Trade Models, Trade Elasticities, and the Gains from Trade

Ina Simonovska
University of California, Davis and NBER

Michael E. Waugh
New York University and NBER

This Version: June 2014

PRELIMINARY

ABSTRACT ———

We argue that the welfare gains from trade in new trade models with various micro-level margins are larger relative to models without these margins. Theoretically, we show that for fixed trade elasticity, different models have identical predictions for bilateral trade flows but different predictions for micro-level price variation. Thus, given data on trade flows and micro-level prices, different models have different implied trade elasticities and welfare gains from trade. Empirically, we quantify the difference by estimating the trade elasticity and the welfare cost of autarky under different assumptions about the underlying model. Relative to an Armington or Krugman trade model, an active extensive margin increases the welfare cost of autarky by 20 to 30 percent; variable markups further increase the cost by 30 percent. Finally, we incorporate into our estimation moment conditions that use trade flow and tariff data, which imply a common trade elasticity; this modification has virtually no effect on our results.

———

1. Introduction

This paper argues both theoretically and empirically that the welfare gains from trade in new trade models with various micro-level margins are larger relative to models without these margins. In a large and important class of trade models, we show that micro-level cross-country price variation is informative about the trade elasticity and that different models have different implications for micro-level price variation, even though their predictions for aggregate trade are identical.[1] These two facts imply that, given data on trade shares and micro-level prices, different assumptions about the underlying model yield different trade elasticities and consequently different welfare gains from trade. Empirically, we quantify these differences and we find significantly larger welfare gains in models with various micro-level margins versus older models without these margins.

Our theoretical analysis focuses on three canonical models of trade: Armington, Eaton and Kortum (2002) (henceforth EK) , and Bernard et al. (2003) (henceforth BEJK). Analyzing these three models is insightful because each model adds an additional micro-level margin of adjustment from a reduction in trade costs. In particular, Armington features an intensive margin only (i.e. reductions in trade costs lead to higher quantities of trade), EK adds an extensive margin (i.e. reductions in trade costs simply lead to additional goods being traded), and BEJK further adds a variable mark-up margin. Furthermore, all these models fit into the class of models examined by Arkolakis, Costinot, and Rodriguez-Clare (2011) where trade elasticity and the share of expenditure on domestic goods are sufficient statistics to measure the welfare cost of autarky.

In a symmetric two-country version of each model, we examine the distribution of cross-country price gaps. Assuming a fixed trade elasticity, we show that the price gap distribution in the Armington model stochastically dominates that in the EK model which further dominates the one in the BEJK model. This ranking of these distributions implies a similar ranking in the expectation of the largest order statistic of price gaps in a finite sample of prices from each model. The statistic is inversely related to the trade elasticity. Hence, to match the same order statistic in the data, the Armington model requires the highest trade elasticity, while the BEJK model needs the lowest. Because this argument is made with the constraint that all models have identical aggregate trade flows, the result together with the sufficient statistic formula of Arkolakis et al. (2011) implies that the welfare cost of autarky is different across models.

We quantify the importance of the theoretical results by estimating the trade elasticities in the multi-country asymmetric versions of the three models. The particular estimation approach that we use builds both on the theory described above and our previous work in Simonovska

---

[1]The class of models includes Armington or Anderson (1979), Krugman (1980), Eaton and Kortum (2002), Bernard, Eaton, Jensen, and Kortum (2003), and Melitz (2003) as articulated in Chaney (2008), which all generate log-linear relationships between bilateral trade flows and trade frictions.

and Waugh (2014). The basic idea behind our estimation strategy is to chose the trade elasticity to match moments between the model and the data about order statistics of bilateral price gaps. While we focus on a broader set of models than in Simonovska and Waugh (2014), the methodological approach is similar. First, we estimate the parameters of the models necessary to simulate micro-level price data using bilateral trade flow data, which guaranties that all models have identical aggregate trade predictions. Second, we use these parameter estimates and a given trade elasticity to simulate micro-level prices from each model and to construct the model-implied moments. Third, we choose the trade elasticity to minimize the distance between each model and the data.

Using price and trade flow data for the year 2004 for the 30 largest countries in terms of gross output, we estimate trade elasticities (with various specifications: exactly identified, overidentified, different weighting matrices) for each of the three models. Across various specifications, the estimate of the trade elasticity is systematically lower for the new trade models relative to the Armington model, and further the estimate for BEJK is lower than for EK. The difference in magnitudes is substantial. The EK estimate is about 20 percent lower than Armington, implying that the welfare cost of autarky is 20 percent higher in the EK model. In comparison, our estimate of the trade elasticity in the BEJK model is about 33 percent lower relative to EK, implying the welfare gains are estimated to be 50 percent larger in BEJK relative to Armington.

We further examine versions of the two canonical endogenous variety models: the monopolistic competition framework of Krugman (1980) and its counterpart that features micro-level heterogeneity—or an extensive margin of trade due to productivity-based firm selection—Melitz (2003). We apply our estimation procedure on the two frameworks and obtain trade elasticity estimates in the Melitz model that are 30 percent lower than Krugman, resulting in 30 percent higher welfare gains from trade in the former.

The reason behind the result is the extensive margin of trade featured in the Melitz model. In both models, relative prices of traded goods equal trade barriers. Whereas in Krugman all goods are sold to all destinations, in the Melitz model there are instances in which some products are not sold everywhere because some firms are not productive enough to reach each market. In this case, a relative price ratio cannot be defined since the product is not available in one of the destinations. Therefore, in order to be able to define relative prices in the Melitz model and derive moments from the relative price distribution used to identify trade elasticities, one must condition only on the firms that sell everywhere. These firms are likely to be from source countries with low trade costs to export. This implies that, in the Melitz model, the average maximal price difference across country pairs is likely to be below the average trade friction across country pairs. Thus, one needs a lower trade elasticity in Melitz relative to Krugman to rationalize the same amount of trade observed in the data. The differences in trade elasticities then translate into differences in the measured welfare gains from trade in the two models.

Our conclusion regarding the ranking of welfare between the Krugman and the Melitz model is in line with the conclusions reached by Melitz and Redding (2014). The authors also argue that the Melitz model generates higher welfare gains from trade than the Krugman model. Their argument, however, is very different from ours and it goes as follows. The Krugman model is a special case of the Melitz model when the distribution of firm productivities becomes degenerate. Naturally, the Pareto shape parameter—or trade elasticity—is the central parameter that governs this distribution. The authors then demonstrate that, if all other parameters of the two models are kept identical, the Melitz model generates higher trade shares and therefore higher welfare gains from trade. Instead, our exercise estimates the parameters of the two models such that they generate *identical* trade shares—the ones observed in the data. Then, the reason why the welfare gains differ is because different trade elasticities are necessary in order to match the same price moments in the data.

A natural question is why we focus on micro-level price variation as a means to estimating trade elasticities. The focus on micro-level price variation is interesting/informative for several reasons. First, all models make concrete and distinct predictions about price variation at the micro level. This is in contrast to moments about the firm/plant size distribution which is indeterminant in the Armington and EK model. Second, micro-level price variation speaks directly to the key economic mechanisms at work in each model. The active extensive margin in the EK or Melitz model or variable markups in BEJK all manifest themselves in different patterns of micro-level price variation. Thus, to understand if these margins matter, one should focus on data that these margins can speak to.

There are estimation approaches that utilize aggregate data and relationships between model and data that are common across models (see, e.g., the discussion in Arkolakis et al. (2011) and Caliendo and Parro (2011)). These are not invalid approaches and provide additional evidence on the elasticity of trade. However, aggregate, model independent approaches lack the ability to discriminate across models (which is of interest in its own right) and empirically do not appear to be that informative about the trade elasticity. To make this later point, we combine a "macro" based estimation of the trade elasticity with our "micro" based estimation of the trade elasticity, by combining our moment conditions with those utilized by Caliendo and Parro (2011) (i.e. triple differenced trade flows and tariffs and a model-independent orthogonality condition). Thus, depending upon aspects of the data, this estimation procedure has the ability to deliver an estimate of the trade elasticity that is common across models. We find that these additional macro moment conditions have no impact on our estimates (see, e.g. third column of Table 3). The reason is that the tariff data is extremely noisy and has little explanatory power for trade flows. Thus, any loss in fit or bias by not satisfying the best fit between trade flows and tariff elasticity estimate is trivial, and hence these moment conditions do not modify our results.

2. Fixed Variety Models of Trade

2.1. Armington

The simplest model of international trade that yields a gravity equation is the Armington model outlined in Anderson and van Wincoop (2003). The framework features N countries populated by consumers with constant elasticity of substitution (CES) preferences and tradable goods that are differentiated by the country of origin. Perfect competition among producers and product differentiation by origin imply that each good is purchased in each destination at a price that equals the marginal cost of production and delivery of the good there.

We consider a more empirically-relevant version of the Armington model that features product differentiation within and across countries.[2] Throughout the paper, let i denote the source country and n the destination. We allow each country to produce an exogenously given measure of tradable goods equal to $1/N$. We assume that products are differentiated. Within each country n, there is a measure of consumers L_n. 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 $\rho > 1$,

$$
U_n = \left[\sum_{i=1}^{N} \int_0^{1/N} x_{ni}(j)^{\frac{\rho-1}{\rho}} dj \right]^{\frac{\rho}{\rho-1}}.
$$

To produce quantity $x_{ni}(j)$ in country i, a firm employs labor using a linear production function with productivity $T_i^{1/\theta}$, where T_i is a country-specific technology parameter and $\theta = \rho - 1$. The perfectly competitive firm from country i incurs a marginal cost to produce good j of $w_i/T_i^{1/\theta}$, 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, $\tau_{ni} \leq \tau_{nk}\tau_{ki}$. We maintain this assumption about the nature of iceberg trade costs in all the models that we describe below.

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

$$
p_{ni}(j) = \frac{\tau_{ni} w_i}{T_i^{1/\theta}}.
$$

Since goods are differentiated, consumers in destination n buy all products from all sources and pay $p_{ni}(j)$ for good j from i. Hence, all tradable goods are traded among the N markets and

---

[2]This model yields identical aggregate predictions to the simple Armington model, but it accommodates a greater number of goods than countries, which is an observation that holds true in the data.

consumers buy a unit measure of goods.

2.2. Eaton and Kortum (2002)

We now outline the environment of the multi-country Ricardian model of trade introduced by Eaton and Kortum (2002)—EK. As in the Armington model, there is a continuum of tradable goods indexed by $j \in [0, 1]$. Preferences are represented by the following utility function

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

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

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

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 $w_i/z_i(j)$. 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 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

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

Hence, the EK framework introduces endogenous tradability into the Armington model outlined above. In particular, countries export only a subset of the unit measure of tradable goods for which they are the most efficient suppliers.

2.3. Bernard, Eaton, Jensen, and Kortum (2003)

Bernard et al. (2003)—BEJK—introduce Bertrand competition into EK's model. The most important implication from this extension is that individual good prices differ from the EK model.

