this suggests that $\hat{\beta}$ is proportional to θ. Figure 1 confirms this.

[7] Another reason for using the moment β is that it is a consistent estimator of θ as argued in Proposition 2.

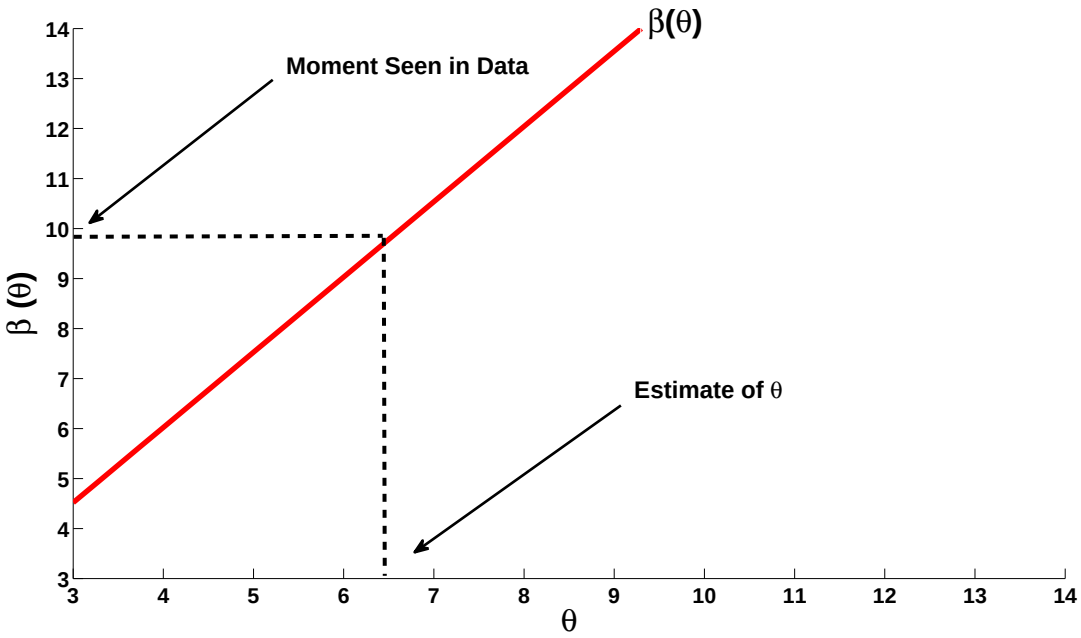


Figure 1: Schematic of Estimation Approach

5.2. Simulation Approach

We show how to recover all parameters of interest up to the unknown scalar θ from trade data only and then we describe our simulation approach. This provides the foundations for the simulated method of moments estimator we propose.

Step 1.—We estimate parameters for the country-specific productivity distributions and trade costs from bilateral trade flow data. We perform this step by following Eaton and Kortum (2002) and Waugh (2009) and deriving the following gravity equation from equation (4) by dividing the bilateral trade share by the importing country's home trade share,

$$\log\left(\frac{X_{ni}/X_n}{X_{nn}/X_n}\right) = S_i - S_n - \theta \log \tau_{ni}, \tag{22}$$

in which S_i is defined as $\log\left[T_i w_i^{-\theta}\right]$ and is the same value in the parameter vector $\mathbf{S}$ in Definition 1. Note that this is a different equation than that used to estimate θ in (5) which is derived by dividing the bilateral trade share by the exporting country's home trade share. S_is are recovered as the coefficients on country-specific dummy variables given the restrictions on how trade costs can covary across countries. Following the arguments of Waugh (2009), trade costs take the following functional form:

$$\log(\tau_{ni}) = d_k + b_{ni} + ex_i + \nu_{ni}. \tag{23}$$

Here, trade costs are a logarithmic function of distance, where d_k with $k = 1, 2, ..., 6$ is the effect of distance between country i and n lying in the kth distance intervals.[8] b_{ni} is the effect of a shared border in which $b_{ni} = 1$, if country i and n share a border and zero otherwise. The term ex_i is an exporter fixed effect and allows for the trade cost to vary in level depending upon the exporter. We assume ν_{ni} reflects other factors and is orthogonal to the regressors and normally distributed with mean zero and standard deviation σ_ν. We use least squares to estimate equations (22) and (23).

Step 2.—The parameters from the first stage gravity regression provide us with all the parameters necessary to simulate trade flows and micro-level prices up to a constant, θ.

The relationship is obvious in the estimation of trade barriers since $\log(\tau_{ni})$ is scaled by θ in (22). To see that we can simulate micro-level prices as a function of θ only, notice that for any good j, $p_{ni}(j) = \tau_{ni} w_i / z_i(j)$. Thus, rather than simulating productivities, it is sufficient to simulate the inverse of marginal costs of production $u_i(j) = z_i(j)/w_i$. In in Appendix B.1 we show that u_i is distributed according to:

$$M_i(u_i) = \exp(-S_i u_i^{-\theta}), \quad \text{with} \quad S_i = \log\left[T_i w_i^{-\theta}\right]. \tag{24}$$

Thus, having obtained the coefficients S_i from the first stage gravity regression, we can simulate the inverse of marginal costs and prices.

To simulate the model, we assumed there is a large number (150,000) of potentially tradable goods. In Section 8.1 we discuss how we made this choice and the motivation behind it. For each country, the inverse marginal costs are drawn from the country-specific distribution (24) and assigned to each good. Then, for each importing country and each good, the low-cost supplier across countries is found, realized prices are recorded, and aggregate bilateral trade shares are computed.

Step 3.—From the realized prices, a subset of goods common to all countries is defined and the subsample of prices is recorded, i.e. we are acting as if we were collecting prices for the international organization that collects the data. We added disturbances to the predicted trade shares with the disturbances drawn from a mean zero normal distribution with the standard deviation set equal to the standard deviation of the residuals from Step 1.

These steps then provide us with an artificial data set of micro-level prices and trade shares mimicking their analogs in the data. Given this artificial data set, we can then compute

[8] Intervals are in miles: $[0, 375)$; $[375, 750)$; $[750, 1500)$; $[1500, 3000)$; $[3000, 6000)$; and $[6000, \text{maximum}]$. An alternative to specifying a trade cost function is to recover scaled trade costs as a residual using equation (5), trade data, and measures of aggregate prices as in Waugh (2010).

moments—as a function of θ— to compare to moments in the data.

5.3. Estimation

Below we describe the moments we try to match and the formalities of our estimation procedure.

5.3.1. Moments

In this section, we define the moments of interest. We perform two estimations: one overidentified procedure with two moments and an exactly identified procedure with one moment.

Define β_k as Eaton and Kortum's (2002) method of moment estimator defined in (12) using the kth order statistic over micro-level price differences. Then the moments we are interested in are:

$$\beta_k = -\frac{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}{\sum_n \sum_i \left(\log \hat{\tau}_{ni}^k + \log \hat{P}_i - \log \hat{P}_n\right)}, \quad k = 1, 2 \tag{25}$$

where $\hat{\tau}_{ni}^k$ is computed as the kth order statistic over micro-level price differences between country n and i. In the exactly identified estimation we use β_1 as the only moment.

We denote the simulated moments by $\beta_1(\theta, u_s)$ and $\beta_2(\theta, u_s)$, which come from the analogous formula as in (25) and are estimated from artificial data generated as described in **Steps 1-3** above. Note that these moments are a function of θ and depend upon a vector of random variables u_s associated with a particular simulation s. There are three components to this vector. First, there are the random productivity draws for production technologies for each good and each country. The second component is the set of goods sampled from all countries. The third component mimics the residuals ν_{ni} from equation (22), which are described in Section 5.2.

Stacking our data moments and averaged simulation moments gives us the following zero function:

$$y(\theta) = \begin{bmatrix} \beta_1 - \frac{1}{S}\sum_{s=1}^{S} \beta_1(\theta, u_s) \\ \\ \beta_2 - \frac{1}{S}\sum_{s=1}^{S} \beta_2(\theta, u_s) \end{bmatrix}. \tag{26}$$

5.3.2. Estimation Procedure

We base our estimation procedure on the moment condition:

$$E\left[y(\theta_o)\right] = 0,$$

where θ_o is the true value of θ. Thus our simulated method of moments estimator is

$$\hat{\theta} = \arg\min_{\theta} \left[y(\theta)' \, \mathbf{W} \, y(\theta)\right], \tag{27}$$

where $\mathbf{W}$ is a 2×2 weighting matrix which we discuss below.

The idea behind this moment condition is that though β_1 and β_2 will be biased away from θ, the moments $\beta_1(\theta, u_s)$ and $\beta_2(\theta, u_s)$ will be biased by the same amount when evaluated at θ_o, in expectation. Viewed in this language, our moment condition is closely related to the estimation of bias functions discussed in MacKinnon and Smith (1998) and to indirect inference as discussed in Smith (2008). The key issue in MacKinnon and Smith (1998) is how the bias function behaves. Figure 1 shows that the bias function is basically linear in the parameter of interest and thus it is well behaved.

For the weighting matrix, we use the optimal weighting matrix suggested by Gouriéroux and Monfort (1996) for simulated method of moments estimators. Because the weighting matrix depends upon our estimate of θ, we used a standard iterative procedure outlined in the next steps.

