Exporter Heterogeneity and Price Discrimination: A Quantitative View*

Jae Wook Jung† UC Davis

Ina Simonovska‡ UC Davis, NBER

Ariel Weinberger§ UC Davis

July 19, 2015

Abstract

We quantify a class of commonly-employed general equilibrium models of international trade and pricing-to-market that feature firm-level heterogeneity and consumers with non-homothetic preferences. We demonstrate theoretically that the models lack the flexibility to match salient features of US firm-level data. Consequently, we outline a theoretical framework that can reconcile the documented price dispersion across firms and markets, while maintaining consistency with cross-sectional observations on firm productivity and sales. We calibrate the model's parameters to match bilateral trade flows across 71 countries as well as the productivity and sales advantages of US exporters over non-exporters. We find that the calibrated model accounts for the majority of the dispersion in prices of tradables across countries of different income levels, while maintaining a tight quantitative fit to firm-level data. Given its additional flexibility, the model quantitatively outperforms the existing alternatives and yields welfare gains for the US that are 14-54% higher, but at the cost of loss of tractability.

JEL Classification: F12, F14, F17, F61

Keywords: variable mark-ups, non-homothetic preferences, heterogeneous firms, international trade, gravity

*We are indebted to Andres Rodriguez-Clare for extremely insightful comments that were critical to the analysis in this paper.

†jwjung@ucdavis.edu

‡inasimonovska@ucdavis.edu.

§aweinberger@ucdavis.edu

1 Introduction

The empirical international trade literature has documented two sets of stylized facts that shed light on the role that heterogeneous firms play in shaping prices and flows of goods across borders. On the one hand, Bernard et al. (2003) demonstrate that: (i) exporters enjoy a large productivity and size advantage over non-exporters; (ii) exporters constitute the minority among producing plants; (iii) exporter sales abroad account for a small portion of their total sales. On the other hand, De Loecker and Warzynski (2012) show that more productive firms, and especially exporters, extract higher mark-ups than less productive ones, while Simonovska (2015) finds that a typical exporter enjoys higher mark-ups in richer over poorer destinations.

We outline a theoretical framework that can reconcile these facts, both qualitatively and quantitatively. In particular, we analyze a general equilibrium model of international trade that features monopolistically competitive heterogeneous firms and identical consumers whose preferences are represented by a generalized CES utility function—Dixit-Stiglitz utility with displaced origin. According to the model, the marginal utility that each consumer derives from positive consumption is bounded, which implies the existence of a choke price above which demand is zero. Hence, only firms with sufficiently high productivity draws that offer their products at prices below the choke price survive in a given market, and the marginal firm realizes zero sales. Iceberg trade barriers raise the cost to serve foreign markets, which implies that a subset of productive firms constitute the set of exporters. These firms enjoy a sales and value-added advantage over non-exporters, while constituting a minority when trade barriers are high. In turn, for the majority of exporters, high trade barriers erode foreign relative to domestic sales. Finally, consumers that reside in countries characterized by higher income levels are less responsive to price changes than those that live in poorer ones, so firms optimally price identical products higher in more affluent markets.

Having reconciled theory and fact, we examine the ability of the model to quantitatively account for the observations in the data. Under the assumption that firm productivities are Pareto distributed, the framework falls within the class of models investigated in Arkolakis et al. (2015) (ACDR); therefore, it is easily quantifiable.[1] We rely on the model's predicted gravity equation of trade as well as bilateral trade-flow data for 71 countries to uncover all, but two of, the parameters necessary to simulate moments at the firm level. The two key remaining parameters constitute the Pareto shape parameter and the utility curvature parameter. We identify these parameters by matching the measured-productivity and sales advantage of U.S. exporters over non-exporters documented in Bernard et al. (2003) (BEJK). The focus on these two moments is intentional as they are directly informative about the two parameters in ques-

[1] We are indebted to Andres Rodriguez-Clare for pointing this out to us.

tion.[2] On the one hand, lower values of the Pareto shape parameter increase the variability in firm productivity, thus raising the exporters' value-added advantage over non-exporters. On the other hand, higher values of the utility curvature parameter raise product substitutability, thus reducing the market power of each firm. More productive firms charge lower prices than less productive ones and earn a higher sales advantage with more substitution across varieties, while enjoying higher value added and mark-ups.