Let $c_{kni}(j) \equiv \tau_{ni} w_i / z_{ki}(j)$ be the cost that the k-th most efficient producer of good j in country i faces in order to deliver a unit of the good to destination n. With Bertrand competition, as with perfect competition, the low-cost supplier of each good serves the market. For good j in market n, this supplier has the following cost $c_{1n}(j) = \min_i \{c_{1ni}(j)\}$. This supplier is constrained not to charge more than the second-lowest cost of supplying the market, which is $c_{2n} = \min \{c_{2ni^*}(j), \min_{i \neq i^*} \{c_{1ni}(j)\}\}$, where i^* satisfies $c_{1ni^*}(j) = c_{1n}(j)$. Hence, the price of good j in market n is

$$p_n(j) = \min \{c_{2n}(j), \bar{m} c_{1n}(j)\},$$

where $\bar{m} = \rho/(\rho - 1)$ is the Dixit-Stiglitz constant mark-up.

Finally, for each country i, productivity, $z_{ki}(j)$ for $k = 1, 2$ is drawn from

$$G_i(z_1, z_2) = \left[1 + T_i(z_2^{-\theta} - z_1^{-\theta})\right] \exp\left(-T_i z_2^{-\theta}\right).$$

Hence, the BEJK model features a key additional component relative to the EK framework: the existence of variable mark-ups. In particular, the most efficient suppliers in this model enjoy the highest mark-ups.

2.4. Trade Flows, Aggregate Prices, and Welfare

The models described above produce identical aggregate outcomes, even though they feature different micro-level behavior. In particular, under the parametric assumption made above, the models yield the same expressions for trade flows, price indices (up to a constant scalar), and welfare gains from trade. Proposition 1 summarizes the result.

Proposition 1 *Given the functional forms for productivity $F_i(\cdot)$, $G_i(\cdot)$ for all $i = 1, ..., N$,*

a. The share of expenditures that n spends on goods from i, X_{ni}/X_n, predicted by the models is

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

b. *The CES exact price index for destination* n, P_n, *predicted by the models is*

$$P_n \propto \Phi_n^{-\frac{1}{\theta}}, \quad where \quad \Phi_n = \sum_{k=1}^{N} T_k (\tau_{nk} w_k)^{-\theta}. \tag{2}$$

c. *The percentage compensation that a representative consumer in* n *requires to move between two trading equilibria predicted by the models is*

$$\frac{P'_n}{P_n} - 1 = 1 - \left(\frac{X'_{nn}/X'_n}{X_{nn}/X_n} \right)^{\frac{1}{\theta}}. \tag{3}$$

We prove a. and b. for the Armington model in the Appendix. The results for the EK and BEJK models are derived in the respective papers. The proof of part c. can be found in Arkolakis et al. (2011). Across the models, the welfare gains from trade are essentially captured by changes in the CES price index that a representative consumer faces. Using the objects from the Proposition above, it is easy to relate the price indices to trade shares and the parameter θ. In particular, expressions (1) and (2) allow us to relate trade shares to trade costs and the price indices of each trading partner via the following equation

$$\frac{X_{ni}/X_n}{X_{ii}/X_i} = \frac{\Phi_i}{\Phi_n} \tau_{ni}^{-\theta} = \left(\frac{P_i \tau_{ni}}{P_n} \right)^{-\theta}, \tag{4}$$

where $\frac{X_{ii}}{X_i}$ is country i's expenditure share on goods from country i, or its home trade share, and $\Phi_n = \sum_k$ The welfare equation follows trivially from this expression.

2.5. The Elasticity of Trade

The key parameter determining trade flows (equation (4)) and welfare (equation (3)) is θ. To see the parameter's importance for trade flows, take logs of equation (4) yielding

$$\log \left(\frac{X_{ni}/X_n}{X_{ii}/X_i} \right) = -\theta \left[\log (\tau_{ni}) - \log(P_i) + \log(P_n) \right]. \tag{5}$$

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 to 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 easy to demonstrate that (3) implies that θ

represents the inverse of the elasticity of welfare with respect to domestic expenditure shares

$$\log(P_n) = -\frac{1}{\theta}\log\left(\frac{X_{nn}}{X_n}\right).$$ (6)

Hence, decreasing the domestic expenditure share by one percent generates a $(1/\theta)/100$-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.

The challenge with estimating the trade elasticity is that one must *separately* disentangle θ from trade costs, which are not observed. To overcome this challenge, we theoretically show that micro-level cross-country price variation or price gaps identifies the trade elasticity. More importantly, different models represent different mappings between price gaps and the trade elasticity, which implies that in order to match the observed price variation in the data, different models require different trade elasticities.

3. Price Gaps and Trade Elasticities

This section shows that different trade models have different implications for the distribution of price gaps. This implies that, given data on trade flows and micro-level prices, different models have different trade elasticities and thus different welfare gains from trade.

To illustrate the argument theoretically, we focus on a symmetric, two country world with the countries denoted as home (h) and foreign (f). Throughout this section, we constrain all the models to have the same aggregate trade flow predictions. In a symmetric two-country world, the only two parameters that govern trade flows (as well as prices) in the models above are θ and τ.[3] Throughout this section, we assume that θ and τ are constrained to fit the bilateral trade share. This implies that if $\theta > \theta'$, then $\tau < \tau'$.

Given the constraint that all models must imply the same aggregate trade flows, we then make the following argument. First, we describe properties of the distribution of relative prices at the micro-level. Second, we order these distributions (across models and for different θs) by first-order stochastic dominance and, in turn, by the largest order statistic or maximal price gap in a finite sample. This then allows us to order the implied θs across models given the observed largest order statistic from a sample of data. Finally, we show that these models are testable in the sense that they provide refutable restrictions on other moments from the price

---

[3]To see this, from the first equality in expression (4) and the definition of Φ_n in (2), notice that, given trade share data and a value for θ, $\log \tau$ can be computed immediately in a symmetric two-country world.

gap distribution.

3.1. Ordering Price Distributions and the Largest Order Statistic Across Models

In this section, we focus on deviations from the law of one price at the micro level, i.e. for an individual variety. Specifically, we focus on the logged price gap between the home and the foreign country:

Definition 1 *For any micro-level good ℓ, define the random variable that equals the logged price gap between the home and the foreign country as $\log \tilde{P}_h \equiv \log P_{h,\ell} - \log P_{f,\ell}$.*

The distribution of price gaps at the micro-level across countries is the key object of interest. Given a model $\mathcal{M}$, denote the cumulative density function of log price gaps, conditional on the price gap being positive as:

$$\forall \text{ models } \mathcal{M}, \quad G_{\mathcal{M}}(\log \tilde{p}_h, \theta) = \begin{cases} 0 & : \log \tilde{P}_h < 0 \\ \log \tilde{p}_h \in [0,1] & : 0 \leq \log \tilde{P}_h \leq \log \tau \\ 1 & : \log \tau < \log \tilde{P}_h. \end{cases} \tag{7}$$

Note that we index the density by parameter θ. This indexing is sufficient because other parameters are either the same across models (i.e. the technology parameters), or are determined by the value θ, e.g. θ determines τ so aggregate trade flows are held constant.

A key property of this density is that all price gaps lie below $\log \tau$. The reason is the following: suppose that $\log P_h(\ell) - \log P_f(\ell) > \log \tau$, then an arbitrage opportunity exists as an agent could import good ℓ from the foreign country at a lower price. Thus, this inequality places an upper bound on the possible price gaps that are observable. This observation is important because the support of the density depends on θ. Different θs imply different τs because aggregate trade constrains the values that these parameters can take. This property allows us to make conclusions as to how parts of the density shift in a model with parameter θ relative to the model with parameter θ'.

How the density shifts as θ changes at any price gap in certain models is analytically intractable (though numerically verifiable). Thus, we assert that the following regularity conditions holds on the price gap distribution.

Regularity Condtion 1 (No Crossing.) *Let the density $G_{\mathcal{M}}(\log \tilde{p}_h, \theta)$ satisfy the following no-crossing property. If $\theta' > \theta$, then for $\log \tilde{p}_h < \log \tau'$:*

$$\text{Prob}_{\mathcal{M}}(\log \tilde{P}_h < \log \tilde{p}_h, \ \theta) \ \leq \ \text{Prob}_{\mathcal{M}}(\log \tilde{P}_h < \log \tilde{p}_h, \ \theta'). \tag{8}$$

This property says the following: Fix a model $\mathcal{M}$. If we lower the elasticity from θ' to θ and constrain aggregate trade to stay the same, the probability of seeing a particular price gap cannot increase. So as the upper end of the support increases from $\log \tau'$ to $\log \tau$, it can not induce a "cross" in the density.

This property clearly holds in the Armington model. In the Armington model, all goods are traded and thus the probability mass of the price gaps lies completely at $\log \tau'$. If we lower the elasticity from θ' to θ, then all the probability mass shifts rightward to $\log \tau$. For the EK model, this holds as well (see online appendix of Simonovska and Waugh (2011)). For the BEJK model, analytically verifying this is difficult; we have, however, verified that this holds numerically for all relevant parameter values.

Below, we define the data features we will focus on. Specifically, the maximal price gap from a sample of price gaps:

Definition 2 *Consider a finite, random sample of positive price gaps* $\log \tilde{P}_{h,1}, \log \tilde{P}_{h,2}, \dots \log \tilde{P}_{h,L}$ *which are ordered from smallest to largest; The maximal price gap in the sample of L goods is* $\log \tilde{P}_{h,L:L}$. *Given a model* $\mathcal{M}$, *parameter* θ, *and density* $G_{\mathcal{M}}(\log \tilde{p}_h, \theta)$, *the expected maximal price gap or largest order statistic is* $\mathbb{E}_{\mathcal{M}}\left(\log \tilde{P}_{h,L:L}, \theta \right)$, *where the subscript* $\mathcal{M}$ *indexes the model and the dependence on the parameter* θ *is noted.*

Focusing on the maximal price gap has several appealing features. First, it has some history of thought as it has been used in Eaton and Kortum (2002) and Simonovska and Waugh (2014) in the estimation of θ. Second, order statistics encode much information about the underlying density. For example, one can show that a sequence of largest order statistics (that is for $L = 1, 2, \dots$) completely characterizes the density G (see, e.g., Arnold, Balakrishnan, and Nagaraja (1992)). We will never have access to an infinite sequence of largest order statistics. However, by utilizing recurrence relationships between largest order statistics of different sample sizes (again see Arnold et al. (1992)), we can construct testable restrictions for each model.

Lemma 1 connects the expected maximal price gap with a stochastic dominance relationship in the density.

Lemma 1 *If the price density,* $G_{\mathcal{M}}(\log \tilde{p}_h, \theta)$, *first-order stochastically dominates the price density* $G_{\mathcal{M}'}(\log \tilde{p}_h, \theta')$, *then* $\mathbb{E}_{\mathcal{M}}\left(\log \tilde{P}_{h,L:L}, \theta \right) > \mathbb{E}_{\mathcal{M}'}\left(\log \tilde{P}_{h,L:L}, \theta' \right)$.