Step 4.— We make an initial guess of the weighting matrix $\mathbf{W}^0$ and solve for $\hat{\theta}^0$. Then we simulate the model to estimate the variance-covariance matrix to generate a new estimate of the weighting matrix with which we then solve for a new $\hat{\theta}^1$. We perform this iterative procedure until our estimates of the weighting matrix and $\hat{\theta}$ converge. Explicit consideration of simulation error are taken into account because we utilize the weighting matrix suggested by Gouriéroux and Monfort (1996).

Step 5.— We compute standard errors using a bootstrap technique. We computed residuals υ_{ni} implied by the estimator in (25) and the fitted values, we resampled the residuals υ_{ni} with replacement and generated a new set of data using the fitted values. Using the data constructed from each resampling b, we computed new estimates β_1^b and β_2^b.

For each bootstrap b, we replaces the moments β_1 and β_2 with bootstrap generated moments β_1^b and β_2^b. To account for simulation error, a new seed is set to generate a new set of model generated moments: $\frac{1}{S}\sum_{s=1}^{S} \beta_1(\theta, u_s)^b$ and $\frac{1}{S}\sum_{s=1}^{S} \beta_2(\theta, u_s)^b$. Defining $y^b(\theta)$ as the difference

in moments for each b as in (26), we solve for:

$$\hat{\theta}^b = \arg\min_{\theta} \left[y^b(\theta)' \, \mathbf{W} \, y^b(\theta) \right]. \tag{28}$$

We repeat this exercise 100 times and compute the estimated standard error of our estimate of $\hat{\theta}$ as:

$$\text{S.E.}(\hat{\theta}) = \left[\frac{1}{100} \sum_{b=1}^{100} (\hat{\theta}^b - \hat{\theta})(\hat{\theta}^b - \hat{\theta})' \right]^{\frac{1}{2}} \tag{29}$$

This procedure for constructing standard errors is similar in spirit to the approach employed in Eaton, Kortum, and Kramarz (2008) who use a simulated method of moments estimator to estimate the parameters of a similar trade model from the performance of French exporters.

5.4. Performance on Simulated Data

In this section, we evaluate the performance of our estimation approach using simulated data when we know the true value of θ.

Table 3 presents the results from the following exercise. We generated an artificial data set with true value of θ equal to 8.28 and 4.00, then applied our estimation routine, and repeated this exercise 100 times. The sequence of artificial data is the same for both the overidentified case and exactly identified case to facilitate comparisons across estimators.

The first row presents the average value of our simulated method of moments estimate which is 8.29 with a standard error of 0.03. For all practical purposes, the estimation routine recovered the true value of θ generating the data. To emphasize the performance of our estimator, the next two rows of Table 3 present the approach of Eaton and Kortum (2002) (which also correspond to the moments used). Though not surprising given the discussion above, this approach generates estimates of θ that are significantly (in both their statistical and economic meaning) higher than the true value of θ of 8.28.

The final two rows present the exactly identified case when we use only one moment to estimate θ. In this case we used β_1. Similarly to the over-identified case, the average value of our simulated method of moments estimate is 8.24 with a standard error of 0.04. Again, this is effectively the true value of θ.

The second column reports the results when the true value of θ is set equal to 4.00. The estimates using our estimator are 3.99 and 3.98 in the overidentified and exactly identified case. Similar to the previous results, these values are effectively the true value of θ. Furthermore,

Table 3: Estimation Results With Artificial Data

Estimation Approach	**True θ = 8.28**	**True θ = 4.00**
Overidentified	Mean Estimate of θ (S.E.)	Mean Estimate of θ (S.E.)
SMM	8.29 (0.03)	3.99 (0.02)
Moment, β_1	12.47 (0.05)	6.03 (0.03)
Moment, β_2	15.20 (0.05)	7.34 (0.03)
Exactly Identified		
SMM	8.24 (0.04)	3.98 (0.02)
Moment, β_1	12.47 (0.05)	6.03 (0.03)

Note: In each simulation there are 19 countries, 150,000 goods and 100 simulations performed. The sequence of artificial data is the same for both the overidentified case and exactly identified case.

the alternative approaches which correspond with the moments we used in our estimation are biased away from the true value of θ.

We also compared our estimation approach relative to a naive statistical approach to bias reduction. By naive we mean approaches that do not depend on the model's explicit distributional assumptions. Robson and Whitlock (1964) proposes a way to reduce the bias when estimating the truncation point of a distribution. This problem is analogous to estimating the trade cost from price differences. This can be seen by inspecting the integral in (14) of Lemma 1. Robson and Whitlock's (1964) approach would suggest (in our notation) an estimator of the trade cost of $2\hat{\tau}^1_{ni} - \hat{\tau}^2_{ni}$, or two times the first-order statistic minus the second order statistic. This makes intuitive sense because it increases the first-order statistic by the difference between the first and second-order statistic. They show that this estimator is as efficient as the first-order statistic but with less bias.[9]

We followed their approach to approximate the trade friction and then used it as an input into the simple method of moments estimator. We compared the results from this estimation procedure to the results obtained using our SMM estimator. Table 4 presents the results. The second row reports the results when using Robson and Whitlock's (1964) approach to reduce the bias in the estimator of the trade friction. This approach reduces the bias relative to using the first-order statistic (Eaton and Kortum's (2002) approach) reported in the third row. It is not, however, a complete solution as the estimates are still meaningfully higher than both

[9]Robson and Whitlock (1964) provide more general refinements using inner order statistics, but methods using inner order statistics will have very low efficiency. Cooke (1979) provides an alternative bias reduction technique but only considers cases when the sample size (L in our notation) is large.

the true value of θ and the estimates from our estimation approach. This suggests that exploiting the structure of the model has content because it outperforms a naive statistical procedure.

Table 4: Comparison to Alternative Statistical Approaches to Bias Reduction

	True θ = 8.28	**True θ = 4.00**
Estimation Approach	Mean Estimate of θ (S.E.)	Mean Estimate of θ (S.E.)
SMM	8.29 (0.03)	3.99 (0.02)
Robson and Whitlock (1964)	10.54 (0.07)	5.11 (0.03)
Moment, β_1	12.47 (0.05)	6.03 (0.03)

Note: In each simulation there are 19 countries, 150,000 goods and 100 simulations performed. The sequence of artificial data is the same for all cases.

Overall, we view these results as evidence supporting our estimation approach and empirical estimates of θ presented in Section 6 below.

6. Empirical Results

In this section, we apply our estimation strategy described in section 5 to several different data sets. The key finding of this section is that our estimation approach yields an estimate around 4 in contrast to previous estimation strategies which yield estimates around 8.

6.1. Baseline Results Using New ICP 2005 Data

6.1.1. New ICP 2005 Data

Our sample contains 123 countries. We use trade flows and production data for the year 2004 to construct trade shares. The price data used to compute aggregate price indices and proxies for trade costs comes from basic-heading level data from the 2005 round of the International Comparison Programme (ICP). The ICP collects price data on goods with identical characteristics across retail locations in the participating countries during the 2003-2005 period.[10] The basic-heading level represents a narrowly-defined group of goods for which expenditure data are available. In the data set there are a total of 129 basic headings, and

[10]The ICP Methodological Handbook is available at http://go.worldbank.org/MW520NNFK0.

we reduce the sample to 62 categories based on their correspondence with the trade data employed. Appendix A provides more details.

On its own, this data set provides two contributions to the existing literature. First, because this is the latest round of the ICP, the measurement issues are less severe than in previous rounds. Furthermore, this data set has a very extensive coverage as it includes as many as 123 developing and developed countries that account for 98 percent of world output.

The ICP provides a common list of "representative" goods whose prices are to be randomly sampled in each country over a certain period of time. A good is representative for a country if it comprises a significant share of a typical consumer's bundle there. Thus, the ICP samples the prices of a common basket of goods across countries, where the goods have been pre-selected due to their highly informative content for the purpose of international comparisons.

Eaton and Kortum's (2002) model gives a natural common basket of goods to be priced across countries. In this model, the agents in all countries consume all goods that lie within a fixed interval, $[0, 1]$. Thus, we consider this common list in the simulated model and randomly sample the prices of its goods across countries, in order to approximate trade barriers, much like it is done in the ICP data.

6.1.2. Results—New ICP 2005 Data

Table 5 presents the results. The first row simply reports the moments that our estimation procedure targets. As discussed, these values correspond with Eaton and Kortum's (2002) estimate of θ.

Table 5: Estimation Results With 2005 ICP Data

	Estimate of θ (S.E.)	β_1	β_2
Data Moments	—	7.75	9.61
Exactly Identified Case	4.12 (???)	7.75	—
Overidentified Case	4.06 (0.01)	7.65	9.62

The second row reports the results for exactly identified estimation and the underlying moment used is β_1. In this instance our estimate of θ is 4.12, more or less than half what Eaton and Kortum's (2002) estimate of θ would imply.

The third row reports the results for the overidentified estimation. The estimate of θ is 4.06 — almost the same as in our exactly identified estimation and again more or less than half of Eaton and Kortum's (2002) estimate. The second and third columns report the resulting moments from the estimation routine which are close to the data moments targeted given that only one parameter was used to match two moments.