We conduct a variety of out-of-sample tests to evaluate the model's fit to data. Consistent with firm-level facts, only the minority of firms are exporters (41%), the majority of these exporters sell mostly domestically, and the standard deviation of log domestic sales is broadly in line with the data. More importantly, we test the model's fit to the observed price dispersion across firms and markets. The model yields markups consistent with the empirical literature—the average markup is 31%. In addition, the model yields a positive relationship bwtween prices of tradable goods and destination per-capita income. Among highly productive exporters that serve at least half the markets in the dataset, a linear regression of log relative price on log per-capita income yields a slope of 0.12, which compares favorably to the statistic obtained from standard cross-country price datasets such as the Economist Intelligence Unit (EIU) and the International Comparison Program (ICP).

Furthermore, because the model falls within the ACDR framework, we can rely on welfare results derived by the authors to compute the welfare gains from trade. We calculate the welfare gains of moving from autarky to the observed trade equilibrium in year 2004. The model predicts that the gains for the US economy would be 4%, while the average gains across 65 countries would amount to 16%.

The discussion above suggests that the generalized CES model represents a plausible quantitative framework that can be used to study the welfare gains from trade in the presence of pro-competitive effects. While the framework's aggregate outcomes are easily quantifiable, the model does not yield analytical representations of firm-level variables, which incurs computational costs to the researcher. In addition, the model shares key qualitative predictions with three existing frameworks: Melitz and Ottaviano (2008) (MO), Behrens et al. (2014) (BMMS), and Simonovska (2015) (SIM).[3] The distinctive features of these models are that consumer preferences are non-homothetic and monopolistically competitive firms are heterogeneous in their productivities, which are Pareto distributed.[4] Therefore, when framed in a multi-country

[2]The focus on the actual statistics reported in BEJK is simply due to data availability constraints.

[3]Feenstra and Weinstein (2010) also examine firm-level price variation in a monopolistically competitive world that features heterogeneous firms; however, the authors rely on translog preferences, which fall within a homothetic class.

[4]Dhingra and Morrow (2012) and Weinberger (2014) examine the welfare consequences of variable markups and incomplete pass-through in similar frameworks.

general equilibrium setting, these models are qualitatively in line with all of the stylized facts described above. If, in addition, the MO model is parameterized appropriately, the three models satisfy the assumption of separable preferences so that the choke price in all the models is identical to the one generated by the generalized CES model.[5]

Given a unique set of gravity variables and choke price that can be computed for all the models, we show that the three existing models cannot jointly account for the observed measured-productivity and sales advantage of exporters over non-exporters, unlike the flexible generalized CES model. We demonstrate theoretically that, in general equilibrium with free entry and under the assumption that firm productivity is unbounded from above and Pareto distributed, the three existing non-homothetic models rely on a single parameter, the Pareto shape parameter, to match the two moments, which renders the task theoretically impossible. Furthermore, for a given Pareto shape parameter value, all models—existing as well as the generalized CES one—yield identical productivity cutoffs and macro aggregates, but differing firm mark-up and sales distributions. Therefore, the models yield differing predictions about price dispersion and ultimately consumer welfare. Consequently, if firm-level data rejects these moments, but the models are used to quantify the welfare gains from trade, the latter could be mismeasured.

To quantify the potential costs of welfare mismeasurement, we replicate the quantitative analysis from above within the context of these models and we compute the predicted welfare gains. First, we find that all three models can match the measured productivity advantage of US exporters over non-exporters (under model-specific values for the Pareto shape parameter), but the result is that the models predict too low of a sales advantage of exporters.[6][7] Quantitatively, the parsimonious frameworks underpredict the welfare gains from trade for the US by 14-54% relative to the flexible one. Second, we show that all three models can match the sales advantage of exporters under relatively high (model-specific) Pareto shape parameters. In this case: (i) the measured productivity distribution in the models is misaligned with respect to the data; and (ii) the welfare gains, which are falling in the value of the Pareto shape parameters, are further underestimated relative to the first case.

The above analysis does not imply that the existing models do not constitute plausible benchmarks for quantitative analysis. Admittedly, we evaluate the performance of the models

[5]In particular, in the MO model, the parameter denoted by η must be set to zero. For completeness, we also examine MO's general non-separable case qualitatively and quantitively.

[6]In contrast, the unrestricted MO model can reconcile the sales and value-added advantage of exporters over non-exporters, but this flexibility comes at a high cost: the calibrated model predicts that exporters are in the majority, which is in contrast to the data, and that the firm sales variance is too low, while the average mark-up and the elasticity of price with respect to income are too high.