Proof: If a random variable first-order stochastically dominates another, the same ranking holds for the distribution of the variables' maxima, that is $G_{\mathcal{M}}(\log \tilde{p}_h, \theta)^L >_{fosd} G_{\mathcal{M}'}(\log \tilde{p}_h, \theta')^L$. This property implies that the expected maximal price gap under model $\mathcal{M}$ is larger than in model $\mathcal{M}'$. ∎

The result here is that if one density dominates another density, this implies that the maximal price gap must be larger in the model that has the density which dominates. Application of

Lemma 1 and the regularity condition above allows us to rank densities according to θ for a given model $\mathcal{M}$.

Lemma 2 *For a given model, $\mathcal{M}$, the parameter θ indexes the price density by first-order stochastic dominance, and strictly indexes the expected maximal price gap. That is, if $\theta' > \theta$ then $G_{\mathcal{M}}(\log \tilde{p}_h, \theta)$ $>_{fosd} G_{\mathcal{M}}(\log \tilde{p}_h, \theta')$ and $\mathbb{E}_{\mathcal{M}}\left(\log \tilde{P}_{h,L:L}, \theta\right) > \mathbb{E}_{\mathcal{M}}\left(\log \tilde{P}_{h,L:L}, \theta'\right)$.*

Proof: First, recall that, when $\log \tau$ and θ are constrained to fit bilateral trade shares, $\theta' > \theta$ implies that $\tau' < \tau$. Then, $\tau' < \tau$ and Regularity Condition 1 implies that, for $\log \tilde{p}_h < \log \tau'$, the cumulative probability in the model with θ is weakly less than the cumulative probability in the model with θ'. For $\log \tilde{p}_h = \log \tau'$, the cumulative probability in the model with θ' is equal to one, and because there is strictly positive probability mass in the region $\log \tilde{p}_h > \log \tau'$ for the model with θ, this implies a strict inequality in probability mass at $\log \tilde{p}_h = \log \tau'$. This implies first-order stochastic dominance. From Lemma 1, the ordering of the order statistics follows. ∎

Lemma 2 allows us, for a given model, $\mathcal{M}$, an ability to strictly rank expected maximal price gaps by θ. One implication of this result is that the expected maximal price gap uniquely identifies the θ parameter. We exploit this result in Proposition 3. The second implication is that it assists us in showing that if all models have the same expected maximal price gap, then they have different θ's. We state the result formally below:

Proposition 2 *Consider the models $G_{\mathcal{ARM}}(\log \tilde{p}_h, \theta_{\mathcal{ARM}})$, $G_{\mathcal{EK}}(\log \tilde{p}_h, \theta_{\mathcal{EK}})$, $G_{\mathcal{BJEK}}(\log \tilde{p}_h, \theta_{\mathcal{BJEK}})$. If the expected maximal price gaps are the same in all models:*

$$\mathbb{E}_{\mathcal{ARM}}\left(\log \tilde{P}_{h,L:L}, \theta_{\mathcal{ARM}}\right) = \mathbb{E}_{\mathcal{EK}}\left(\log \tilde{P}_{h,L:L}, \theta_{\mathcal{EK}}\right) = \mathbb{E}_{\mathcal{BEJK}}\left(\log \tilde{P}_{h,L:L}, \theta_{\mathcal{BEJK}}\right), \tag{9}$$

then

$$\theta_{\mathcal{ARM}} > \theta_{\mathcal{EK}} \text{ and } \theta_{\mathcal{ARM}} > \theta_{\mathcal{BEJK}}.$$

Moreover, if $G_{\mathcal{EK}}(\log \tilde{p}_h, \theta) >_{fosd} G_{\mathcal{BJEK}}(\log \tilde{p}_h, \theta)$, then

$$\theta_{\mathcal{ARM}} > \theta_{\mathcal{EK}} > \theta_{\mathcal{BEJK}}.$$

Proof: The proof proceeds by contradiction in the following two cases with the focus on the EK model. The same argument applies with respect to the BEJK model.

- Case #1. Suppose not and that $\theta_{\mathcal{ARM}} = \theta_{\mathcal{EK}} = \theta$. First, note that for the same θ,

$$G_{\mathcal{ARM}}(\log \tilde{p}_h, \theta) >_{fosd} G_{\mathcal{EK}}(\log \tilde{p}_h, \theta), \tag{10}$$

11

because for any $\log \tilde{p}_h < \log \tau$, the probability mass in the Armington model is zero as all goods are traded and hence all positive price differences are equal to the trade friction. This is strictly less than in the EK model as for any $\log \tilde{p}_h < \log \tau$, the probability mass is greater than zero because there are non-traded goods. This argument implies the distribution of price gaps in the Armington model first-order stochastically dominates the distribution in the EK model. Application of Lemma 1 then implies the expected maximal price gaps should be ranked, which is a contradiction.

- Case #2. Suppose not and that $\theta_{\mathcal{EK}} > \theta_{\mathcal{ARM}}$. First, notice that this implies that $G_{\mathcal{ARM}}(\log \tilde{p}_h, \theta_{\mathcal{ARM}})$ $>_{fosd} G_{\mathcal{ARM}}(\log \tilde{p}_h, \theta_{\mathcal{EK}})$ from Lemma 2. This then implies that,

$$\mathbb{E}_{\mathcal{ARM}}\left(\log \tilde{P}_{h,L:L}, \theta_{\mathcal{ARM}}\right) > \mathbb{E}_{\mathcal{ARM}}\left(\log \tilde{P}_{h,L:L}, \theta_{\mathcal{EK}}\right) > \mathbb{E}_{\mathcal{EK}}\left(\log \tilde{P}_{h,L:L}, \theta_{\mathcal{EK}}\right) \quad (11)$$

and the last inequality follows from the observation made in Case #1 above, that for the same θ, the Armington price gap distribution strictly dominates the EK price gap distribution.

Finally note that the same exact arguments apply for the BEJK model, where some goods are non-traded as in the EK model. Thus, if the Armington model, EK model, and BEJK model, all have the same expected maximal price gap, then the elasticity in the Armington model must be strictly greater than the elasticity in EK or BEJK. ∎

Figure 1 illustrates the intuition behind this result. The top panel plots the cumulative distribution function of log price gaps for the Armington (solid blue line), the EK (dashed red line) and the BEJK (solid black line) model—for the same θ.

In the Armington model, notice that all the mass lies on $\log \tau$. This is because all goods are traded, i.e. there is no active extensive margin in the Armington model. If the home country imports the good from the foreign country, then the price gap exactly reflects the trade friction. The distribution of price gaps in the EK model is different as it places a positive mass in the region between zero and the trade friction and hence the density in EK is stochastically dominated by that of Armington. This is because there are endogenously non-traded goods, i.e. there is an active extensive margin. In the EK model, there are instances where consumers in *both* countries find it cheaper to consume a variety from the local producer rather than importing the good. If a good is non-traded, the price difference across locations reflects differences in costs which are strictly less than the trade friction. This implies a positive mass in the region between zero and the trade friction.

Proposition 2 says that because the price gap density in Armington dominates the density in EK, the only way for the expected maximal price gap to be the same is if the EK model has a lower θ. The bottom panel of Figure 1 illustrates this by plotting the densities when they have
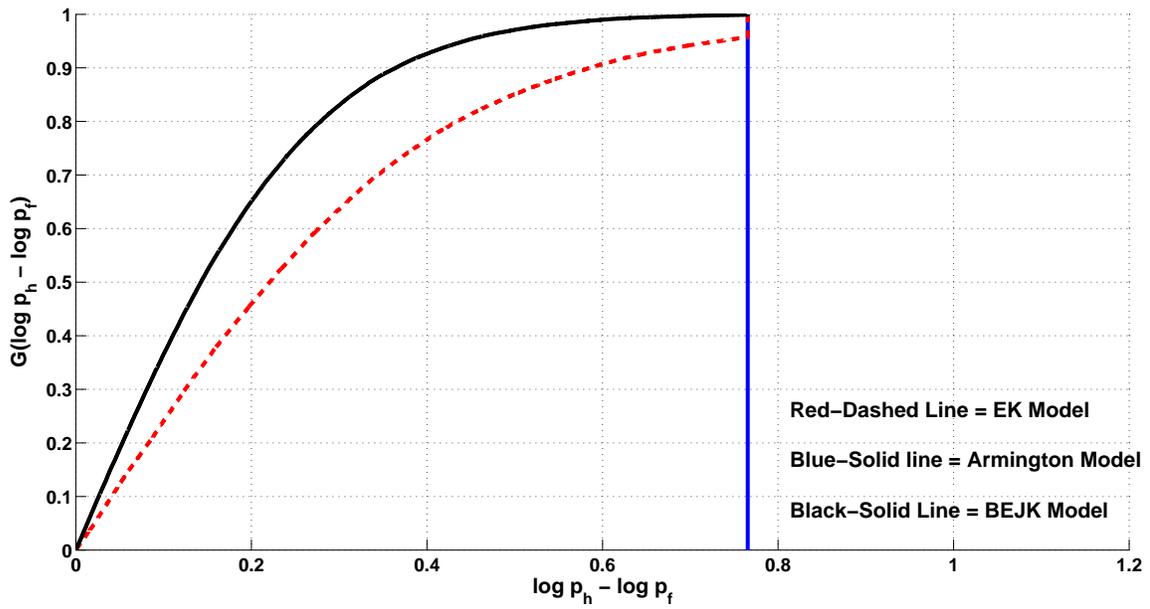
(a) **Same** θ
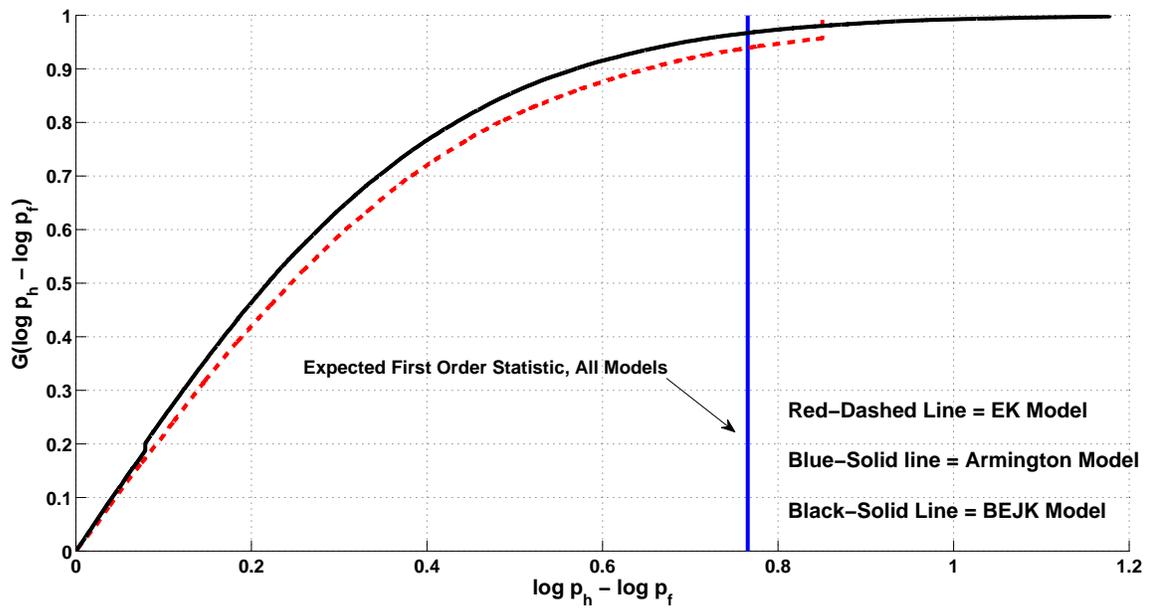
(b) **Same Largest Order Statistic**

Figure 1: Price Gap Distribution: Armington, EK, BEJK.