6.2. Estimates Using Eaton and Kortum's (2002) Data

In this section, we apply our estimation strategy to the same data used in Eaton and Kortum (2002) as another check of our estimation procedure. Their data set consists of bilateral trade data for 19 OECD countries in 1990 and 50 prices of manufactured goods for all countries. The prices come from an earlier round of the ICP, which considered only OECD countries. Similar to our data, the price data is at the basic heading level and is for goods with identical characteristics across retail locations in the participating countries.

6.2.1. Results—Eaton and Kortum's (2002) Data

Table 6 presents the results. The first row simply reports the moments that our estimation procedure targets. The entry in the third column corresponds with β_2 which is Eaton and Kortum's (2002) baseline estimate of θ.

Table 6: Estimation Results With EK (2002) Data

	Estimate of θ (S.E.)	β_1	β_2
Data Moments	—	5.93	8.28
Exactly Identified Case	3.93 (0.09)	5.93	—
Overidentified Case	4.42 (0.06)	6.64	8.10

The second row reports the results for exactly identified estimation and the underlying moment used is β_1. In this instance our estimate of θ is 3.93 which is again more or less than half what the Eaton and Kortum's (2002) estimate of θ would imply. The standard error of our estimate is fairly tight.

The third row reports the results for the overidentified estimation. Here our estimate of θ is 4.42. Again, this is substantially below Eaton and Kortum's (2002) estimate. Unlike our

results in Table 5 with newer data, the overidentified case seems to be giving a different value than the exactly identified case. This contrasts with the monte carlo evidence that suggests the estimation procedure should not deliver very different estimates. Furthermore, comparing the data moments in the top row versus the implied moments in the second and third columns of the third row suggest the estimation routine is facing challenges fitting the observed moments. We view this as pointing towards a problem with measurement error in the data as Eaton and Kortum (2002) suggested.

6.3. Discussion

Our estimation results compare favorably with alternative estimates of θ which *do not* use the max over price data to approximate trade costs. For example, estimates of θ using firm level data as in Bernard, Eaton, Jensen, and Kortum (2003) and Eaton, Kortum, and Kramarz (2008) are in the range of 3.6 to 4.8—exactly in the range of values we find. Eaton and Kortum (2002) provide an alternative estimate of θ using wage data and find a value of 3.6. Burstein and Vogel (2009) estimate θ matching moments regarding the skill intensity of trade and find a value of 4. Simonovska (2010) uses a non-homothetic model of trade featuring variable mark-ups and calibrates θ to a level of 3.8 which allows her model to match average mark-ups in OECD countries.

Donaldson (2009) estimates θ as well and his approach is illuminating relative to the issues we have raised. His strategy to approximating trade costs is to study differences in the price of salt across locations in India. In principal, his approach is subject to our critique as well, i.e. how could price differences in one good be informative about trade frictions? However, he argues convincingly that in India salt was produced in only a few locations and exported everywhere. Thus by examining salt, Donaldson (2009) has found a "binding" good. Using this approach, he finds estimates in the range of 3.8-5.2, again consistent with the range of our estimates of θ.

Finally, it should be noted that the elasticity of trade, θ, is closely related to the elasticity of substitution between foreign and domestic goods, the Armington elasticity, which determines the behavior between trade flows and relative prices across a large class of models. Recently, Ruhl (2008) presents a comprehensive discussion of the puzzle regarding this elasticity. In particular, he argues that international real business cycle models need low elasticities, in the range of 1 to 2, to match the quarterly fluctuations in trade balances and the terms of trade, but static applied general equilibrium models need high elasticities, between 10 and 15, to account for the growth in trade following trade liberalization. Using very disaggregate data, Romalis (2007), Broda and Weinstein (2006), and Hummels (2001) pro-

vide estimates for the Armington elasticity parameter across a large number of industries. Romalis's (2007) estimates range between 4-13, Hummels's (2001) estimates range between 3-8, while the most comprehensive work of Broda and Weinstein (2006), who provide tens of thousands of elasticities using 10-digit HS US data, results in a median value of 3.10.

Given our estimates of θ, it is straightforward to back out the Armington elasticity ρ within the context of the model of Anderson and van Wincoop (2004), where $\rho = \theta + 1$. Using our estimates of the elasticity of trade, the implied Armington elasticity ranges between 4.93-5.49. This utility parameter also appears in the heterogeneous firm framework of Melitz (2003) parameterized by Chaney (2008). Together with the elasticity of trade, θ, the utility parameter governs the distribution of firm sales arising from the model, which has Pareto tales with a slope given by $\theta/(\rho - 1)$. Luttmer (2007) discusses firm-level evidence that this slope takes on the value of 1.65, which given our estimates of θ, provides the range of $3.38 - 3.72$ for ρ. Hence, the Armington elasticity implied by our estimates ranges between $3.38 - 5.49$, which falls within the low end of the ranges of estimates of existing studies. Thus, our results close the gap between (implied) estimates of the elasticity of trade stemming from a rich demand structure and the ones obtained from the general equilibrium structure of a Ricardian model of trade.

7. Why Estimates of θ Matter: The Welfare Gains From Trade

The elasticity parameter θ is key in measuring the welfare gains from trade across all models outlined in this paper. In section 2.1, we argued that θ^{-1} represents the elasticity of welfare with respect to trade. Recall expression (6), reproduced below for convenience:

$$\log(P_n) = -\frac{1}{\theta} \log\left(\frac{X_{nn}}{X_n}\right)$$

Consider a trade liberalization episode that results in a 1 percent fall in domestic expenditure share, generating $(1/\theta)/100$ percent increase in consumer welfare. Using the estimates for θ obtained from the original procedure and the improved simulated method of moments procedure, roughly 8 and 4, respectively, the welfare gains from trade would be mis-measured by a hundred percent. Namely, an estimate for θ of 8 would generate 0.125 percent welfare increase for a percent fall in the domestic share, while an estimate of 4 suggests a 0.25 percent welfare gain from trade, twice as high as the original calculation. These differences illustrate the importance to obtain better estimates of the elasticity of trade.

8. Robustness

8.1. The Number of Goods

The estimation routine requires us to take a stand on the number of goods in the economy. We argue that the appropriate way to view this issue is simply how to numerically approximate the infinite number of goods in model. Thus the number of goods chosen should be judged on the accuracy of its approximation relative to it's computational cost. It should not be judged on the basis of how many goods actually exist in the "real world" because this value is impossible to know or discipline.

To understand our argument recall that our estimation routine is based on a moment condition that compares a biased estimate from the data versus a biased estimate using artificial data. In Section 3.4 we argued that bias depends largely on the expected value of the max over a finite sample of price differences, i.e. the integral of the left hand side of equation (14). Thus when we compute the biased estimate using artificial data, we are effectively computing this integral via simulation.[11] This suggests that the number of goods should be chosen in a way that delivers an accurate approximation. Furthermore, a way to judge if the number of goods selected delivers an appropriate approximation is to increase the number of goods until the estimate does not change too much.

Table 7 performs this analysis. It shows how our estimate of θ varies as the number of goods in the economy changes using the Eaton and Kortum (2002) data and 2004 ICP data. For the Eaton and Kortum (2002) data, notice that our estimates are relatively similar across all the different numbers of goods employed ranging from 4.14 to 3.93. Moreover, the estimates are effectively the same after the number of goods is above 100,000 suggesting that this is a reasonable starting point.

The results with the 2004 ICP data vary more depending upon the number of goods used. The reason is the much larger number of countries. With more countries there is more competition to be the low cost supplier for any one good, i.e. the low cost good is selected from a wider set. This changes the distribution of equilibrium prices and maximal price differences in a way that requires a larger number of goods to accurately compute the integrals in (14). While the change from 4.22 to 4.12 when going from 100,000 and 150,000 goods is numerically large, computational costs forces our hand to settle on 150,000 goods as as the number of goods in the economy.[12]

[11] An alternative estimation strategy would be to use different numerical methods to compute the integrals (14) and then adjust the Eaton and Kortum (2002) estimator given this value.

[12] The reason is that 150,000 goods is near the maximum amount of goods feasible while still being able to

Table 7: Results with Different # of Goods

Number of Goods	5,000	25,000	100,000	150,000*
EK (2002) Data, Exactly Identified Case, $\hat{\theta}$	4.14	3.99	3.93	3.93
Fraction of Wrong Zeros	0.10	0.03	0.005	0.003
Fraction of Correct Zeros	—	—	—	—
2004 ICP Data, Exactly Identified Case, $\hat{\theta}$	5.54	4.67	4.22	4.12
Fraction of Wrong Zeros	0.46	0.31	0.21	0.18
Fraction of Correct Zeros	0.85	0.72	0.55	0.50

Table 7 also reports a side effect of using a low number of goods — zero trade flows between countries in the model in places where there are observed trade flows in the data. Table 7 reports the fraction of zeros that the model produces in instances where their are positive trade flows observed in the data. With only only 5,000 goods using the 2004 ICP data set, almost half of the instances where trade flows are observed in the data the model generates a zero. While not as severe, 10 percent of positive trade flows are assigned zeros with the Eaton and Kortum (2002) data. Results of this nature suggests increasing the number of goods to minimize the number of wrong zeros as well.

8.2. Country-Specific Taxes and Distribution Costs

The price data used in our estimation is collected at the retail level. As such, it necessarily reflects local (distribution) costs and sales taxes. It turns out that these market frictions do not affect our estimates of the elasticity parameter, for as long as they are country- but not good-specific. The basic way to see this is to note that any country specific effects will cancel out in in the denominator of (13). This is another reason for using β as a moment in our estimation routine rather than some other moment.

execute the simulation routine in parallel on a multi-core machine which allows a speed up of just under a factor of 8.