[7]Feenstra (2014) examines the role that the upper bound of the Pareto distribution for productivity plays in governing welfare in a large class of models. In results available upon request, we show that incorporating a Pareto productivity distribution that is bounded above into all the models examined in this paper further restricts the sales advantage of exporters over non-exporters.

along a limited number of dimensions. A natural question arises: why focus the analysis on these moments in the data? For example, another strategy could be to identify the models' parameters by matching moments from the price distribution in each model as suggested by Simonovska and Waugh (2014b). We believe that, by focusing on moments that do not relate to prices, we put the existing (and the generalized CES) models to a more difficult test because these models were developed precisely so as to reconcile pricing observations in the data. In addition, the cross-sectional moments from firm data that we examine constitute the core moments traditionally examined by the quantitative international trade literature of micro-level heterogeneity (see BEJK, Arkolakis (2010), Behrens et al. (2014), and ACDR among others). We leave it for future work to examine the performance of these models along other dimensions.

We contribute to a large and important existing literature that emphasizes the role that heterogeneous productivity firms play in shaping international prices and trade flows. Most notably, to reconcile the documented facts about US exporters, BEJK develop a model in which suppliers with heterogeneous productivities compete within and across countries à la Bertrand and sell to consumers with homothetic (CES) preferences. A key prediction of the model is that, on average, more efficient suppliers have greater cost advantage over their rivals, set higher mark-ups and enjoy higher measured productivity. Furthermore, on average, the more productive firms compete with more productive rivals, charge lower prices, and enjoy higher sales. Their model is not only able to quantitatively account for the first set of documented facts, but it is also highly tractable and lends itself useful for quantitative analysis. The models examined in the present paper are distinct from theirs in that consumer preferences are assumed to be non-homothetic and the market structure is monopolistically competitive. These two features allow us to derive testable predictions about individual firms' price discrimination practices within and across countries and to quantitatively account for the positive relationship between prices of tradables and per-capita income documented by Alessandria and Kaboski (2011) and by Simonovska (2015).[8]

Finally, our paper relates to a large literature that quantitatively examines the role of heterogeneous firms in the global economy. Most notably, the workhorse model by Melitz (2003), parameterized as in Chaney (2008), features firms with heterogeneous productivity levels that incur fixed domestic and (in addition) export market access costs, thus accounting for the size and productivity advantage of exporters over non-exporters. Dispensing with Melitz's (2003) fixed market access cost formulation and incorporating into the model Arkolakis's (2010)

[8] Notably, deBlas and Russ (2015) and Edmond et al. (2014) generalize the BEJK model and derive richer predictions about aggregate pass-through and welfare, albeit under the same market structure assumption as BEJK.

advertising technology that allows firms to reach only a fraction of consumers in each market enables the framework to rationalize the co-existence of few exporters with exporters' tiny sales per export market. Adding to the last model firm- and market-specific shocks enables Eaton et al. (2011) to reconcile the entry and sales patterns across different types of firms and markets that were documented for French exporters in Eaton et al. (2004) and Eaton et al. (2011).

The unifying feature of the models described above is the assumption that consumer preferences are of the homothetic (CES) form. The formulation makes the models tractable, but it yields the theoretical prediction that firms charge identical prices across different destinations, net of trade costs. However, Waugh (2010) relies on the gravity equation of trade predicted by a large class of models, including the ones described above, and shows that, if trade costs are estimated from bilateral trade data, they are, at best, uncorrelated with destination income. This implies that the models yield no systematic link between a country's level of development and prices of tradables, which is at odds with the data.[9] Nonetheless, the theoretical tools developed by these pioneering studies can be incorporated into the generalized CES model in order to explore the important firm- and market-specific dimensions that we abstract from in the present paper.

We organize the remainder of the paper as follows. We provide the details of the framework that will apply to all non-homothetic heterogeneous-firm models and outline the flexible model in Section 2. We describe the solution algorithm for the model and generate moments for calibration in Section 3. We outline the existing non-homothetic models and we examine their predicted distributions of sales and mark-ups in Section 4. We quantify the models in Section 5. We conclude in Section 6.

2 Generalized CES Model

In this section, we outline the generalized CES model. We begin by describing the market structure, which applies to all existing models, and we proceed to derive predictions from the generalized model.

2.1 Framework

The environment is static. Goods are differentiated by the producers' identity as well as the country of origin. I countries are engaged in trade of final goods, where I is finite. Let i represent an exporter and j an importer, that is, i is the source country, while j is the

[9] The author demonstrates that an alternative trade-cost specification proposed by Eaton and Kortum (2002) results in estimates of trade barriers that are systematically negatively related to destination income, which results in a negative relationship between per-capita income and prices of tradables.

destination country. In every country, i, there exists a pool of potential entrants who pay a one time cost, $f_e > 0$, in order to get a single draw from a distribution, $G_i(\phi)$, with support $[b_i, \infty)$. We assume that $G_i(\phi) = 1 - \left(\frac{b_i}{\phi}\right)^\theta$, so that $dG_i(\phi) = \frac{\theta (b_i)^\theta}{\phi^{\theta+1}}$. Therefore, the (unbounded above) productivity distribution is Pareto with shape parameter θ and lower bound b_i, which governs average productivity in i. A measure J_i of firms that are able to cover their marginal cost of production enter the world market and a subset of entrants, N_{ij}, produce and sell to market j. Thus, a subset of entrants immediately exit and, in equilibrium, the expected profit of an entrant is zero. Hence, aggregate profit rebates to each consumer are also zero. Assuming that each consumer has a unit labor endowment which, when supplied (inelastically) to the local labor market earns a wage rate of w_i—per-capita income in country i necessarily equals w_i.