13

the same expected largest order-statistic. Here, the EK density is spread to the left with a larger upper bound to ensure that the largest order statistic is the same as in the Armington model.

What about the relationship between EK and BEJK? Generically, we can say the following: if the density in EK strictly stochastically dominates BEJK and the largest order statistics are the same, then by logic of Proposition 2 this implies that $\theta_{\mathcal{EK}} > \theta_{\mathcal{BJEK}}$ and vice versa. A stronger statement is currently not possible as it is difficult to theoretically construct the BEJK price gap distribution because of the presence of variable markups. However, in the case in which $\rho = 1$ and the Dixit-Stiglitz mark-up is infinite, we can show that the probability that the log price gap reaches the boundaries is strictly lower in the BEJK than in the EK model. This implies that the mass has to be distributed between these two end-points—meaning, in the non-traded good region. However, the exact shape of the distribution cannot be easily characterized.

We verify computationally that distributions are indeed ranked, with EK stochastically dominating BEJK. The top panel in Figure 1 makes clear that for the same θ, the price gap density in EK dominates the density in BEJK. More generally we have verified that this relationship holds numerically for all relevant parameter values. And our estimates of θ in a multi-country setting with asymmetries always result in a ranking of $\theta_{\mathcal{EK}} > \theta_{\mathcal{BJEK}}$. Furthermore, the bottom panel of Figure 1 verifies that if the largest order statistic is the same in EK and BEJK, then the BEJK model must have a lower theta. This can be seen as the density in BEJK is spread even further to the left with a larger upper bound to ensure that the largest order statistic equals that in the EK and Armington models.

Variable markups and how they correlate with costs in the BEJK model explain why the price gap density in EK strictly dominates the density in BEJK.[4] In the EK model, price gaps correspond to cost gaps and these cost gaps are identical in the two models. In BEJK, cost gaps do not correspond with price gaps as producers are able to price at a markup over marginal cost depending on other latent competitors. Thus, the price gap in BEJK reflects both markups, cost differences (if the good is non-traded), and trade frictions (if the good is traded). In BEJK, markups are negatively correlated with marginal costs. This results in the EK distribution being a mean preserving spread of the price gap distribution in BEJK.

Consider the following example to illustrate this point. Take a producer in the foreign country who has a very high productivity and is the lowest cost producer in his country of origin as well as the home country. This producer will likely charge a high markup (call it m') in the foreign country because she has a very low cost relative to her latent competitors as that is her country of origin. In contrast, she will likely charge a relatively lower markup (call it m'') in the home country because her cost advantage is eroded from the trade friction faced when

---

[4]We suspect in a model where markups were positively correlated with marginal costs, the ranking of the price gap densities would be reversed.

exporting to the home country. This implies that the price gap, $\log \tilde{p}_h$, will be *less than* the trade friction even though the good is traded, i.e. the price gap equals $m'' - m' + \log \tau < \log \tau$. This observation implies that the BEJK distribution will have less mass around the log of the trade friction because of producers differentially marking up their products across locations.

3.2. Estimating Trade Elasticities and the Gains from Trade

We now connect Proposition 2 with an econometrician's inference about θ given a sample of data (i.e. aggregate trade flows and a sample of micro-level prices) and discuss how this inference depends upon assumptions about the underlying model. Proposition 3 suggests a simple method of moments estimator and states that the estimated θ will (i) depend on the econometrician's assumptions about the underlying data generating process and (ii) will be ordered across models.

Proposition 3 *Consider the following estimator of θ that chooses $\hat{\theta}$ to minimize the distance between the observed largest order statistic $\log \tilde{P}_{h,L:L}$ and the expected largest order statistic $\mathbb{E}_{\mathcal{M}}\left(\log \tilde{P}_{h,L:L}\right)$. That is,*

$$\hat{\theta}_{\mathcal{M}} = \arg\min g(\theta_{\mathcal{M}})' g(\theta_{\mathcal{M}})$$

$$\text{where} \quad g(\theta_{\mathcal{M}}) = \left(\log \tilde{P}_{h,L:L} - \mathbb{E}_{\mathcal{M}}\left(\log \tilde{P}_{h,L:L}, \ \theta_{\mathcal{M}}\right)\right).$$

Then:

1. *The estimate $\hat{\theta}_{\mathcal{ARM}}$ is strictly greater than the estimate $\hat{\theta}_{\mathcal{EK}}$ and $\hat{\theta}_{\mathcal{BEJK}}$.*

2. *The estimate of the welfare cost of autarky in the Armington model is strictly less than the estimate of the welfare cost of autarky in EK and BEJK.*

3. *If the density in EK stochastically dominates the density in BEJK, then the estimate $\hat{\theta}_{\mathcal{EK}} > \hat{\theta}_{\mathcal{BEJK}}$ and the welfare gains from trade in EK are strictly less than the gains in BEJK.*

Proof: The first statement follows from: (i) Lemma 2 implies that the order statistic uniquely determines the θ for a given model and (ii) Proposition 2 implies that across models, the θ must differ. The second statement follows from the fact that equation (6) is the same across models and that, by construction, all models have the same predictions for aggregate trade flows, but different θ. ∎

A couple of comments are in order about this result. First, the estimator is quite simple in this context. However, our estimation strategy in Section 4 builds off the intuition from this result. Moreover, we include alternative moment conditions to enrich the estimation, to be able to test

over-identifying restrictions, and to exploit empirical regularities we see in the data. We discuss these moment conditions in Section 4.

Second, compare Proposition 3 to the discussion in Section 6 in Arkolakis et al. (2011). In this section, they propose an alternative estimator of the trade elasticity that delivers a common estimate of θ across models. This estimator uses aggregate trade flows and some measure of trade frictions (say tariffs). Given that all these models have the same connection between aggregate trade flows (i.e. equation 5) and assuming that the orthogonality condition is model independent, the estimate of θ as well as the gains from trade will be the same across models.

4. Estimating the Elasticity

This section describes how we estimate the trade elasticity given a data set that features micro-level prices and bilateral trade flows across countries. The basic idea is captured in Proposition 3. In particular, we use moment conditions that compare statistics from a sample of micro-level prices with their expected values from each model, where the latter are functions of θ and depend on assumptions about the model that generated the data. Because the expected values do not have closed form expressions, we use simulation methods to approximate the expected value and we discuss how we are able to simulate prices from the different models. Finally, we provide monte-carlo evidence that our estimation procedure works well.

4.1. Estimation

We will focus on three sets of results (i) an exactly identified case, (ii) an overidentified case with an identity weighting matrix, and (iii) an overidentified case with an optimal weighting matrix. Below we discuss each of these cases in turn.

Exactly Identified Case. In the exactly identified case we will focus on properties of the logarithm of the maximal price gap adjusted by average prices in the two locations. We define the value d_{ni} as

$$d_{ni} = \log \hat{\tau}_{ni} + \log \hat{P}_i - \log \hat{P}_n, \tag{12}$$

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

$$\log \hat{P}_i = \frac{1}{L} \sum_{\ell=1}^{L} \log(p_i(\ell)),$$

and L is the number of micro-level prices in the sample. Like the statistics discussed in Section 3, we focus initially on the maximal price gap. One caveat is that we adjust this by average

16

price differences in the two locations. This is important as it helps control for country specific factors that may be present in the price data (see Simonovska and Waugh (2014) for a detailed discussion).

Given the d_{ni}s from data, we define the following function

$$h_1(\theta, d_{ni}, \mathcal{M}) = d_{ni} - \frac{1}{S}\sum_{s=1}^{S} d_{ni}(\theta, u_s, \mathcal{M}), \tag{13}$$

which compares the observed value d_{ni} to its expected value. Because we do not have a closed form expression for the expected value of d_{ni}, we construct this value via simulation. Given a model $\mathcal{M}$ and a value of θ, we simulate prices, construct synthetic data sets, and then construct simulated values for $d_{ni}(\theta, u_s, \mathcal{M})$. The expected value of d_{ni} is the average across simulations. The next section discusses the simulation procedure in more detail. Finally, note that the subscript 1 distinguishes this function from the vector valued function in the overidentified case.

In the exactly identified case, our estimation is based on the following orthogonality restriction

$$\mathbb{E}(h_1(\theta_o, d_{ni}, \mathcal{M})) = 0. \tag{14}$$

That is, at the true value θ_o, in expectation the average difference between the observed d_{ni} and its expected value should be equal to zero. An estimate of θ is found by computing the sample average of $h_1(\theta, \mathcal{M})$ as

$$g(\theta, \mathcal{M}) = \left(\frac{1}{N^2 - N}\right)\sum_{n}\sum_{i} h_1(\theta, d_{ni}, \mathcal{M}) \tag{15}$$

where $N^2 - N$ is the number of d_{ni}s we can construct. Our estimate of θ is the value that minimizes the quadratic form of $g(\theta, \mathcal{M})$. Or mathematically:

$$\hat{\theta}(\mathcal{M}) = \arg\min_{\theta} g(\theta, \mathcal{M})'g(\theta, \mathcal{M}). \tag{16}$$

Overidentified Case. The overidentified case focuses on two additional moments: (i) the price gap in the 85th percentile adjusted by average prices and (ii) the covariance of d_{ni} with the logarithm of bilateral distance.

The focus on the price gap in the 85th percentile is a way to incorporate more information about the underlying distribution of price gaps and, hence, information about the underlying model. For example, as discussed in Section 3, the Armington model places a restriction on the data that the max should be the same as the price gap in the 85th percentile. The price gap in the 85th percentile is not the only statistic that has this property. We explored other percentiles (e.g.

90 or 75) and found virtually no effect on our results.