8.3. Measurement Error

Measurement error in the price data is a concern. Error of this nature may artificially generate *larger* maximal price differences than implied by the underlying model. This would result in estimates of θ that are biased downwards. Because measurement error can potentially affect our estimates, we are addressing this concern in the near future.

8.4. Mark-ups

The price data used in our estimation likely reflects retail mark-ups, which are not only country-, but also retailer-specific. In order to check whether such variable mark-ups affect our results, we make use of a richer price dataset. In particular, we obtain price data provided by the EIU Worldwide Cost of Living Survey, which spans 77 of the original 123 countries we consider. More importantly, the data comprise of 111 tradable goods per country, and the price of each product is recorded once in a supermarket and once in a mid-price store. We repeat our exercise by first using the prices of items collected in mid-price stores, which appear to be cheaper on average, and then the prices found in supermarkets or chain stores. We postpone the results until section 8.5.1 below, since the level of detail in the EIU data allows us to also address aggregation bias, described there.

8.5. Aggregation

8.5.1. Data Approach

The basic-heading data employed in our analysis constitute fairly disaggregate price data, however, the data do not represent individual good price observations. For example, a price observation titled "rice" contains the average price across different types of rice sampled, for example basmati rice, wild rice, whole-grain white rice, etc. Since estimating the elasticity parameter necessitates arriving at a measure of trade barriers via the maximum price difference across observed good prices for each pair of countries, the elasticity estimate may be biased upwards due to a downward bias in trade barrier estimates arising from aggregation. To see this, suppose that for importer n, basmati rice is the binding good that allows us to estimate the trade cost of importing from country i. In the ICP data however, we only observe the average price of rice which reflects prices of multiple varieties of rice. Hence, the difference between the average prices of rice between the two countries is necessarily smaller than the price difference of basmati rice, should the remaining types of rice be more equally priced across the two countries. In this case, trade barriers are underestimated and

consequently the elasticity of trade is biased upwards.[13]

In order to alleviate the aggregation problem, we present estimates of the elasticity parameter stemming from the good-level price dataset provided by the Economist Intelligence Unit (EIU), which spans a subset of 77 countries from our original dataset, but provides prices for 111 individual tradable goods in two types of retail stores.

Table 8: Estimation Results With EIU Data

	Estimate of θ (S.E.)	β_1	β_2
Cheap Stores			
Data Moments	—	4.17	5.11
Exactly Identified Case	2.47 (???)	4.17	—
Overidentified Case	2.48 (???)	4.19	5.09
Expensive Stores			
Data Moments	—	4.39	5.23
Exactly Identified Case	2.60 (???)	4.39	—
Overidentified Case	2.54 (???)	4.29	5.21

The results in table 8 suggest that aggregation causes a downward bias on trade barrier estimates, resulting in elasticity of trade estimates that are biased upwards. Indeed, when we use the highly disaggregate EIU dataset, the elasticity of trade falls to $2.55 - 2.66$. However, retail mark-ups do not seem to bias the estimates, since the elasticity of trade is not dramatically different whether high- or low-end store prices are used.

8.5.2. Model Approach

The aggregation problems discussed above are only reflected in the first-stage of our analysis which applies the Eaton and Kortum (2002) methodology to the actual ICP data. Once we invoke our simulation methodology which makes use of a particular model that features price heterogeneity, we employ simulated good-level prices in order to estimate trade barriers and therefore the elasticity of trade. From the point of view of the Eaton and Kortum (2002) model, aggregating the goods in the basic-heading manner employed by the ICP

[13]Our aggregation argument is different from the argument by Imbs and Mejean (2009) who demonstrate that imposing elasticities across disaggregated sectors of the economy to be equivalent results in lower elasticity of substitution estimates than ones obtained by allowing for heterogeneity.

is not possible. Recall that this one-sector model features a continuum of goods and each good is bought from the cheapest source. So, while varieties of this good which are potentially supplied from a number of sources could be aggregated into a basic heading, only the cheapest one of these varieties is actually supplied to the market and its price is recorded in the data. Hence each good in the data represents a particular basic heading and further aggregation that is consistent with the ICP methodology is impossible.

An aggregation argument potentially goes through in the monopolistic competition framework of Melitz (2003) and Chaney (2008). In these models, varieties produces by firms with identical productivity draws from different source countries can be thought of as varieties of a good produced with a particular productivity level. Thus, a basic heading price represents the average price of all varieties produced by firms with a particular productivity draw originating from different source countries. The implications of the monopolistic competition micro-structure on the estimates of the elasticity of trade are explored in Simonovska and Waugh (2010).

9. Conclusion

In this paper, we estimate the elasticity of trade using a simulated method of moments estimator that exploits micro-price data. Our estimates of the elasticity of trade for 123 countries that comprise most of world output range between 2.55 and 4.49. These results imply a doubling in the measured welfare gains from trade, relative to existing estimates obtained following the approach of Eaton and Kortum (2002).

The analysis in our paper has broader implications than merely arriving at unbiased estimates of the elasticity of trade. Results from Arkolakis, Costinot, and Rodriguez-Clare (2009) suggest that heterogenous firm and production models provide no value added for aggregate outcomes over models which abstract from heterogeneity. Our methodological approach suggests otherwise. In this paper, we exploited the structure of the Eaton and Kortum (2002) model to provide an estimate of the elasticity of trade, which is the key parameter to measuring the welfare gains from trade. Moreover, we demonstrated that model-independent bias reduction methods cannot fully eliminate the bias in estimates of the elasticity of trade obtained using a simple method of moments estimator. Finally, our approach to estimating this key parameter would not have been possible in models without heterogenous outcomes.

Thus, while the Eaton and Kortum (2002), Melitz (2003) and Chaney (2008) models may provide no new additional gains from trade, their structure allows us to provide a better

estimate of the elasticity of trade than a simple Armington model would have allowed. The ability to use both measurement and theory in ways that alternative models would not allow is an important component of the value added that new heterogenous firm and production models of international trade provide.

References

ALVAREZ, F., AND R. J. LUCAS (2007): "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade," *Journal of Monetary Economics*, 54(6), 1726–1768.

ANDERSON, J., AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93, 170–192.

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

ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (2009): "New Trade Models, Same Old Gains?," Working Paper 15628, National Bureau of Economic Research.

BERNARD, A., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268–1290.

BRODA, C., AND D. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121(2).

BURSTEIN, A., AND J. VOGEL (2009): "Globalization, Technology, and the Skill Premium," Discussion paper.

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

COOKE, P. (1979): "Statistical inference for bounds of random variables," *Biometrika*, 66(2), 367.

DAVIDSON, R., AND J. MACKINNON (2004): *Econometric Theory and Methods*. New York: Oxford University Press.

DONALDSON, D. (2009): "Railroads of the Raj: Estimating the Economic Impact of Transportation Infrastructure," *unpublished paper, MIT*.

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

EATON, J., S. KORTUM, AND F. KRAMARZ (2008): "An Anatomy of International Trade: Evidence from French Firms," NBER Working Paper 14610,.

GOURIÉROUX, C., AND A. MONFORT (1996): *Simulation-based econometric methods*, CORE lectures. Oxford University Press.

HAYASHI, F. (2000): *Econometrics*. Princeton.

HUMMELS, D. (2001): "Toward a Geography of Trade Costs," *Purdue University, unpublished mimeo*.

IMBS, J., AND I. MEJEAN (2009): "Elasticity Optimism," *HEC Lausanne, unpublished mimeo*.

KRUGMAN, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70(5), 950–959.

LUTTMER, E. G. J. (2007): "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 122 (3)(3), 1103–1144.

MACKINNON, J. G., AND A. A. SMITH (1998): "Approximate bias correction in econometrics," *Journal of Econometrics*, 85(2), 205–230.

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

MUENDLER, M.-A. (2009): "Converter from SITC to ISIC," *University of California - San Diego, unpublished mimeo*.

ROBSON, D., AND J. WHITLOCK (1964): "Estimation of a truncation point," *Biometrika*, 51(1-2), 33.

ROMALIS, J. (2007): "NAFTA's and CUSFTA's Impact on International Trade," *Review of Economics and Statistics*, 89(3), 416–35.

RUHL, K. (2008): "The International Elasticity Puzzle," *unpublished paper, NYU*.

SIMONOVSKA, I. (2010): "Income Differences and Prices of Tradables," Working Paper 16233, National Bureau of Economic Research.

SIMONOVSKA, I., AND M. WAUGH (2010): "The Elasticity of Trade: Estimates and Evidence II," *mimeo*.

SMITH, A. (2008): "Indirect inference," *The New Palgrave Dictionary of Economics, 2nd Edition (forthcoming)*.

WAUGH, M. (2010): "Bilateral Trade, Relative Prices, and Trade Costs," *unpublished paper, NYU*.

WAUGH, M. E. (2009): "International Trade and Income Differences," Staff Report 435, Federal Reserve Bank of Minneapolis.