Labor is the only factor of production. The production function of a monopolistically-competitive firm with productivity draw ϕ is $x(\phi) = \phi l$, where l is the amount of labor used toward the production of final output. Moreover, each firm from country i wishing to sell to destination j faces an iceberg transportation cost incurred in terms of labor units, $\tau_{ij} \geq 1$, with $\tau_{ii} = 1 (\forall i)$. For a firm with productivity draw ϕ, we define the marginal cost to produce in country i and ship to country j by $c_{ij} = \frac{w_i \tau_{ij}}{\phi}$. Then, the cdf of firms from i who can deliver their good to j at a unit cost below c is:

$$\mu_{ij}(c) = \left(\frac{c_{ij}}{\hat{c}_{ij}}\right)^\theta, \text{ where } \hat{c}_{ij} = \frac{w_i \tau_{ij}}{b_i}. \tag{1}$$

We show below that there exists a good produced by a firm from country i with cost draw $\bar{c}_{ij}$ whose demand is zero in market j and the firm producing it earns zero profits from its sale. Then the subset N_{ij} of entrants from i whose cost is below the threshold $\bar{c}_{ij}$ is:

$$N_{ij} = J_i(\mu_{ij}(\bar{c}_{ij})), \tag{2}$$

The total measure of goods consumed in j is the sum across sources that sell to it: $N_j = \sum_{i=1}^I N_{ij}$.

We maintain the above assumptions throughout the remainder of the paper. In Section 4 we explore the role that utility parameterizations as in the SIM, MO (with $\eta = 0$), and BMMS models play in governing cutoff costs as well as mark-up and sales distributions of firms. In Appendix E we repeat the analysis for the non-separable/flexible MO model. Below, we outline the generalized CES model.

2.2 Consumer and Firm Problems

The generalized CES utility function corresponds to a Dixit-Stiglitz utility function with a displaced origin. One special case of it is the CES utility function found in Melitz (2003). Another special case of it was introduced into a heterogeneous-firm model of international trade by Simonovska (2015). Relative to the SIM specification, the generalized CES utility function features an extra parameter that governs the curvature of the utility function and therefore the elasticity of substitution across goods. Specifically, the utility function is:

$$U^c = \left(\int_{\omega \in \Omega} (q^c(\omega) + \bar{q})^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where $q(\omega)$ is individual consumption of good ω, Ω is a compact set containing all potentially produced goods, $\bar{q} > 0$ is a (country-aspecific) constant, and $\sigma \geq 1$ governs curvature.

The preference relation described above is non-homothetic. Moreover, the marginal utility from consuming a good is bounded from above at any level of consumption. Hence, a consumer does not have positive demand for all goods as there exists a choke price such that demand is zero for all goods whose prices exceed it.

It can be shown that demand for a good ω originating in i that is consumed in a positive amount in j is:

$$q_{ij}(\omega) = L_j \left(\frac{(w_j + \bar{q}P_j)(p_{ij}(\omega))^{-\sigma}}{P_{j\sigma}^{1-\sigma}} - \bar{q} \right) \tag{4}$$

where $p_{ij}(\omega)$ is the good's price, and $P_j = \sum_{i=1}^{I} \int_{\omega \in \Omega_{ij}} p_{ij}(\omega) d\omega$ and $P_{j\sigma}^{1-\sigma} = \sum_{i=1}^{I} \int_{\omega \in \Omega_{ij}} (p_{ij}(\omega))^{1-\sigma} d\omega$ are aggregate price statistics.