The covariance between the d_{ni} statistic and the logarithm of distance was chosen for two reasons. First, as we show in Section 5.1, there is a strong correlation between maximal price gaps and distance. Given the strong role that distance plays in explaining trade flows, we feel this is a natural statistic to target. Second, this also helps guard against estimating the elasticity off of measurement error in the data. If price gaps were purely a result of measurement error, then they should not be correlated with distance. Thus by focusing on the covariance of price gaps and distance, we are focusing on a moment that measurement error should not affect.

Given these moments, the function $h(\theta, \mathbf{d_{ni}}, \mathcal{M})$ is now a 3×1 vector. We are slightly abusing notation here by using the bold-face value $\mathbf{d_{ni}}$ to denote $h(\theta, \mathbf{d_{ni}}, \mathcal{M})$ dependence on the data. Our orthogonality restriction is

$$\mathbb{E}(h\left(\theta_o, \mathbf{d_{ni}}, \mathcal{M}\right)) = 0. \tag{17}$$

An estimate of θ is found by computing the sample average of $h(\theta, \mathcal{M})$ as

$$g(\theta, \mathcal{M}) = \left(\frac{1}{N^2 - N}\right) \sum_n \sum_i h(\theta, \mathbf{d_{ni}}, \mathcal{M}), \tag{18}$$

and choosing the θ to minimize the quadratic form

$$\hat{\theta}(\mathcal{M}) = \arg \min_\theta g(\theta, \mathcal{M})' W^{-1} g(\theta, \mathcal{M}), \tag{19}$$

where W is a positive definite weighting matrix.

For the weighting matrix, we will present two sets of results: one using the identity matrix and another using the optimal weighting matrix suggested by Gouriéroux and Monfort (1996) and described in Adda and Cooper (2003). Because the optimal weighting matrix depends on our estimate of θ, we use a continuous-updating estimator of W which continually updates the weighting matrix within the minimization routine (see e.g. the discussion of this estimator in Hansen, Heaton, and Yaron (1996)).

We compute standard errors using a parametric bootstrap technique (see e.g. Davison and Hinkley (1997)). We add error terms to the trade data (given our parametric assumption on the error term in (21)), simulate a sample of prices given our estimate of θ and the assumed underlying model, and then implement our estimation routine on the simulated data. We repeat this procedure 999 times to construct 90-10 percentile confidence intervals.

In the instances where we use an optimal weighting matrix, we also perform tests of overiden-

tifying restrictions. Specifically, we construct the test statistic

$$(N^2 - N)g(\hat{\theta}, \mathcal{M})'W^{-1}g(\hat{\theta}, \mathcal{M}) \qquad (20)$$

which is the standard "J-statistic" used to test the null-hypothesis that the data generating process is correct. Asymptotically, this test statistic is distributed chi-squared with a certain number of degrees of freedom, though in small samples this may not provide an accurate approximation (see e.g. Hansen et al. (1996)). To avoid these difficulties, we use a parametric bootstrap to construct the finite sample distribution of the test statistic in (20) under the null hypothesis that the data generating process is correct (see Davison and Hinkley (1997)).

Relation to Previous Work. The statistic d_{ni} that we use to estimate θ is developing a history of thought, so a couple of comments are necessary. First, EK used this statistic to directly estimate θ. If the maximal price difference accurately revealed the trade friction, then by examining (4) one can see how averaging across normalized trade shares relative to the average d_{ni} yields an estimate of θ. Using this method, they arrived at the value 8.28. As Simonovska and Waugh (2014) point out, the maximal price gap generally underestimates the trade friction, and hence overestimates the value of θ. Simonovska and Waugh (2014) used indirect inference to correct this bias by matching the observed EK estimate of θ to the model implied value. Using this method, they arrived at the value around four.

We depart from our previous work for the following reasons. First, the new procedure fits transparently within the GMM framework which allows us to formally test the overidentifying restrictions of various models and allows us to incorporate additional moment conditions that existing work uses to estimate the trade elasticity (see e.g. Caliendo and Parro (2011)). Second, we avoid a difficulty that arose when using both simulated trade flows and prices to construct the model implied EK estimate. One difficulty was that the simulated trade flows depended on the number of goods in the economy and for the simulated trade flows to mimic observed trade flows accurately this had to be a very large number. By just simulating prices, we avoid the dependence on the number of goods in the economy and achieve a very large speed up in computation time. Finally, in the exactly identified case our new approach gives nearly the exact same answer for the EK model as the approach used in Simonovska and Waugh (2014), which is reassuring.

4.2. Simulation Approach

A key step in the estimation procedure is the approximation of $\mathbb{E}(d_{ni})$ with $\frac{1}{S}\sum_{s=1}^{S} d_{ni}(\theta, u_s, \mathcal{M})$. Below we describe in detail how we simulate prices from each model and in turn how we approximate $\mathbb{E}(d_{ni})$.

There are basically three steps. The distribution of prices depends on θ and the country-specific

technology parameters and bilateral trade costs. We use the "gravity" equation implied by all models to estimate the technology parameters and trade costs (up to a scalar θ) from trade flow data. This ensures that all models will have the same aggregate predictions for trade flows. Second, we show how the estimates from the gravity equation are sufficient to simulate micro-level prices given a specified model. Third, we specify a protocol as to how the data is sampled. These three steps allow us to create synthetic data sets from which we can then approximate $\mathbb{E}(d_{ni})$.

Step 1.—We estimate the parameters for the country-specific productivity distributions and trade costs from bilateral trade-flow data. We follow closely the methodologies proposed by Eaton and Kortum (2002) and Waugh (2010). First, we derive the gravity equation from expression (1) by dividing the bilateral trade share by the importing country's home trade share,

$$\log\left(\frac{X_{ni}/X_n}{X_{nn}/X_n}\right) = S_i - S_n - \theta \log \tau_{ni} + \nu_{ni}, \tag{21}$$

where S_i is defined as $\log\left[T_i w_i^{-\theta}\right]$. Note that (21) is a different equation from expression (4), which is derived by dividing the bilateral trade share by the exporting country's home trade share, and is used to estimate θ. S_i's are recovered as the coefficients on country-specific dummy variables given the restrictions on how trade costs can covary across countries. We assume that ν_{ni} reflects other factors and is orthogonal to the regressors and normally distributed with mean zero and standard deviation σ_ν. Following the arguments of Waugh (2010), trade costs take the following functional form

$$\log(\tau_{ni}) = \rho \log m_{ni} + b_{ni} + ex_i. \tag{22}$$

Here, trade costs are a logarithmic function of distance, where m_{ni} is the distance in miles between country n and i and ρ is the distance elasticity. 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 level to vary depending upon the exporter. We use least squares to estimate equations (21) and (22).

Step 2.—The parameter estimates obtained from the first-stage gravity regression are sufficient to simulate micro-level prices in each model up to a constant, θ.

Step 2a.—The simulation of the Armington model is simple as there is no micro-level heterogeneity in this framework. $\exp(S_i^{-\theta})$ represents the common marginal cost of production across all producers in country i.

Step 2b.—To simulate micro-level prices in the EK model we follow closely Simonovska and Waugh (2014). Notice that for any good j, the model implies that $p_{ni}(j) = \tau_{ni} w_i / z_i(j)$. 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 **Simonovska and Waugh (2014)**, we show that u_i is distributed according to:

$$M_i(u_i) = \exp\left(-\tilde{S}_i u_i^{-\theta}\right), \quad \text{with} \quad \tilde{S}_i = \exp(S_i) = T_i w_i^{-\theta}. \tag{23}$$

Thus, having obtained estimates of S_i from the gravity regression, we can simulate the inverse of marginal costs. To construct prices at which goods are traded, we take the inverse marginal costs that are drawn from the country-specific distributions above and are assigned to each good. Then, for each importing country and each good, the lowest-cost supplier across countries is found and realized prices are recorded.

Step 2c.—To see that we can simulate micro-level prices in the BEJK model as a function of θ only, we draw on an argument in BEJK. To simulate their model, BEJK reformulate the model in terms of efficiency. In particular, given two productivity draws for good j in country i, $z_{1i}(j)$ and $z_{2i}(j)$, they define the following objects

$$u_{1i}(j) = T_i z_{1i}(j)^{-\theta}$$
$$u_{2i}(j) = T_i z_{2i}(j)^{-\theta}$$

BEJK demonstrate that these objects are distributed according to

$$Pr[u_{1i} \leq u_1] = 1 - \exp(-u_1)$$
$$Pr[u_{2i} \leq u_2 | u_{1i} = u_1] = 1 - \exp(-u_2 + u_1)$$

To simulate trade flows and prices from the model, we define the following variables

$$v_{1i}(j) = \left(\frac{u_{1i}(j)}{T_i w_i^{-\theta}}\right)$$
$$v_{2i}(j) = \left(\frac{u_{2i}(j)}{T_i w_i^{-\theta}}\right)$$

Applying the pdf transformation rule, it is easy to demonstrate that $v_{1i}(j)$ is distributed according to

$$Pr[v_{1i} \leq v_1] = 1 - \exp(-\tilde{S}_i v_1). \tag{24}$$

Similarly,

$$Pr[v_{2i} \leq v_2 | v_{1i} = v_1] = 1 - \exp(-\tilde{S}_i v_2 + \tilde{S}_i v_1), \tag{25}$$

where $\tilde{S}_i$ is defined above.

Thus, to simulate the model we draw minimum unit costs from (24), and conditional on these draws, we draw the second lowest unit costs from (25). Hence, having obtained the estimates of S_i from the first-stage gravity regression, we can simulate the inverse of marginal costs.

To construct prices at which goods are traded, we take the inverse marginal costs that are drawn from the country-specific distributions above and are assigned to each good. Then, for each importing country and each good, the two lowest-cost suppliers across countries are found and realized prices are recorded.

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. The three models give a natural common basket of goods to be priced across countries. In these models, agents in all countries consume all goods that lie within a fixed interval, $[0, 1]$. Thus, we consider this common list in the simulated models and we randomly sample the prices of its goods across countries, in order to approximate trade barriers, much like it is done in the data.

These steps then provide us with an artificial data set of micro-level prices that mimic their analogs in the data. Given this artificial data set, we can then compute moments—as functions of θ—and compare them to the moments in the data.

4.3. Performance on Simulated Data

Estimation Results With Artificial Data, Underlying $\theta = 4$

	Model	Estimate of θ	Mean d_{ni}	$\text{Cov}(d_{ni}, \log m_{ni})$	Mean d_{ni}^{85th}
Exactly Identified	Armington	**3.99** [3.83, 4.26]	1.27	—	—
	EK	**3.99** [3.80, 4.18]	0.96	—	—
	BEJK	**3.99** [3.84, 4.15]	0.63	—	—
Over Identified	Armington	**3.97** [3.82, 4.19]	1.21	0.20	0.24
	EK	**3.99** [3.90, 4.10]	0.96	0.10	0.37
	BEJK	**4.00** [3.88, 4.10]	0.63	0.04	0.25

Note: Value is the median estimate across simulations. In each simulation there are 30 countries and 100 simulations are performed. Values within brackets report 90th and 10th percentiles.