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

A. Data Appendix

A.1. Trade Shares

To construct trade shares, we used bilateral trade flows and production data as follows:

$$\frac{X_{ni}}{X_n} = \frac{\text{Imports}_{ni}}{\text{Gross Mfg. Production}_n - \text{Total Exports}_n + \text{Imports}_n},$$

$$\frac{X_{nn}}{X_n} = 1 - \sum_{k \neq n}^{N} \frac{X_{ni}}{X_n}.$$

Putting the numerator and denominator together is simply computing an expenditure share by dividing the value of goods country n imported from country i by the total value of goods in country n. The home trade share $\frac{X_{nn}}{X_n}$ is simply constructed as the residual from one minus the sum of all bilateral expenditure shares.

To construct $\frac{X_{ni}}{X_n}$, the numerator is the aggregate value of manufactured goods that country n imports from country i. Bilateral trade flow data are from UN Comtrade for the year 2004. We obtain all bilateral trade flows for our sample of 123 countries at the four-digit SITC level. We then used concordance tables between four-digit SITC and three-digit ISIC codes provided by the UN and further modified by Muendler (2009).[14] We restrict our analysis to manufacturing bilateral trade flows only, namely, those that correspond with manufactures as defined in ISIC Rev.#2.

The denominator is gross manufacturing production minus total manufactured exports (for the whole world) plus manufactured imports (for only the sample). Gross manufacturing

[14]The trade data often report bilateral trade flows from two sources. For example, the exports of country A to country B can appear in the UN Comtrade data as exports reported by country A or as imports reported by country B. In this case, we take the report of bilateral trade flows between countries A and B that yields a higher total volume of trade across the sum of all SITC four-digit categories.

production data are the most serious data constraints we faced. We obtain manufacturing production data for 2004 from UNIDO for a large sub-sample of countries. We then imputed gross manufacturing production for countries for which data are unavailable as follows: We first obtain 2004 data on manufacturing (MVA) and agriculture (AVA) value added as well as population size (L) and GDP for all countries in the sample. We then impute the gross output (GO) to manufacturing value added ratio for the missing countries using coefficients resulting from the following regression:

$$\log\left(\frac{MVA}{GO}\right) = \beta_0 + \beta_{GDP}C_{GDP} + \beta_L C_L + \beta_{MVA}C_{MVA} + \beta_{AVA}C_{AVA} + \epsilon,$$

where β_x is a 1x3 vector of coefficients corresponding to C_x, an Nx3 matrix which contains $[\log(x), (\log(x))^2, (\log(x))^3]$ for the sub-sample of N countries for which gross output data are available.

A.2. Prices

The ICP price data we employ in our estimation procedure is reported at the basic-heading level. A basic heading represents a narrowly-defined group of goods for which expenditure data are available. For example, basic heading "1101111 Rice" is made up of prices of different types of rice and the resulting value is an aggregate over these different types of rice. This implies that a typical price observation of "Rice" contains different types of rice as well as different packaging options that affect the unit price of rice within and across countries.

According to the ICP Handbook, the price of the basic heading "Rice" is constructed using a transitive Jevons index of prices of different varieties of rice. To illustrate this point, suppose the world economy consists of 3 countries, A, B, C and 10 types of rice, 1-10. Further suppose that consumers in country 1 have access to all 10 types of rice; those in country 2 only have access to types 1-5 of rice; and those in country 3 have access to types 4-6 of rice. Although all types of rice are not found in all 3 countries, it is sufficient that each pair of countries shares at least one type of rice.

The ICP obtains unit prices for all available types of rice in all three countries and records a price of 0 if the type of rice is not available in a particular country. The relative price of rice between countries A and B, based on goods available in these two countries, $p_{AB}^{A,B}$, is a geometric average of the relative prices of rice of types $1-5$

$$p_{AB}^{A,B} = \left[\prod_{j=1}^{5} \frac{p_A(j)}{p_B(j)}\right]^{\frac{1}{5}}.$$

Similarly, one can compute the relative price of rice between countries A and C (B and C) based on varieties available in both A and C (B and C). The price of the basic heading "Rice" reported by the ICP is:

$$p_{AB} = \left[p_{AB}^{A,B} p_{AB}^{A,B} \frac{p_{AC}^{A,C}}{p_{BC}^{B,C}} \right]^{\frac{1}{3}},$$

which is a geometric average that features not only relative prices of rice between countries A and B, but also cross-prices between A and B linked via country C. This procedure ensures that prices of basic headings are transitive across countries and minimizes the impact of missing prices across countries.

Thus, a basic heading price is a geometric average of prices of varieties that is directly comparable across countries.

B. Proofs

Below we describe the steps to proving Lemmata 1 and 2. The key part in Lemma 1 is deriving the distribution of the maximal log price difference. We then prove Propositions 1 and 2.

B.1. Proof of Lemma 1, Lemma 2, and Proposition 1

FIrst, we derive the distribution of the maximal log price difference. The key insight is to work with direct comparisons of goods' prices (i.e. do not impose equilibrium and work from the equilibrium price distribution), compute the distribution of log price differences and then the distribution of the maximal log price difference.

Having obtained the distribution of the maximum log price difference, we show that the expected value of the maximum log price difference is biased in a finite sample and the estimator $\hat{\beta}$ is biased.

B.1.1. Preliminaries

In deriving the distribution of maximum log price differences, we will work with a relabeling of the production functions and exponential distributions following an argument in Alvarez and Lucas (2007). They relate the pdf's of the exponential and Frèchet distributions.

The claim is that if $z_i \sim \exp(T_i)$, then $y_i \equiv z_i^{-\frac{1}{\theta}} \sim \exp(-T_i y_i^{-\theta})$. To see this, notice that since $y_i = h(z_i)$ is a decreasing function, it must be that $f(z_i)dz_i = -g(y_i)dy_i$, where f, g are the pdf's of z_i, y_i respectively. The result will allow us to characterize moments of the log price difference by invoking properties of the exponential distribution.

B.1.2. Proof of Lemma 1

The proof of Lemma 1 follows.

Let $z_k^{-\frac{1}{\theta}} \sim \exp(-T_k(z_k^{-\frac{1}{\theta}})^{-\theta})$ be the productivity associated with good z, drawn from the Frèchet pdf in country k. By the argument above, the underlying distribution of z_k is exponential. The price for good z produced in country k and supplied to country i is $p_{ik} \equiv w_k \tau_{ik} z_k^{\frac{1}{\theta}}$. The relative price ratio of good z between countries n and i is:

$$v_{ni}(z) = \frac{\min\left\{\min_{k\neq i}[w_k\tau_{nk}z_k^{\frac{1}{\theta}}], w_i\tau_{ni}z_i^{\frac{1}{\theta}}\right\}}{\min\left\{\min_{k\neq n}[w_k\tau_{ik}z_k^{\frac{1}{\theta}}], w_n\tau_{in}z_n^{\frac{1}{\theta}}\right\}} \tag{b.1}$$

Take this object to the power of θ:

$$(v_{ni}(z))^{\theta} = \frac{\min\left\{\min_{k\neq i}[w_k^{\theta}\tau_{nk}^{\theta}z_k], w_i^{\theta}\tau_{ni}^{\theta}z_i\right\}}{\min\left\{\min_{k\neq n}[w_k^{\theta}\tau_{ik}^{\theta}z_k], w_n^{\theta}\tau_{in}^{\theta}z_n\right\}} \tag{b.2}$$

We want to characterize the distribution of (b.2), so we will first derive the pdf's of its components. Define $\tilde{z}_{ik} = w_k^{\theta}\tau_{ik}^{\theta}z_k$. Since $z_k \sim \exp(T_k)$, it must be that $\tilde{z}_{ik} \sim \exp(T_k w_k^{-\theta}\tau_{ik}^{-\theta})$. Let $\tilde{\lambda}_{ik} \equiv T_k w_k^{-\theta}\tau_{ik}^{-\theta}$.

Next, we derive the distribution of $\tilde{z}_i \equiv \min_{k\neq n}[w_k^{\theta}\tau_{ik}^{\theta}z_k] = \min_{k\neq n}[\tilde{z}_{ik}]$. Since each $\tilde{z}_{ik} \sim \exp(\tilde{\lambda}_{ik})$ and independent across countries k, $\tilde{z}_i \sim \exp(\sum_{k\neq n}\tilde{\lambda}_{ik})$. Define $\tilde{\lambda}_i \equiv \sum_{k\neq n}\tilde{\lambda}_{ik}$. Repeat the procedure for importer n in the numerator.