Relabeling each good by the productivity of its supplier and substituting for the demand function using expression (4), the profit maximization problem of a firm with cost draw c_{ij} is:

$$\pi_{ij}(c_{ij}) = \max_{p_{ij} \geq 0} \quad (p_{ij} - c_{ij}) L_j \left(\frac{(w_j + \bar{q}P_j)(p_{ij})^{-\sigma}}{P_{j\sigma}^{1-\sigma}} - \bar{q} \right) \tag{5}$$

The total profits of the firm are simply the summation of profits flowing from all destinations it sells its good to:

$$\pi_i(c_{i1}, ..., c_{iI}) = \sum_{j=1}^{I} \pi_{ij}(c_{ij}).$$

From (5), notice that there exists a good produced by a firm with cost draw $\bar{c}_{ij}$ whose demand is zero in market j and whose price equals the choke price in that destination. The firm producing this good earns zero profits from its sale. This cutoff cost is the maximum cost that enables firms to serve market j:

$$\begin{aligned} q_{ij}(\bar{c}_{ij}) &= 0 \\ \iff (\bar{c}_{ij})^{\sigma} &= \frac{w_j + \bar{q}P_j}{\bar{q}P_{j\sigma}^{1-\sigma}} \end{aligned} \tag{6}$$

Since the RHS of (6) is entirely determined by the market conditions of the destination j, it must be that $\bar{c}_{ij} = \bar{c}_{jj} \equiv \bar{c}_j$, $\forall i \neq j$.

Each firm takes as given the aggregate price statistics and wages. Taking FOCs of (5) and using expression (6) yields the following implicit equation that characterizes the unique optimal price that a firm charges for a good supplied in a positive amount:

$$(1-\sigma)p_{ij} + \sigma c_{ij} = p_{ij}^{\sigma+1}(\bar{c}_j)^{-\sigma} \tag{7}$$

Furthermore, the quantity of a good sold and a firm's sales are:

$$x_{ij}\left(c_{ij}\right) = L_j\bar{q}\left[\left(\frac{\bar{c}_j}{p_{ij}(c_{ij})}\right)^{\sigma} - 1\right] \tag{8}$$

$$r_{ij}\left(c_{ij}\right) = p_{ij}\left(c_{ij}\right)x_{ij}\left(c_{ij}\right) = L_j\bar{q}\left[(\bar{c}_j)^{\sigma}\left(p_{ij}(c_{ij})\right)^{1-\sigma} - p_{ij}(c_{ij})\right] \tag{9}$$

By restricting the domain of (7) to $c_{ij} \in [0, \bar{c}_j]$, notice that a firm with cost $c_{ij} \to 0$ will charge a price $p_{ij}(c_{ij}) \to 0$, which is the only non-negative price that satisfies the FOC for $\sigma > 1$. Any firm with a higher cost draw will charge a higher price. Finally, the firm with a cost draw $\bar{c}_j$ charges a price that equals its marginal cost, a solution that satisfies (7) and yields zero profits. For any $c_{ij} \in [0, \bar{c}_j]$, the RHS of (7) is increasing and convex in p_{ij} while the LHS is linear and decreasing. For each $c_{ij} \in [0, \bar{c}_j]$ there exists a unique $p_{ij} \in [0, \bar{c}_j]$, which characterizes the unique real number that constitutes the optimal price that a firm charges.[10]

A few key predictions of the model follow. There is incomplete pass-through of costs onto prices. Using the implicit function theorem in expression (7), it follows that $dp_{ij}/dc_{ij} > 0$, so high cost firms charge higher prices. However, price rises by less than proportionally with cost. To see this, define the mark-up as $m_{ij} = p_{ij}/c_{ij}$ and use the implicit function theorem once again to find the sign for dm_{ij}/dc_{ij}. In Appendix A we show that the range for mark-ups is $[1, \sigma/(\sigma-1)]$, and that in this range $dm_{ij}/dc_{ij} < 0$.

Notice that the upper bound on mark-ups is the Dixit-Stiglitz markup found in the Melitz

[10] We ignore complex solutions and other solutions which occur in the negative quadrant since they violate the non-negativity constraint on prices.

(2003) model. This is in contrast to the implications that follow from imposing the restriction that $\sigma = 1$, which corresponds to the SIM model. In that case, expression (7) above yields a closed-form solution for individual firm prices and a prediction that the most productive firm charges an infinite mark-up. The upper bound on mark-ups in the generalized CES model, together with the fact that sales are unbounded above (see expression (9)), has the following implications: the most productive firm can only have infinite sales if it demands infinite labor. Because this would bid up the wage to infinity, only measure zero of firms can attain this outcome. In contrast, in the $\sigma = 1$ model, the infinitely productive firms charge infinitely high mark-ups, without having to demand infinite labor. This means that there can be many such firms—thus firm size is quite large in the tail, which bounds above differences in sizes between exporters and non-exporters. We explore the quantitative implications of this limitation in Section 5.