To provide evidence that the estimation procedure recovers the underlying parameter for each model, we apply the methodology on data simulated by each model under a known θ. For exposition purposes, we let θ be the same across the models, and we set it equal to four.

Our estimation routine recovers the value of θ in all three models, both in the exactly identified and overidentified case. Given these results, we proceed to estimate the parameter using our methodology and actual trade-flow and price data.

5. Results

5.1. Description of Data

In this section, we apply our estimation strategy to real data. Our sample contains the thirty largest countries in the world (in terms of absorption). We use trade flows and production data for the year 2004 to construct trade shares. The price data used to compute moments on price gaps and aggregate price levels come from basic-heading-level data from the 2005 round of the International Comparison Programme (ICP). The dataset has been employed in a number of empirical studies. For example, Bradford (2003) and Bradford and Lawrence (2004) use the ICP price data in order to measure the degree of fragmentation, or the level of trade barriers, among OECD countries. In addition, the authors provide an excellent description of the data-collection process. Eaton and Kortum (2002) use a similar dataset for the year 1990 in their estimation.

The ICP collects price data on goods with identical characteristics across retail locations in the participating countries during the 2003-2005 period.[5] The basic-heading level represents a narrowly-defined group of goods for which expenditure data are available. The data set contains a total of 129 basic headings, and we reduce the sample to 62 categories based on their correspondence with the trade data employed. Simonovska and Waugh (2014) provide a more detailed description of the ICP data.

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 of 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. It is precisely this sampling procedure that we mimic in the simulation methodology described above.

The table below shows that price gaps in the ICP data are informative about trade barriers as they covary with typical gravity variables that determine observed trade patterns. The first row regresses the object d_{ni}, computed using the first-order statistic over price gaps, on the logged distance between country pairs and a dummy variable for contiguity. Clearly, maximum price

---

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

gaps are rising in distance and are lower for countries that share a border. The second row repeats the exercise using the object that relies of the 85th percentile of price gaps. While the qualitative patterns are identical, the magnitudes are much lower. Given the systematic relationship between moments from the price gap distribution and gravity variables, we conclude that price gaps do not simply capture measurement error, but rather they contain valuable information about true trade barriers.

Price Differences and Trade Determinants

Dependent Variable	Log Distance	Border	Mean Value
d_{ni}	0.13 [0.11, 0.15]	−0.11 [−0.18, −0.02]	2.5
d_{ni}^{85th}	0.055 [0.047, 0.063]	−0.035 [−0.060, −0.010]	1.4

Note: The first two columns report regression of dependent variables on distance, border indicator and country fixed effects. The last column reports the mean value, exponentiated.

5.2. Results with ICP Data

Estimation Results

	Exactly Identified	Overidentified, Identity	Overidentified, Optimal	
Model	Estimate of θ	Estimate of θ	Estimate of θ	J-stat P-value
Armington	**5.24** [5.02, 5.45]	**5.06** [4.84, 5.26]	**3.94** [3.73, 4.08]	< 0.01
EK	**4.17** [4.00, 4.34]	**4.13** [3.99, 4.29]	**4.05** [3.97, 4.14]	0.25
BEJK, $\rho = 2.5$	**3.32** [3.18, 3.48]	**3.28** [3.16, 3.43]	**3.15** [3.08, 2.96]	< 0.01
BEJK, est. $\hat{\rho}$	− − −	**3.21** [2.90, 3.64]	**2.74** [2.62, 2.89]	< 0.01

Note: Values within brackets report 90th and 10th confidence intervals. J-stat P-value reports probability of observing a J-statistic as large as observed. Estimates of ρ for BEJK model are 2.25 and 1.37.

The table above presents the results for the three specifications. The first column presents the results from the exactly identified case which most closely corresponds with our theoretical results in Proposition 3. Consistent with the theoretical results, the estimate of θ is systematically lower for the new trade models relative to the Armington model, and further the estimate for BEJK is lower than for EK. The difference in magnitudes is substantial. The EK estimate is about

20 percent lower than Armington, implying that the welfare cost of autarky is 20 percent higher in the EK model. In comparison, our estimate of the trade elasticity in the BEJK model is about 33 percent lower relative to EK, implying the welfare gains are estimated to be 50 percent larger in BEJK relative to Armington.

The last three columns present the results for the overidentified case. Note that with the weighting matrix equal to the identity matrix, the results are effectively identical to the results in the exactly identified case.

When the optimal weighting matrix is used, results do begin to change. Specifically, our estimate for the Armington model declines substantially relative to the previous two columns and the ranking of θ reverses. This may seem problematic relative to our results in Proposition 3, yet in fact it is highly symptomatic of the fact that the Armington model is misspecified. This intuition is formalized in the P-value for the J-statistic which tests the null hypothesis that the data generating process is true. For the Armington model, the null hypothesis is rejected at a level of less than one percent. Given the starkness of the Armington model, this should be a reassuring result.

For the EK and BEJK models, the results with the optimal weighting matrix are quantitatively similar to the other results. More interesting are the results for the J-statistic. For the EK model, we fail to reject the null hypothesis that the data generating process is true. For the BEJK model, there is evidence that it is misspecified as the null hypothesis that the BEJK model generated the data is rejected at the 10 percent level and it is right on the margin at the 5 percent level.

Interestingly, the issue with the BEJK model is that it has a difficult time fitting the covariance between the d_{ni} and distance. The problem is due to variable markups. Variable markups naturally induce "noise" between price gaps (even if the good is traded in which case price gaps reflect the distance cost), which make it harder for the BEJK model to replicate the strong covariance observed in the data between price gaps and distance. We discovered this by taking out the covariance moment and only estimating to the two price gaps. In this specification, we failed to reject the BEJK model.

6. Endogenous Variety Models

The analysis thus far has focused on three canonical models of trade where the set of varieties is fixed. Two other canonical models that feature an endogenous set of varieties produced/consumed are the frameworks of Krugman (1980)—Krugman—and Melitz (2003)—Melitz. These two models satisfy the same aggregate restrictions as the three models analyzed thus far and therefore generate the same formula for the welfare gains from trade. Below, we outline versions of these two models and we repeat the quantitative exercises from above. We find that the Melitz model yields a lower estimate of the trade elasticity than the Krugman

25

model, thus generating higher welfare gains from trade.

6.1. Krugman (1980)

Preferences are identical to the Armington model, but the measure of goods produced by any country i is endogenous,

$$W_n = \left[\sum_{i=1}^{N} \int_{J_i} x_{ni}(j)^{\frac{\rho-1}{\rho}} dj \right]^{\frac{\rho}{\rho-1}}.$$

In the above expression, J_i denotes the set of goods produced by country i.

In particular, we assume that a producer in country i has monopoly power over a blueprint to produce a single good j. In order to produce the good, a firm in i uses the following production technology

$$y_i(j) = T_i^{1/\theta} l_i(j) + f_i.$$

Thus a firm incurs a marginal cost of $w_i/T_i^{1/\theta}$ as well as a fixed cost f_i, both in labor units. There is an unbounded pool of potential producers and the mass of firms is pinned down via a zero average profit condition.

We make the assumption that fixed costs are proportional to the mass of workers in country i, L_i. Under the proportionality assumption on fixed costs, the mass of entrants is identical across countries and proportional to $1/(1+\theta)$.

Conditional on entry, all consumers buy all products from all sources. Moreover, due to the proportionality assumption on entry costs, each country contributes a measure of $1/N$ toward the unit measure of consumed products as in the Armington model. In contrast to Armington, the price $p_{ni}(j)$ for good j from i in destination n is given by the product of the marginal cost of production and delivery and the Dixit-Stiglitz constant mark-up $\bar{m}$.

6.2. Melitz (2003)

Finally, we outline a variant of the Melitz (2003) model parameterized as in Chaney (2008). In the exposition, we follow closely Arkolakis, Demidova, Klenow, and Rodríguez-Clare (2008).

The preference relation as well as the market structure are identical to the Krugman (1980) model described above. There are two novel features. First, in order to sell its product to market n, a firm from any country i incurs an additional market access cost, a_n, which is proportional to the mass of workers in the destination, L_n. Second, marginal costs of production are not identical across producers within a country. In particular, upon paying the entry cost f_i, a

26

firm from i draws a productivity realization for good j, $z_i(j)$, from the country-specific Pareto distribution with pdf

$$\mu_i(z) = T_i z^{-\theta}. \tag{26}$$

As in the Frechet distribution, the parameter θ in the Pareto distribution governs the variability of productivity among firms.

The two additional assumptions imply that there exist cost thresholds that limit the participation of certain firms to certain markets. In particular, if a firm obtains a cost draw above the threshold cost that characterizes the "most accessible" destination, it exits immediately without operating. Thus, because of free entry, in equilibrium expected profits of a firm must be equal to the fixed entry cost. In addition, more productive firms serve more markets and the "toughest" markets attract the most efficient producers from all sources.

As in the Krugman (1980) model, firms charge a constant mark-up over their marginal cost of production. However, unlike the Krugman (1980) model, the prices of goods offered by a given country differ according to the efficiency of the individual producer. More importantly, even though entry costs are proportional to the mass of workers, the measure of firms that serve each market is no longer constant. Thus, selection among exporters alters the composition of the bundle of consumed goods.

6.3. Trade Flows, Aggregate Prices, and Welfare

In this section, we show that the two models with an endogenous set of varieties produce identical aggregate outcomes to the models with an exogenous set of varieties. In particular, under the parametric assumption made above, all five models yield the same expressions for trade flows, price indices (up to a constant scalar), and welfare gains from trade. Proposition 4 summarizes the result. The proof can be found in the Appendix.

Proposition 4 *Given the functional forms for productivity $F_i(\cdot)$, $G_i(\cdot)$, and $\mu_i(\cdot)$ for all $i = 1, ..., N$, and assuming that (i) fixed entry costs are proportional to the mass of workers in the country of origin, $f_i \propto L_i$ for all $i = 1, ..., N$; (ii) market access costs are proportional to the mass of workers in the destination market, $a_n \propto L_n$ for all $n = 1, ..., N$;*

a. The share of expenditures that n spends on goods from i, X_{ni}/X_n, predicted by all five models is

$$\frac{X_{ni}}{X_n} = \frac{T_i(\tau_{ni} w_i)^{-\theta}}{\sum_{k=1}^{N} T_k(\tau_{nk} w_k)^{-\theta}}. \tag{27}$$

27

b. The CES exact price index for destination n, P_n, predicted by all five models is

$$P_n \propto \Phi_n^{-\frac{1}{\theta}}, \quad where \quad \Phi_n = \sum_{k=1}^{N} T_k (\tau_{nk} w_k)^{-\theta}. \tag{28}$$

c. The percentage compensation that a representative consumer in n requires to move between two trading equilibria predicted by all five models is

$$\frac{P_n'}{P_n} - 1 = 1 - \left(\frac{X_{nn}'/X_n'}{X_{nn}/X_n} \right)^{\frac{1}{\theta}}. \tag{29}$$

6.4. Price Gap Distribution

The Krugman and the Melitz model yield identical predictions about trade shares and relative prices of a given good across countries. Recall that in both models, if a variety is produced in one country and exported to another, then the relative price would simply reflect relative trade barriers as the cost of production is identical since it is linked to the exporting firm. The only difference across the two models would therefore arise due to the selection of firms into different markets. In the Krugman model, all successful domestic producers necessarily export, whereas in the Melitz model, only a fraction of domestic firms engage in exporting. This is the extensive margin that the Melitz model adds on top of the intensive margin featured in Krugman.