Given these definitions, (b.2) can be rewritten as:

$$(v_{ni}(z))^{\theta} = \frac{\min\left\{\tilde{z}_n, w_i^{\theta}\tau_{ni}^{\theta}z_i\right\}}{\min\left\{\tilde{z}_i, w_n^{\theta}\tau_{in}^{\theta}z_n\right\}} \tag{b.3}$$

Define $\epsilon_{ni}(z) = \log(v_{ni}(z))$. Taking logs of expression (b.3) gives:

$$\begin{aligned}\theta\epsilon_{ni}(z) = \quad & \min\left\{\log(\tilde{z}_n), [\theta\log(w_i) + \theta\log(\tau_{ni}) + \log(z_i)]\right\} \\ - \quad & \min\left\{\log(\tilde{z}_i), [\theta\log(w_n) + \theta\log(\tau_{in}) + \log(z_n)]\right\}\end{aligned} \tag{b.4}$$

Next, we argue that $\theta\epsilon_{ni}(z) \in [-\theta\log(\tau_{in}), \theta\log(\tau_{ni})]$. For any good z, $\theta\epsilon_{ni}(z)$ can satisfy one and only one of the following three cases:

1. Countries n and i buy good z from two different sources. Then,

$$\theta\epsilon_{ni}(z) = \log(\tilde{z}_n) - \log(\tilde{z}_i) \tag{b.5}$$

2. Country n buys good z from country i. Assuming that trade barriers don't violate the triangle inequality, it must be that i buys the good from itself. Then,

$$\theta\epsilon_{ni}(z) = \theta\log(w_i) + \theta\log(\tau_{ni}) + \log(z_i) - \theta\log(w_i) - \log(z_i) = \theta\log(\tau_{ni}) \tag{b.6}$$

3. Country i buys good z from n. Then it must be that n buys the good from itself, so:

$$\theta\epsilon_{ni}(z) = \theta\log(w_n) + \log(z_n) - \theta\log(w_n) - \theta\log(\tau_{in}) - \log(z_n) = -\theta\log(\tau_{in}) \tag{b.7}$$

We claim that the following ordering occurs: $-\theta\log(\tau_{in}) \leq \log(\tilde{z}_n) - \log(\tilde{z}_i) \leq \theta\log(\tau_{ni})$. To show this, we need to consider the following two scenarios:

1. Countries n and i buy good z from the same source k. Then,

$$\begin{aligned}\log(\tilde{z}_n) - \log(\tilde{z}_i) &= \log(w_k^\theta \tau_{nk}^\theta z_k) - \log(w_k^\theta \tau_{ik}^\theta z_k) \\ &= \theta(\log(\tau_{nk}) - \log(\tau_{ik}))\end{aligned} \tag{b.8}$$

   Clearly,

$$\theta(\log(\tau_{nk}) - \log(\tau_{ik})) \geq -\theta\log(\tau_{in}) \iff \tau_{in}\tau_{nk} \geq \tau_{ik}$$

   where the latter inequality is true under the triangle inequality assumption.

   Similarly,

$$\theta(\log(\tau_{nk}) - \log(\tau_{ik})) \leq \theta\log(\tau_{ni}) \iff \tau_{nk} \leq \tau_{ni}\tau_{ik}$$

   again true by triangle inequality.

2. Country n buys good z from source a and country i from source b, $a \neq b$. We want to show that $-\theta\log(\tau_{in}) \leq \log(w_a^\theta \tau_{na}^\theta z_a) - \log(w_b^\theta \tau_{ib}^\theta z_b) \leq \theta\log(\tau_{ni})$.

Since n imported from a over b, it must be that:

$$w_a^\theta \tau_{na}^\theta z_a \leq w_b^\theta \tau_{nb}^\theta z_b \tag{b.9}$$

Similarly, since i imported from b over a, it must be that:

$$w_b^\theta \tau_{ib}^\theta z_b \leq w_a^\theta \tau_{ia}^\theta z_a \tag{b.10}$$

To find the upper bound, take logs of (b.9) and subtract $\log(w_b^\theta \tau_{ib}^\theta z_b)$ from both sides:

$$\log(w_a^\theta \tau_{na}^\theta z_a) - \log(w_b^\theta \tau_{ib}^\theta z_b) \leq \log(w_b^\theta \tau_{nb}^\theta z_b) - \log(w_b^\theta \tau_{ib}^\theta z_b) \tag{b.11}$$

It suffices to show that the right hand side is itself below the upper bound, since by transitivity so is the left hand side (which is the object of interest).

$$\begin{aligned} &\log(w_b^\theta \tau_{nb}^\theta z_b) - \log(w_b^\theta \tau_{ib}^\theta z_b) \leq \theta \log(\tau_{ni}) \\ \iff &\theta \log(\tau_{nb}) - \theta \log(\tau_{ib}) \leq \theta \log(\tau_{ni}) \\ \iff &\tau_{nb} \leq \tau_{ni}\tau_{ib} \end{aligned} \tag{b.12}$$

which is true by triangle inequality.

The argument for the lower bound is similar. Take logs of (b.10), multiply by -1 (and reverse inequality) and add $\log(w_a^\theta \tau_{na}^\theta z_a)$ to both sides:

$$\log(w_a^\theta \tau_{na}^\theta z_a) - \log(w_b^\theta \tau_{ib}^\theta z_b) \geq \log(w_a^\theta \tau_{na}^\theta z_a) - \log(w_a^\theta \tau_{ia}^\theta z_a) \tag{b.13}$$

It suffices to show that the right hand side is itself above the lower bound, since by transitivity so is the left hand side (which is the object of interest).

$$\begin{aligned} &\log(w_a^\theta \tau_{na}^\theta z_a) - \log(w_a^\theta \tau_{ia}^\theta z_a) \geq -\theta \log(\tau_{in}) \\ \iff &\theta \log(\tau_{na}) - \theta \log(\tau_{ia}) \geq -\theta \log(\tau_{in}) \\ \iff &\tau_{in}\tau_{na} \geq \tau_{ia} \end{aligned} \tag{b.14}$$

which is true by triangle inequality.

Hence, $\theta\epsilon_{ni}(z) \in [-\theta \log(\tau_{in}), \theta \log(\tau_{ni})]$.

Next, we proceed to derive the distribution of $\theta\epsilon_{ni}(z) = \log(\tilde{z}_n) - \log(\tilde{z}_i)$. First we derive the pdf's of its two components.

Let $y_i \equiv \log(\tilde{z}_i)$. Then $\tilde{z}_i = \exp(y_i)$. The pdf of y_i must satisfy:

$$\begin{aligned} f(y_i)dy_i = g(\tilde{z}_i)d\tilde{z}_i \quad &\Rightarrow \quad f(y_i) = \tilde{\lambda}_i \exp(-\tilde{\lambda}_i \tilde{z}_i)\frac{d\tilde{z}_i}{dy_i} \\ &\Rightarrow \quad f(y_i) = \tilde{\lambda}_i \exp(-\tilde{\lambda}_i \exp(y_i)) \exp(y_i) \\ &\Rightarrow \quad F(y_i) = 1 - \exp(-\tilde{\lambda}_i \exp(y_i)) \end{aligned} \tag{b.15}$$

The same holds for n.

Now that we have the pdf's of the two components, we can define the pdf of $\epsilon \equiv \theta\epsilon_{ni}(z) \in [-\theta\log(\tau_{in}), \theta\log(\tau_{ni})]$ as follows:

$$f(\epsilon) \equiv f_{y_n - y_i}(x) = \int_{-\infty}^{\infty} f_{y_n}(y) f_{y_i}(y - x)dy \tag{b.16}$$

where we have used the fact that y_n and y_i are independently distributed, hence the pdf of their difference is the convolution of the pdf's of the two random variables.

Substituting the pdf's of y_n and y_i into (b.16) yields:

$$\begin{aligned} f(\epsilon) =& \int_{-\infty}^{\infty} \tilde{\lambda}_n \exp(-\tilde{\lambda}_n \exp(y)) \exp(y) \tilde{\lambda}_i \exp(-\tilde{\lambda}_i \exp(y - \epsilon)) \exp(y - \epsilon)dy \\ =& \frac{-\tilde{\lambda}_n \tilde{\lambda}_i}{(\tilde{\lambda}_n \exp(\epsilon) + \tilde{\lambda}_i)^2} \left[\frac{\tilde{\lambda}_n \exp(y + \epsilon) + \tilde{\lambda}_i \exp(y) + \exp(\epsilon)}{\exp\left\{\exp(y)(\tilde{\lambda}_n + \tilde{\lambda}_i \exp(-\epsilon))\right\}} \right]_{y=-\infty}^{y=+\infty} \end{aligned} \tag{b.17}$$

Let $v(y)$ be the expression in the bracket.

$$\lim_{y \to -\infty} v(y) = \frac{0 + 0 + \exp(\epsilon)}{\exp\{0\}} = \exp(\epsilon) \tag{b.18}$$

For the upper bound, we use l'Hopital rule:

$$\begin{aligned} \lim_{y \to \infty} v(y) &= \lim_{y \to \infty} \frac{\tilde{\lambda}_n \exp(y + \epsilon) + \tilde{\lambda}_i \exp(y)}{\exp\left\{\exp(y)(\tilde{\lambda}_n + \tilde{\lambda}_i \exp(-\epsilon))\right\} \exp(y)(\tilde{\lambda}_n + \tilde{\lambda}_i \exp(-\epsilon))} \\ &= \lim_{y \to \infty} \frac{\tilde{\lambda}_n \exp(\epsilon) + \tilde{\lambda}_i}{\exp\left\{\exp(y)(\tilde{\lambda}_n + \tilde{\lambda}_i \exp(-\epsilon))\right\} (\tilde{\lambda}_n + \tilde{\lambda}_i \exp(-\epsilon))} \\ &= 0 \end{aligned} \tag{b.19}$$

Thus (b.17) becomes:

$$f(\epsilon) = \frac{\tilde{\lambda}_n \tilde{\lambda}_i \exp(\epsilon)}{(\tilde{\lambda}_n \exp(\epsilon) + \tilde{\lambda}_i)^2} \tag{b.20}$$

The corresponding cdf is:

$$F(\epsilon) = 1 - \frac{\tilde{\lambda}_i}{\tilde{\lambda}_n \exp(\epsilon) + \tilde{\lambda}_i} \tag{b.21}$$