2.3 Equilibrium

To derive aggregate predictions, we follow the insights by Arkolakis et al. (2015) (ACDR) and we define a new variable $t_{ij} \equiv \frac{p_{ij}}{\bar{c}_j}$. This strategy will greatly simplify the analysis as all integrals will become country-aspecific, thus resulting in very tractable aggregate equilibrium objects. To start, we rewrite the implicit function that characterizes optimal prices as:

$$(1-\sigma)t_{ij} - t_{ij}^{\sigma+1} = -\sigma\left(\frac{c_{ij}}{\bar{c}_j}\right) \tag{10}$$

To solve the model in general equilibrium, we characterize average profits of firms from country i, total sales from a source i in destination j, and the aggregate price statistics (details of the derivation can be found in Appendix A):

$$\pi_i = \sum_{\upsilon=1}^{I}\int_0^{\bar{c}_\upsilon} \pi_{i\upsilon}(c)d\mu_{i\upsilon}(c) = \sum_{\upsilon=1}^{I} \bar{c}_\upsilon \left(\frac{\bar{c}_\upsilon}{\hat{c}_{i\upsilon}}\right)^{\theta} \frac{L_\upsilon \bar{q}\theta}{\sigma^{\theta+1}}\left(\frac{1}{\sigma-1}\right)^{1-\theta}\beta_1 \tag{11}$$

$$T_{ij} = \int_0^{\bar{c}_j} J_i r_{ij}(c)d\mu_{ij}(c) = J_i\bar{c}_j\left(\frac{\bar{c}_j}{\hat{c}_{ij}}\right)^{\theta}\frac{L_j\bar{q}\theta}{\sigma^{\theta}}\left(\frac{1}{\sigma-1}\right)^{1-\theta}\beta_2 \tag{12}$$

$$P_j = \sum_{\upsilon=1}^{I}\int_0^{\bar{c}_j} J_\upsilon p_{\upsilon j}(c)d\mu_{\upsilon j}(c) = \sum_{\upsilon=1}^{I} J_\upsilon \frac{\bar{c}_j^{\theta+1}b_\upsilon^{\theta}}{(\tau_{\upsilon j}w_\upsilon)^{\theta}}\beta_P \tag{13}$$

$$P_{j\sigma}^{1-\sigma} = \sum_{\upsilon=1}^{I}\int_0^{\bar{c}_j} p_{\upsilon j}(c)^{1-\sigma}d\mu_{\upsilon j}(c) = \sum_{\upsilon=1}^{I} J_\upsilon \frac{\bar{c}^{\theta+1-\sigma}b_\upsilon^{\theta}}{(\tau_{\upsilon j}w_\upsilon)^{\theta}}\beta_{\sigma P} \tag{14}$$

where β_1, β_2, β_P and $\beta_{\sigma P}$ are constants that depend on θ and σ. The change of variables was an essential step that allowed us to arrive at the closed-form solutions for the integrals above as it changed the range of integration to be country-aspecific and in particular equal to $[0, 1]$.

Next, we use the aggregates above to characterize equilibrium wages and cost cutoffs. The free-entry (FE) and income-spending (IS) conditions pin down the measure of entrants:

$$w_i f_e = \pi_i \tag{15}$$

$$w_i L_i = \sum_{v=1}^{I} T_{vi} \tag{16}$$

Specifically, plugging (11) and (12) into (15) and (16) yield the measure of entrants: $J_i = \beta_J L_i / f_e$, where β_J is a constant (see Appendix A). To solve for wages, we first impose the income-spending equality (via trade balance), which yields the following characterization for trade shares:

$$\lambda_{ij} = \frac{T_{ij}}{\sum_{v=1}^{I} T_{vj}} = \frac{L_i \left(w_i \tau_{ij} / b_i\right)^{-\theta}}{\sum_{v=1}^{I} L_v \left(w_v \tau_{vj} / b_v\right)^{-\theta}} \tag{17}$$

We then combine market clearing with the expression for trade shares to obtain the following implicit solution for wages:

$$L_i w_i = \sum_{j=1}^{I} L_j w_j \lambda_{ij} \tag{18}$$

$$\iff \frac{w_i^{\theta+1}}{b_i^{\theta}} = \sum_{j=1}^{I} \left(\frac{w_j L_j}{\tau_{ij}^{\theta} \sum_{v=1}^{I} L_v b_v^{\theta} \left(w_v \tau_{vj}\right)^{-\theta}} \right) \tag{19}$$

Lastly, substituting aggregate objects into the cutoff costs yields:

$$\bar{c}_j = \left[(\beta_{\sigma P} - \beta_P) \beta_j \bar{q} f_e^{-1} \right]^{-\frac{1}{\theta+1}} \left[\frac{w_j}{\sum_{v=1}^{I} \frac{L_v b_v^{\theta}}{(\tau_{vj} w_v)^{\theta}}} \right]^{\frac{1}{\theta+1}}. \tag{20}$$

Cost cutoffs are a product of a (model-specific constant) and a general equilibrium object that involves wages, population sizes, trade costs, productivities as well as the Pareto shape parameter θ. In Section 4 below, we show that this cost cutoff expression (modulo proportionality constant) holds for all existing models with separable preferences. This implies that, for a given set of “macro variables” and θ, we can construct cutoff costs in each model, relative to a benchmark country.