What remains to analyze is the circumstance under which productivity-based selection will affect the price gap distribution across countries. Consider a symmetric two-country version of each model. In the Krugman model, the price gap distribution is identical to the one derived for the Armington model. For the Melitz model, on the other hand, there are instances in which some products are sold domestically but not abroad. In this case, however, a relative price ratio cannot be defined since the product is not available in one of the destinations. Therefore, in order to be able to define relative prices in the Melitz model, we need to condition only on the firms that sell both at home and abroad. In this case, however, the price gap distribution collapses to the one in the Armington model and the two models yield identical trade elasticity estimates and welfare gains.

The estimates of the trade elasticity, however, differ across the models when the assumptions about two symmetric countries are relaxed. In order to understand the argument, it is useful to first describe the simulation procedure for each model, to which we turn next.

6.5. Simulation of Prices for Endogenous Variety Models

Since the endogenous variety models yield an identical gravity equation of trade as the exogenous variety models, the first stage of the estimation is identical. What remains is to describe the micro price-level simulation.

Step 2d.—The simulation of the Krugman model is trivial because it does not feature micro-level heterogeneity. In this model, $\tilde{S}_i$ fully characterizes the common marginal cost of production across all producers in country i. Moreover, all produced goods appear on the common list, and each country contributes a fraction $1/N$ of them.

Step 2e.— The simulation of the Melitz (2003) model is more intricate. In the Eaton and Kortum (2002) and Bernard et al. (2003) models—which feature heterogeneity—a good is indexed by $j \in [0,1]$ or an integer when discretized on the computer. In the Melitz (2003) model, different countries are consuming and producing different goods. To carry out the simulation, we discretize the continuum to a set of potentially produced goods, which equals to 300,000*N. We proceed to determine the subset of goods produced domestically for each country. Then, we simulate inverse marginal costs for each good. Inverse marginal costs are determined following Eaton, Kortum, and Kramarz (2011) who show how normalized inverse marginal costs can be sampled from a parameter free uniform distribution.

Then, we determine the set of exported varieties for all country-pairs and we compute their prices. We proceed to define the common set of goods whose prices will be sampled. The set includes the goods produced by firms with low enough cost draws so that they serve all N destinations. Let M_i represent the measure of goods from i that appear on the common list and let $M = \sum_m M_m$ be the common list. Then, the share of country i's goods on the common list is

$$
\frac{M_i}{M} = \frac{\min_l \tilde{S}_i \tau_{li}^{-\theta} \left[\sum_k \tilde{S}_k \tau_{lk}^{-\theta} \right]^{-1}}{\sum_m \min_l \tilde{S}_m \tau_{lm}^{-\theta} \left[\sum_k \tilde{S}_k \tau_{lk}^{-\theta} \right]^{-1}}.
\tag{30}
$$

Notice that this share varies across countries. This variation is precisely the reason behind the different elasticity estimates that we obtain from the two models. In the Krugman model, the maximal bilateral price difference is likely to equal the unobserved trade friction (exactly like in the Armington model). In the Melitz model, only goods sold by firms that enter all markets are able to make the sample of prices that data collectors would consider. Moreover, the firms that enter all markets are likely to be from source countries with low trade costs to export. This implies that, in the Melitz model, the average maximal price difference across country pairs is likely to be below the average trade friction across country pairs. Thus, one needs a lower trade elasticity in Melitz relative to Krugman to rationalize the same amount of trade observed in the data. The differences in trade elasticities then translate into differences in the measured welfare

gains from trade in the two models.

6.6. Results with ICP Data

Estimation Results, Endogenous Variety Models

Model	Exactly Identified	Over-Identified, Identity	Over Identified, Optimal	
	Estimate of θ	Estimate of θ	Estimate of θ	J-stat P-value
Krugman	**5.24**	**5.06**	**3.94**	< 0.01
	[5.02, 5.45]	[4.84, 5.26]	[3.73, 4.08]	
Melitz	**3.69**	**3.69**	**3.65**	0.35
	[3.32, 4.21]	[3.37, 4.09]	[3.36, 3.86]	

The table above reports the estimates of θ for the two endogenous variety models obtained using the ICP data. The first column reports the results from the exactly identified estimation. The Melitz model yields an estimate of θ that is roughly 30 percent lower than the estimate for the Krugman model. Hence, adding an extensive margin of trade increases the welfare gains from trade by 30 percent.

The last three columns report the results from the overidentified estimation. Once again a similar picture emerges as in the estimation of the exogenous variety models. The overidentified estimation that uses the optimal weighting matrix yields an estimate of θ for the model that features an extensive margin—Melitz in this case—that is roughly the same as in the overidentified case with an identity matrix and the exactly identified case. However, when using the optimal weighting matrix, the estimate of θ falls considerably for the model that features only an intensive margin of trade—in this case Krugman. Once again, the J statistic allows us to reject the hypothesis that the model with an intensive margin only has generated the data. In contrast, we cannot reject the null that the Melitz model has generated the data.

6.7. Melitz and Redding (2014)

Our estimation yields a lower trade elasticity estimate for the Melitz model than the Krugman model, which translates into higher welfare gains from trade in the former. Recently, Melitz and Redding (2014) also argue that the Melitz model generates higher welfare gains from trade than the Krugman model. Their argument, however, is very different from ours and it goes as follows. The Krugman model is a special case of the Melitz model where the distribution of firm productivities becomes degenerate. Naturally, θ is the central parameter that governs this distribution. The authors then demonstrate that, if all other parameters of the two models

are kept identical, the Melitz model generates higher trade shares and therefore higher welfare gains from trade.

Instead, our exercise estimates the parameters of the two models such that they generate *identical* trade shares—the ones observed in the data. Then, the reason why the welfare gains differ is because different θ's are necessary in order to match the price moments in the data.

7. The Elasticity of Trade and Tariffs

How do our estimates of trade elasticities compare to estimates that one would obtain using a tariff based approach? This is a particularly relevant question in light of the results of Arkolakis et al. (2011). As they suggest, a tariff based approach with a moment condition that is not model specific will result in an estimate of θ that is model independent , which implies that the welfare gains from trade are the same across models.

To explore the implications of our method/results in light of this question we modified our estimation approach by following Caliendo and Parro's (2011) which uses triple differenced trade flows and triple differenced tariffs. For notational convince, denote the log of triple differenced trade flows by $\tilde{x}_{nih}$ for country triple n, i, h and the log of triple differenced tariffs by $\tilde{t}_{nih}$. Finally denote N_t as the number of country triple observations.

Caliendo and Parro's (2011) estimation is based on the following moment condition:

$$\mathbb{E}\left\{\tilde{t}_{nih}\left(\tilde{x}_{nih} - \theta\tilde{t}_{nih}\right)\right\} = 0, \tag{31}$$

that is the covariance of triple differenced tariffs is orthogonal to the residual. In other words, their estimation is just OLS. In our context, we will stack moment conditions in the following way. Define

$$h_{nih}(\theta, \tilde{x}_{nih}, \tilde{t}_{nih}) = \tilde{t}_{nih}\left(\tilde{x}_{nih} - \theta\tilde{t}_{nih}\right). \tag{32}$$

Then averaging across triples yields

$$g_T(\theta) = \left(\frac{1}{N_T}\right) \sum_n \sum_i \sum_h h_{nih}(\theta, \tilde{x}_{nih}, \tilde{t}_{nih}), \tag{33}$$

where N_T is the number of triples. Caliendo and Parro's (2011) estimate of θ is the value that minimizes the quadratic form of $g_T(\theta)$. Or mathematically:

$$\hat{\theta}_T = \arg\min_\theta g_T(\theta)' g_T(\theta). \tag{34}$$

where the notation $\hat{\theta}_T$ denotes that this is the estimate from the tariff based approach.

To combine this with our estimation method, we will simply stack the g functions that contain our preferred moments and the tariff-based moments. So define $\tilde{g}(\theta, \mathcal{M})$ as

$$\tilde{g}(\theta, \mathcal{M}) = \left\{ \begin{array}{c} g_T(\theta) \\ g(\theta, \mathcal{M}) \end{array} \right\} \tag{35}$$

where $g(\theta, \mathcal{M})$ is defined in equation (15) from our exactly identified estimation approach. Then the estimate of θ which merges our approach with that of Caliendo and Parro's (2011) is the value that minimizes the quadratic form of $\tilde{g}(\theta)$. Or mathematically:

$$\hat{\theta}(\mathcal{M}) = \arg\min_\theta \tilde{g}(\theta, \mathcal{M})'\tilde{g}(\theta, \mathcal{M}). \tag{36}$$

where we will focus attention on an estimate obtained using a weighting matrix that is equal to the identity matrix.

To implement this approach we used the trade flow and tariff data from Parro (2013) for both capital and non-capital goods. We then took a trade weighted average of tariffs across these two sectors to arrive at an aggregate tariff measure.

The first column of Table 3 presents the results using only the tariff based estimation approach. Using this data we find an estimate of θ of around 4.6. This result is very much consistent with the results that Parro (2013) reports. The second column presents the results from our exactly identified estimation, under a value for ρ that equals 1.37 in the BEJK model.

Table 3: Estimates of θ with Tariffs

	Estimation Method		
Model	Tariffs	Price Gaps	Tariffs + Price Gaps
Armington	4.63	5.24	5.24
EK	4.63	4.17	4.17
BEJK, ρ=1.37	4.63	2.74	2.74

The third column of Table 3 reports our estimate of θ when both the price gap moment conditions and the moment conditions from Caliendo and Parro's (2011) method are used. The key result is that the third column is identical to the second column. Incorporating tariff based moments has virtually zero effect on our estimates.