Given that ϵ is bounded, we can compute the truncated pdf as:

$$\begin{aligned} f_T(\epsilon) &= \frac{f(\epsilon)}{F(\theta \log(\tau_{ni})) - F(-\theta \log(\tau_{in}))} \\ &= \gamma^{-1} \frac{\tilde{\lambda}_n \tilde{\lambda}_i \exp(\epsilon)}{(\tilde{\lambda}_n \exp(\epsilon) + \tilde{\lambda}_i)^2}, \end{aligned} \tag{b.22}$$

where:

$$\gamma = -\frac{\tilde{\lambda}_i}{\tilde{\lambda}_n \exp(\theta \log(\tau_{ni})) + \tilde{\lambda}_i} + \frac{\tilde{\lambda}_i}{\tilde{\lambda}_n \exp(-\theta \log(\tau_{in})) + \tilde{\lambda}_i} \tag{b.23}$$

Similarly, the truncated cdf is:

$$F_T(\epsilon) = \gamma^{-1} \int_{-\theta \log(\tau_{in})}^{\epsilon} f(t) dt \tag{b.24}$$

Now that we have these distributions, we compute order statistics from them, which allow us to characterize the trade barriers estimated from price data. We use the following result: Given L observations drawn from pdf $h(x)$, the pdf of the r-th order statistic (where $r = L$ is the max and $r = 1$ is the min) is:

$$h_r(x) = \frac{L!}{(r-1)!(L-r)!} H(x)^{r-1} (1 - H(x))^{L-r} h(x) \tag{b.25}$$

The pdf of the max reduces to:

$$h_{\max}(x, L) = L H(x)^{L-1} h(x)$$

With this pdf defined, we can compute the expectation of the maximum statistic:

$$E[\max_{z \in L}(x_z)] = \int_{-\infty}^{\infty} x h_{\max}(x, L) dx \tag{b.26}$$

Recall that we are interested in computing the expectation of the maximum logged price difference between countries n and i. But, so far, we have derived the truncated pdf and cdf of $\epsilon = \theta \log(v_{ni}(z))$. Our object of interest is actually $\log(v_{ni}(z)) = \frac{1}{\theta}\epsilon$. The expectation of this object, which represents the maximum log price difference, for L draws, is given by:

$$E[\max_{z \in L}(\log(p_n(z)) - \log(p_i(z)))] = \frac{1}{\theta}\int_{-\theta \log(\tau_{in})}^{\theta \log(\tau_{ni})} \epsilon f_{\max}(\epsilon, L) d\epsilon \tag{b.27}$$

where:

$$\begin{aligned} f_{\max}(\epsilon, L) =& L F_T(\epsilon)^{L-1} f_T(\epsilon) \\ =& L \left[\gamma^{-1}\int_{-\theta \log(\tau_{in})}^{\epsilon} f(t)dt\right]^{L-1} \gamma^{-1} \frac{\tilde{\lambda}_n \tilde{\lambda}_i \exp(\epsilon)}{(\tilde{\lambda}_n \exp(\epsilon) + \tilde{\lambda}_i)^2} \end{aligned} \tag{b.28}$$

Hence, the expectation of the maximum of the log price difference is proportional to $1/\theta$, where the proportionality object comes from gravity,

$$E[\max_{z \in L}(\log(p_n(z)) - \log(p_i(z)))] = \Psi_{ni}(L; \mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}), \tag{b.29}$$

where:

$$\Psi_{ni}(L; \mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}) \equiv \frac{1}{\theta}\int_{-\theta \log(\tau_{in})}^{\theta \log(\tau_{ni})} \epsilon f_{\max}(\epsilon, L) d\epsilon \tag{b.30}$$

and the values $\mathbf{S}$ and $\tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}$ correspond with the definitions outlined in Definition 1. It is worth emphasizing the nature of this integral: other than the scalar in the front, it depends completely on objects that can be recovered from the standard gravity equation in (22).

Finally, one can rewrite equation (b.30) via integration by parts as:

$$E[\max_{z \in L}(\log(p_n(z)) - \log(p_i(z)))] = \log \tau_{ni} - \frac{1}{\theta}\int_{-\theta \log(\tau_{in})}^{\theta \log(\tau_{ni})} F_{\max}(\epsilon, L) d\epsilon \tag{b.31}$$

$$\log \tau_{ni} = E[\max_{z \in L}(\log(p_n(z)) - \log(p_i(z)))] + \frac{1}{\theta}\int_{-\theta \log(\tau_{in})}^{\theta \log(\tau_{ni})} F_{\max}(\epsilon, L) d\epsilon \tag{b.32}$$

which implies the following strict inequality

$$\log \tau_{ni} > E[\max_{z \in L}(\log(p_n(z)) - \log(p_i(z)))] = \Psi_{ni}(L; \mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}) \tag{b.33}$$

where the strict inequality simply follows from the inspection of the CDF $F_{\max}(\epsilon, L)$ which

has positive mass below the point $\theta \log(\tau_{ni})$. This then proves claim 1. in Lemma 1.

To prove claim 2. in Lemma 1, we compute the difference in the expected values of log prices between two countries. We show that they are equal to the (scaled) difference in the price parameters Φ.

Rather than working with the distribution described above, it is more convenient to directly compute the expectation of log prices using the equilibrium price distribution. Note that Eaton and Kortum (2002) show that the cdf and pdf of prices in country i are $G(p) = 1 - \exp(-\Phi_i p^\theta)$ and $g(p) = p^{\theta-1}\theta\Phi_i \exp(-\Phi_i p^\theta)$, respectively.

For any country i, define the expectation of logged prices as

$$E[\log(p_i(z))] = \int_0^\infty \log(p) g(p) dp \tag{b.34}$$

Substituting the pdf of prices and then utilizing some algebra to find an appropriate change in variables, expression (b.34) yields

$$\begin{aligned} E[\log(p_i(z))] &= \int_0^\infty \log(p) p^{\theta-1}\theta\Phi_i \exp(-\Phi_i p^\theta) dp \\ &= \int_0^\infty \log(p)\theta\Phi_i \exp(\theta \log(p)) \exp(-\Phi_i \exp(\theta \log(p))) \frac{dp}{p} \end{aligned}$$

Our change of variables will set $x = \log(p)$, which yields $dx/dp = 1/p$. Then, integration by change of variables allows us to rewrite the above as

$$\begin{aligned} E[\log(p_i(z))] &= \int_0^\infty \log(p)\theta\Phi_i \exp(\theta \log(p)) \exp(-\Phi_i \exp(\theta \log(p))) \frac{dp}{p} \\ &= \int_0^\infty x\theta\Phi_i \exp(\theta x) \exp(-\Phi_i \exp(\theta x)) \frac{\theta}{\theta} dx \end{aligned}$$

Let $y = \theta x$, so that $dy/dx = \theta$, then another change of variables gives

$$E[\log(p_i(z))] = \frac{1}{\theta} \int_0^\infty y\Phi_i \exp(y) \exp(-\Phi_i \exp(y)) dy$$

Let $t = \Phi_i \exp(y)$, so that $dt/dy = \Phi_i \exp(y)$ and $y = \log(t/\Phi_i)$. Then

$$\begin{aligned} E[\log(p_i(z))] =& \frac{1}{\theta} \int_0^\infty \log\left(\frac{t}{\Phi_i}\right) \exp(-t) dt \\ =& \frac{1}{\theta} \left\{ \int_0^\infty \log(t) \exp(-t) dt - \int_0^\infty \log(\Phi_i) \exp(-t) dt \right\} \\ =& -\frac{1}{\theta} \{\gamma + \log(\Phi_i)\} \end{aligned} \tag{b.35}$$

where γ is the Euler-Mascheroni constant. Finally, using (b.35) and taking the expected difference in log prices between country n and country i, the scaled Euler-Mascheroni constant cancels between the two countries and leaves the following expression

$$\begin{aligned} E[\log(p_n(z))] - E[\log(p_i(z))] \quad &= -\frac{1}{\theta} \{\log(\Phi_n) - \log(\Phi_i)\} \\ &\equiv \Omega_{ni}(\mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}), \end{aligned} \tag{b.36}$$

which then proves claim 2. in Lemma 1.

B.1.3. Proof of Lemma 2 and Proposition 1

To prove Lemma 2 and Proposition 1 we invert Eaton and Kortum's (2002) estimator for the elasticity of trade:

$$\frac{1}{\hat{\beta}} = -\frac{\sum_n \sum_i \left(\log \hat{\tau}_{ni} + \log \hat{P}_i - \log \hat{P}_n \right)}{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)} \tag{b.37}$$

Given the assumption that the trade data are fixed, equation (b.37) is linear in the random variables $\log \hat{\tau}_{ni}$ and $(\log \hat{P}_n - \log \hat{P}_i)$. With this observation, taking expectations of these random variables yields

$$E\left(\frac{1}{\hat{\beta}}\right) = \frac{1}{\theta} \left\{ -\frac{\sum_n \sum_i (\theta \Psi_{ni}(L) - (\log \Phi_i - \log \Phi_n))}{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)} \right\} < \frac{1}{\theta}, \tag{b.38}$$

by substituting in for the expectation of the maximum log price difference using (b.30), and the difference in expectations of log prices using (b.36). Inspection of the bracketed term

above implies the following strict inequality must hold,

$$1 \; > \; \left\{ -\frac{\sum_n \sum_i (\theta \Psi_{ni}(L) - (\log \Phi_i - \log \Phi_n))}{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)} \right\} \; > \; 0, \tag{b.39}$$

with the reason being that $\Psi_{ni}(L) < \log \tau_{ni}$ from Lemma 1, otherwise the bracketed term would correspond exactly with equation (5) in logs and thus equal one. Now inverting the expression above and applying Jensen's inequality results in the following:

$$E(\hat{\beta}) \; \geq \; \left[E\left(\frac{1}{\hat{\beta}}\right) \right]^{-1} = \theta \left\{ -\frac{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}{\sum_n \sum_i (\theta \Psi_{ni}(L) - (\log \Phi_i - \log \Phi_n))} \right\} > \theta. \tag{b.40}$$

With the strict inequality following from (b.38) and (b.39). This proves Proposition 1.