2.4 Income Per-Capita and Prices

A key prediction of the model is that prices of identical goods are higher in richer destinations. To see this, apply the implicit function theorem to expression (7) to verify that $dp_{ij}/d\bar{c}_j > 0$. Furthermore, differentiating the cutoff in expression (20) with respect to the destination's per-capita income, w_j, yields a positive elasticity. Combining the two elasticities, as verified in Appendix A.3, implies that the price of each good is increasing in the per-capita income of the destination.

3 Solution Algorithm

As discussed in the introduction, the qualitative predictions of the model are shared by three existing models in the literature. Hence, the contribution of the generalized model would draw from its quantitative fit to data. In order to derive quantitative cross-sectional and aggregate predictions, we need to solve the model via simulation. In principle, the numerical algorithm is cumbersome as it involves a large number of firms and countries. However, if we follow insights by Eaton et al. (2011) and relabel firm-level indicators from costs into "inherent efficiency", we can simulate the latter from a parameter-free uniform distribution only once and re-use them throughout the entire quantitative analysis.

We are interested in deriving predictions for firms from all countries $i = 1, ..., I$ that operate in equilibrium; namely, firms with costs $c_{ii} \in [0, \bar{c}_i]$. In the presence of trade costs, all exporters are a subset of these firms. We proceed as follows. We draw 5,000,000 realization of the uniform distribution on the $[0, 1]$ domain, $U \sim [0, 1]$, we order them in increasing order, and find the maximum realization, denoted by $u_{\max}$. Define $s = \frac{u}{u_{\max}}$. Then, we let $c_{ii} = s^{\frac{1}{\theta}}\bar{c}_i$. Notice that $c_{ii} \in [0, \bar{c}_i]$ by construction, and it has a pdf of $d\mu_{ii}(c_{ii}) = \theta c_{ii}^{\theta-1}\left(\frac{b_i}{w_i}\right)^{\theta}$; yet the normalization allows us to utilize all draws in the subsequent analysis. Multiplying each c_{ii} by the appropriate trade cost yields the cost to serve each market. Comparing the cost (inclusive of trade costs) to the cost cutoff for each source-destination pair determines the set of exporters to every destination.

3.1 From Model to Data

To simulate predictions from the model, we need (i) values for the following parameters $\Theta \equiv (\theta, \sigma, \{L_i\}_{i=1,...,I-1}, \{\tau_{ij}\}_{i,j=1,...,I})$ and (ii) values for the endogenous objects, most notably wages and cost cutoffs ($\{w_i, \bar{c}_i\}_{i=1,...,I-1}$). We normalize population L_i relative to a numeraire country. Similarly, we compute wages and cost cutoffs relative to a numeraire country.

To reduce the computational burden, we leverage equation (17). Since trade shares yield a standard log-linear gravity equation, we first specify a parametric functional form that relates bilateral trade barriers to country-pair-specific characteristics, and we estimate the coefficients that govern the function directly from bilateral trade data, up to a constant, θ. Specifically, motivated by Waugh (2010), we assume a functional form for bilateral trade costs:

$$\log(\tau_{ij}) = \alpha + ex_i + \gamma_h d_h + \gamma_d log(dist)_{ij} + \gamma_t \log(td_{ij}) + \gamma_g cepii_{ij}$$

where α is a constant, ex_i is an exporter-specific fixed effect, d_h is an indicator that takes on the value of one if trade is internal ($i = j$), $dist$ is the distance between country pairs, td_{ij} is the time zone difference between country pairs, and $cepii_{ij}$ is a matrix that contains vectors of indicator variables that capture country-pair-specific characteristics collected from CEPII including: shared border, language (official), language (ethnic), common colony, colony in 1945, legal system, currency, free trade agreements, hegemony of origin/destination over the trade partner. We substitute this functional form for trade costs into the gravity equation, we take logs, and we estimate the coefficients via OLS. In particular, we estimate the following: $\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = S_j - S_i - \theta \log(\tau_{ij})$, with $S_i \equiv \log(L_i(w_i/b_i)^{-\theta})$ $\forall i$ recovered from country-specific fixed effects.[11] We normalize all trade costs relative to their domestic counterparts so that $\tau_{ii} = 1$ ($\forall i$).