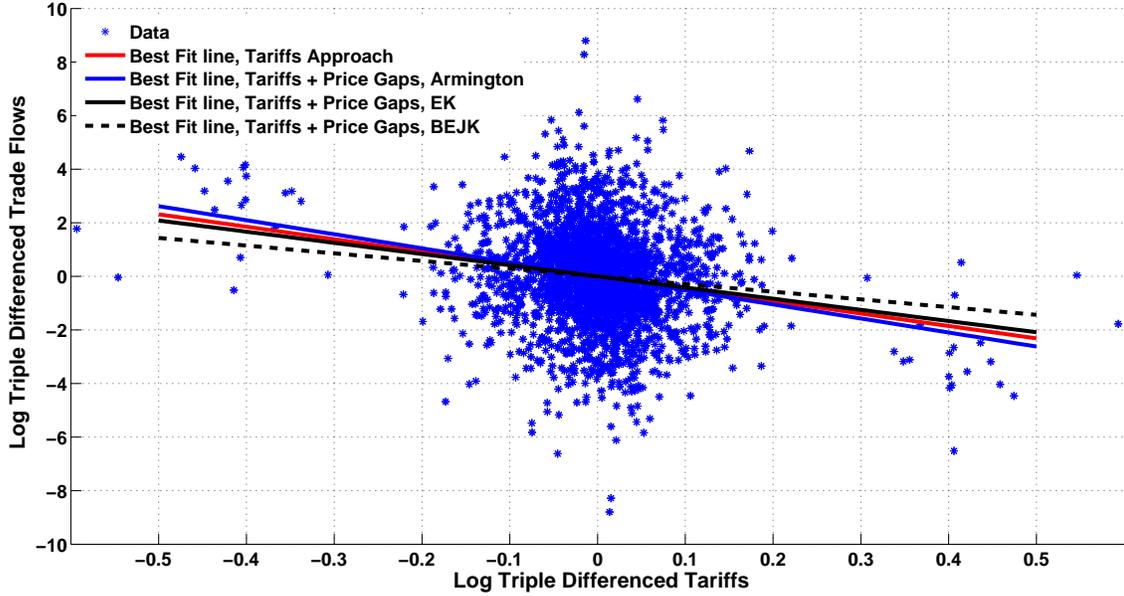The reason for this is that the tariff based method provides very little information about the

Figure 2: Trade Flows and Tariffs

exact nature of the trade elasticity. To see this, Figure 2 simply plots the data with triple differenced tariffs on the x-axis and triple differenced trade flows on the y-axis. Moreover, plotted on top of the data are the "best fit" lines implied by the different estimation methods.

The key thing to notice is that the data are exceptionally noisy and that the pure tariff based approach does not describe the data that well. This implies that large deviations (say from 4.6 to 2.8) from the least square best fit provide negligible losses in mean square error. This means that tariffs provide little valuable information with regards to θ. Thus when the tariff based approach is combined with our estimation method, the tariff data does not assist in identifying θ. Hence the reason the third column of Table 3 is exactly the same as the second column.

8. Conclusion

We argue that the welfare gains from trade in new trade models with various micro-level margins are larger relative to models without these margins. First, we make a theoretical argument: Assuming a fixed trade elasticity, different models have identical predictions for bilateral trade flows but different predictions for cross-country price variation. Thus, for given data on trade flows and prices, different models have different trade elasticities and thus different welfare gains from trade. Second, we estimate the trade elasticity and the welfare cost of autarky under different assumptions about the underlying model. Relative to an Armington trade model, an active extensive margin increases the welfare cost of autarky by 20 percent; variable markups further increase the cost by 30 percent. In addition, relative to a Krugman model, an active

extensive margin as in Melitz increases the welfare cost of autarky by 30 percent. Third, we incorporate into our estimation moment conditions that use trade flow and tariff data, which imply a common trade elasticity; this modification has virtually no effect on our results.

References

ADDA, J. AND R. W. COOPER (2003): *Dynamic economics: quantitative methods and applications*, MIT press.

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

ANDERSON, J. E. (1979): "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69, 106–16.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (2011): "New Trade Models, Same Old Gains?" *American Economic Review (forthcoming)*.

ARKOLAKIS, C., S. DEMIDOVA, P. KLENOW, AND A. RODRÍGUEZ-CLARE (2008): "Endogenous Variety and the Gains from Trade," *The American Economic Review*, 98 (2), 444–450.

ARNOLD, B. C., N. BALAKRISHNAN, AND H. N. NAGARAJA (1992): *A first course in order statistics*, vol. 54, Siam.

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

BRADFORD, S. (2003): "Paying the Price: Final Goods Protection in OECD Countries," *The Review of Economics and Statistics*, 85, 24–37.

BRADFORD, S. AND R. Z. LAWRENCE (2004): *Has Globalization Gone Far Enough: The Costs of Fragmented Markets*, no. 349 in Peterson Institute Press: All Books, Peterson Institute for International Economics.

CALIENDO, L. AND F. PARRO (2011): "Estimates of the Trade and Welfare Effects of NAFTA," *mimeo*.

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

DAVISON, A. C. AND D. HINKLEY (1997): *Bootstrap methods and their application*, vol. 1, Cambridge university press.

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

EATON, J., S. KORTUM, AND F. KRAMARZ (2011): "An Anatomy of International Trade: Evidence from French Firms," *Econometrica (forthcoming)*.

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

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

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

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

MELITZ, M. J. AND S. J. REDDING (2014): "New Trade Models, New Welfare Implications," Working Paper 18919, National Bureau of Economic Research.

PARRO, F. (2013): "Capital-skill complementarity and the skill premium in a quantitative model of trade," *American Economic Journal: Macroeconomics*, 5, 72–117.

SIMONOVSKA, I. AND M. WAUGH (2011): "Different Trade Models, Different Elasticities?" *New York University and Princeton University, mimeo*.

SIMONOVSKA, I. AND M. E. WAUGH (2014): "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 92, 34 – 50.

WAUGH, M. E. (2010): "International Trade and Income Differences," *American Economic Review*, 100, 2093–2124.

A. Proofs

The proof to Proposition 4 follows.

Proof

We only need to show the results for the endogenous variety models.

a. and b. We derive the trade shares and price indices for the versions of the Krugman (1980) and Melitz (2003) models that we consider in the present paper.

For the Krugman (1980) model, maximizing W_i subject to a standard budget constraint yields the following CES demand function

$$x_{ni}(j) = \left(\frac{p_{ni}(j)}{P_n}\right)^{-\rho} \frac{w_n L_n}{P_n}, \tag{37}$$

$$P_n = \left[\sum_{i=1}^{N} \int_{J_i} p_{ni}(j)^{1-\rho} dj\right]^{\frac{1}{1-\rho}} \tag{38}$$

Moreover, maximizing firm profits and accounting for the demand in (37) yields

$$p_{ni}(j) = \frac{\bar{m}\tau_{ni} w_i}{T_i^{1/\theta}} \tag{39}$$

Free entry implies that the expected profit of a firm is zero, after paying the fixed entry cost. Using the firm's optimal pricing rule and quantity demanded, which must equal to the quantity produced, yields

$$\sum_{n=1}^{N} \frac{(p_{ni}(j))^{1-\rho}}{P_n^{1-\rho}} \frac{w_n L_n}{\rho} - w_i f_i = 0 \tag{40}$$

Let M_i denote the measure of entrants in country i, which corresponds to the measure of set J_i. In order to compute the measure of entrants, combine the free entry condition from (40) with the labor market clearing condition for country i,

$$M_i \sum_{n=1}^{N} \frac{(p_{ni}(j))^{1-\rho}}{P_n^{1-\rho}} w_n L_n + f_i = L_i, \tag{41}$$

to obtain $M_i = L_i/(\rho f_i)$. Using the definition $\theta = \rho - 1$ and assuming that $L_i \propto f_i$ yields $M_i \propto 1/(1+\theta)$. Relying on this result, use the equilibrium measure of entrants M_i and the optimal pricing rule (39) into (38) to obtain that $P_n \propto \Phi_n^{-1/\theta}$.

Finally, the share of expenditure that n spends on goods from i is given by

$$\frac{X_{ni}}{\sum_{k=1}^{N} X_{nk}} = \frac{M_i \frac{(p_{ni}(j))^{1-\rho}}{P_n^{1-\rho}} \frac{w_n L_n}{\rho}}{\sum_{k=1}^{N} M_k \frac{(p_{nk}(j))^{1-\rho}}{P_n^{1-\rho}} \frac{w_n L_n}{\rho}} \tag{42}$$

Using the definition $\theta = \rho - 1$, the equilibrium measure of entrants M_i under the proportionality assumption, the optimal pricing rule (39) and the exact price index (38) in (42) yields the expression in the text.

For the Melitz (2003) model, following the same steps as for the Krugman (1980) model obtains the demand function in (37) and the price index in (38). Let the marginal cost of a firm with

efficiency draw $z_i(j)$ of producing good j in i and delivering it to n be

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

We will express all equilibrium objects in terms of cost draws in the remainder of the Appendix. As in the Krugman (1980) model, the optimal pricing rule is given by

$$p_{ni}(j) = \bar{m} c_{ni}(j) \tag{43}$$

Using the firm's optimal pricing rule and quantity demanded, which must equal to the quantity produced, yields the firm's profit function

$$\pi_{ni}(c_{ni}(j)) = w_n L_n \left(\frac{\rho}{\rho-1}\right)^{-\rho} \frac{1}{\rho-1} \left(\frac{c_{ni}(j)}{P_n}\right)^{1-\rho} - w_n a_n$$

The cost cutoff for a firm from i to sell good j to n, $\bar{c}_{ni}(j)$, satisfies $\pi_{ni}(\bar{c}_{ni}(j)) = 0$ and is given by

$$\bar{c}_{ni}(j) = \left[\frac{a_n}{L_n} \frac{P_n^{1-\rho}}{\left(\frac{\rho}{\rho-1}\right)^{-\rho} \frac{1}{\rho-1}}\right]^{\frac{1}{1-\rho}}$$

Since $\bar{c}_{ni}(j) = \bar{c}_{nn}(j) \; \forall i$, denote cost cutoffs by the destination to which they apply, $\bar{c}_n(j)$.

To determine the measure of entrants, repeat the procedure applied to the Krugman (1980) model to obtain $M_i = L_i/f_i(\rho-1)/(\rho\theta)$. Only those firms with cost draws below $\bar{c}_i$ serve the domestic market and the remainder immediately exit. Using the Pareto distribution, the measure of domestic firms that sell j in country i is $M_{ii}(j) = M_i T_i w_i^{-\theta}(\bar{c}_i(j))^\theta$. Similarly, the measure of firms from i that sell j to n is $M_{ni}(j) = M_i T_i (w_i \tau_{ni})^{-\theta}(\bar{c}_n(j))^\theta$.

Furthermore, assuming that $L_i \propto f_i$ yields $M_i \propto (\rho-1)/(\rho\theta)$. Relying on this result, use the equilibrium measure of firms that serve destination n from any source i, M_{ni}, the optimal pricing rule (43), and the Pareto distribution into (38) to obtain that $P_n \propto \Phi_n^{-1/\theta}$.

Finally, to derive the trade share, use the optimal pricing rule (43), quantity demanded, which must equal to the quantity produced, the equilibrium measure of firms that serve each destination, and the Pareto distribution.

c. We refer the reader to Arkolakis et al. (2011) for derivations concerning welfare in all the models.