B.2. Proof of Proposition 2

In this subsection, we prove Proposition 2. To prove the claims in Proposition 2, we start with claim 1.

To prove claim 1., we argue that the sample maximum of scaled log price differences is a consistent estimator of the scaled trade cost. In particular, we argue that as the sample size becomes infinite the probability that the sample scaled trade cost is arbitrarily close to the true scaled trade cost is one.

To see this, consider an estimate of the scaled trade barrier, given a sample of L goods' prices,

$$\theta \log \hat{\tau}_{ni}^{L} = \theta \left\{ \max_{\ell=1,\ldots,L} (\log p_n(\ell) - \log p_i(\ell)) \right\}. \tag{b.41}$$

The cdf of this random variable is the integral of its pdf, which is given in expression (b.28), over the compact interval in which the scaled logged price difference lies, $[-\theta \log \tau_{in}, \theta \log \tau_{ni}]$. Denote this cdf by F_{max}^{L}. From (b.28), $F_{max}^{L} \equiv (F_T)^L$, where F_T is the truncated distribution of the scaled log price difference over the domain $[-\theta \log \tau_{in}, \theta \log \tau_{ni}]$. By definition, F_T and F_{max}^{L} take on values between zero and one, as they are cdfs. In particular, for any realization $x < \theta \log \tau_{ni}$, $F_T(x) < 1$. For any $L > 1$, $F_{max}^{L}(x) = (F_T(x))^L \leq F_T(x) < 1$.

Take $L \to \infty$. Then, for any $x \in [-\theta \log \tau_{in}, \theta \log \tau_{ni})$, $F_{max}^{L} = (F_T)^L$ becomes arbitrarily close to zero since $F_T < 1$. Hence, all the mass of the cdf F_{max}^{L} becomes concentrated at $\theta \log \tau_{ni}$. Thus, as the sample size becomes infinite, the estimated scaled trade barrier converges to

the true scaled trade barrier, in probability. Rescaling everything by $\frac{1}{\theta}$ then implies

$$\operatorname*{plim}_{L\to\infty} \Psi_{ni}(L; \mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}) \;=\; \log(\tau_{ni}). \tag{b.42}$$

This proves claim 1. of Proposition 2.

To show consistency of the estimator $\hat{\beta}$, we argue that

$$\operatorname*{plim}_{L\to\infty} \hat{\beta}\left(L; \mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)\right) = \theta, \tag{b.43}$$

or equivalently that, $\forall \delta > 0$,

$$\lim_{L\to\infty} \Pr\left[\parallel \hat{\beta}\left(L; \mathbf{S}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{n}}, \tilde{\boldsymbol{\tau}}_{\boldsymbol{i}}, \frac{X_{ni}/X_n}{X_{ii}/X_i}\right) - \theta \parallel < \delta\right] = 1. \tag{b.44}$$

Basically we will argue that, by sampling the prices of an ever increasing set of goods, and applying the estimator β over these prices, with probability one, we will obtain estimates that are arbitrarily close to θ.

Inverting the expression for the estimator $\hat{\beta}$ in expression (13), rearranging, and multiplying and dividing by the scalar θ yields

$$\frac{1}{\hat{\beta}} = \frac{1}{\theta} \frac{\sum_n \sum_i \left(\theta \log \hat{\tau}_{ni} - \theta[\log \hat{P}_n - \log \hat{P}_i]\right)}{-\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}. \tag{b.45}$$

By assumption, the denominator is trade data and is not a random variable.

In the numerator, $\log \hat{P}_n - \log \hat{P}_i$ is the difference in the average of logged prices for countries n and i, given a sample of L goods. In particular,

$$\log \hat{P}_n - \log \hat{P}_i \equiv \frac{1}{L}\sum_{\ell=1}^{L} \log p_n(\ell) - \frac{1}{L}\sum_{\ell=1}^{L} \log p_i(\ell) \tag{b.46}$$

We refer the reader to Davidson and MacKinnon (2004) for a proof of the well known result that the sample average is both an unbiased and consistent estimator of the mean. Since the difference operator is continuous, the difference in the sample average of logged price is an unbiased and consistent estimator of the difference in mean logged prices. Finally, multiplying these sample averages by a scalar θ, a continuous operation, ensures convergence to true difference in the price terms Φ.

We have argued that the two components in the numerator converge in probability to their true parameter counterparts, as the sample size becomes infinite. Taking the difference of these two components, summing over all country pairs (n, i), and dividing by the scalar $\left[-\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)\right]$, all of which are continuous operations, allows us to conclude that

$$\underset{L\to\infty}{\text{plim}} \frac{\sum_n \sum_i \left(\theta \log \hat{\tau}_{ni}^L - \theta[\log \hat{P}_n - \log \hat{P}_i]\right)}{-\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}$$

$$= \underset{L\to\infty}{\text{plim}} \frac{\sum_n \sum_i \left(\theta \max_{\ell=1,\ldots,L}(\log p_n(\ell) - \log p_i(\ell)) - \theta\left[\frac{1}{L}\sum_{\ell=1}^L \log p_n(\ell) - \frac{1}{L}\sum_{\ell=1}^L \log p_i(\ell)\right]\right)}{-\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}$$

$$= \frac{-\sum_n \sum_i (\theta \log \tau_{ni} - [\log \Phi_i - \log \Phi_n])}{\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right)}. \tag{b.47}$$

To complete the argument, consider the log of expression (5) which involves Φ. Summing this expression over all (n, i) country pairs gives:

$$\sum_n \sum_i \log\left(\frac{X_{ni}/X_n}{X_{ii}/X_i}\right) = -\sum_n \sum_i (\theta \log \tau_{ni} - [\log \Phi_i - \log \Phi_n]). \tag{b.48}$$

Substituting expression (b.48) in the denominator of (b.47) above makes the fraction in that expression equal to unity. Hence $1/\hat{\beta}$ converges to $1/\theta$ in probability. Since, for $\beta \in (0, \infty)$, $1/\hat{\beta}$ is a continuous function of $\hat{\beta}$, $\hat{\beta}$ converges to θ in probability. This proves claim 2. of Proposition 2.

Claim 3. of Proposition 2 follows from the fact that $\hat{\beta}$ is a consistent estimator of θ (see Hayashi (2000) for a discussion).

B.3. Deriving the Inverse Marginal Cost Distribution

To simulate the model, we argue that by using the coefficients **S** estimated from the gravity regression (22), we have enough information to simulate prices and trade flows. The key insight is that the S's are sufficient to characterize the inverse marginal cost distribution. Thus, we can sample from this distribution, and then compute equilibrium prices and trade flows.

To see this argument let $z_i \sim F_i(z_i) = \exp(-T_i z_i^{-\theta})$ and define $u_i \equiv z_i/w_i$. The pdf of z_i is

$f_i(z_i) = \exp(-T_i z_i^{-\theta})\theta T_i z_i^{-\theta-1}$. To find the pdf of the transformation u_i, $m_i(u_i)$, use the fact that $f_i(z_i)dz_i = m_i(u_i)du_i$, or $m_i(u_i) = f_i(z_i)(du_i/dz_i)^{-1}$. Let $\tilde{S}_i = T_i w_i^{-\theta}$. Using $f_i(z_i)$, $\tilde{S}_i$, and the fact that $du_i/dz_i = 1/w_i$, we obtain:

$$\begin{aligned} m_i(u_i) = f_i(z_i)\left(\frac{du_i}{dz_i}\right)^{-1} &= \exp(-T_i z_i^{-\theta})\theta T_i z_i^{-\theta-1}\left(\frac{1}{w_i}\right)^{-1} \\ &= \exp\left(-T_i z_i^{-\theta}\frac{w_i^{-\theta}}{w_i^{-\theta}}\right)\theta T_i z_i^{-\theta-1}\left(\frac{1}{w_i}\right)^{-1}\frac{w_i^{-\theta}}{w_i^{-\theta}} \\ &= \exp\left(-\tilde{S}_i\frac{z_i^{-\theta}}{w_i^{-\theta}}\right)\theta\tilde{S}_i\frac{z_i^{-\theta-1}}{w_i^{-\theta-1}} \\ &= \exp\left(-\tilde{S}_i u_i^{-\theta}\right)\theta\tilde{S}_i u_i^{-\theta-1} \end{aligned}$$

Clearly $m_i(u_i)$ is the pdf that corresponds to the cdf $M_i(u_i) = \exp(-\tilde{S}_i u_i^{-\theta})$, which concludes the argument.