Without specifying a value for θ, all trade costs (in logs) are scaled by θ. We take these trade costs as well as estimated country-specific coefficients and we recover implied trade shares: $\log\left(\frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{jj}}\right) = \hat{S}_j - \hat{S}_i - \theta \log(\hat{\tau}_{ij})$. To recover the wages, we substitute estimated trade shares into (18) and use data for L_i.[12] All L_i are computed relative to a numeraire country; hence, the wage rate for that country is set to unity. Finally, we back out cost cutoffs from expression (20). Notice that the summation terms in the cost cutoffs (modulo proportionality factor), as defined in equation (20), can be rewritten in terms of the gravity objects, S_i, only.

Given estimates of S_i, data on population, and estimated wages, we could back out the technology parameters b_i with the definition of $S_i = \log(L_i(w_i/b_i)^{-\theta})$. For each guess of θ, the technology parameters would satisfy $\hat{b}_i = \left(\frac{\exp(\hat{S}_i)}{L_i}\right)^{1/\theta} \hat{w}_i$, where $\hat{w}_i$ is the equilibrium wage rate computed above. At this step, a value for θ would be needed, which would suggest that the vector of technology parameters would have to be jointly estimated with the key parameters, θ and σ. However, given our strategy to identify these two key parameters from moments that relate to US exporters and non-exporters, the technology parameters are not necessary. Moreover, given the predictions of the model that we are interested in, the technology

[11] See Simonovska and Waugh (2014a) for a discussion regarding the separate identification of S_i and ex_i.

[12] An alternative strategy is to use per-capita income directly from the data. Quantitative results obtained using this method are very similar and they are available upon request from the authors.

parameters need not be computed. It is for this reason that these parameters were not included in the set of parameters described at the outset of this section.

To summarize, the first step in our estimation will be to calculate bilateral (asymmetric) trade costs, scaled by any θ, and recover the implied trade shares. The second step will be the computation of wages from market clearing, predicted trade shares, and population data. This allows us to also characterize cost cutoffs relative to a benchmark country. The third and final step will be to simulate firms and identify σ and θ.

3.1.1 Identifying θ and σ

To identify these two parameters, we derive the following two moments from the model that we perfectly match in the US data as documented by BEJK: 1) measured productivity advantage of exporters over non-exporters (in logs) and 2) domestic sales advantage of exporters over non-exporters.

Measured Productivity Advantage. To derive the first moment, we first need to characterize measured productivity. Following BEJK, this object is labor productivity, or total value added divided by employment. The value added for each s from i is:

$$\begin{aligned} va_i(s) &= \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) p_{i\upsilon}(s) x_{i\upsilon}(s) \\ &= \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \bar{c}_\upsilon L_\upsilon \bar{q} (t_{i\upsilon}^{1-\sigma}(s) - t_{i\upsilon}(s)) \end{aligned}$$

where $\delta_{i\upsilon}(s)$ is an indicator that takes on the value of one if firm s from i sells in market υ and $t_{i\upsilon}$ is the change of variables introduced earlier. There are no intermediate goods in this model, so only labor is used for production. Therefore, the value added is the same as the revenue for each s.

Employment of the same s is

$$\begin{aligned} emp_i(s) &= \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \frac{\tau_{i\upsilon} x_{i\upsilon}(s)\, c_{i\upsilon}(s)}{\tau_{i\upsilon} w_{i\upsilon}} \\ &= \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \bar{c}_\upsilon L_\upsilon \bar{q} \frac{t_{i\upsilon}^{\sigma+1}(s) + (\sigma-1) t_{i\upsilon}(s)}{\sigma} \frac{t_{i\upsilon}^{-\sigma}(s) - 1}{w_i} \end{aligned}$$

Notice that due to the iceberg transportation cost τ_{iv}, a firm must produce $\tau_{iv}x_{iv}$ so that x_{iv} units reach the consumer. The measured productivity of s is the ratio of the two objects,

$$mp_i(s) = \log\left(\frac{va_i(s)}{emp_i(s)}\right)$$

It remains to compute the average measured productivity for exporters and non-exporters. To separate firms into these two groups, we use expression (20) to compute the cost draw necessary for a firm from i to reach its easiest export destination: $\tilde{c}^x_{ii} \equiv \max_{v\neq i} \frac{\bar{c}_v}{\tau_{iv}}$. Hence, all firms from i with cost draws below this cutoff necessarily export to at least one destination, while the remaining firms serve the domestic market only. To aggregate over exporters and non-exporters, it is useful to employ the change of variables introduced earlier. Let us characterize $\tilde{t}_{ii}$: the value for t_{ii} that corresponds to the cutoff cost that differentiates exporters from non-exporters. Focus on expression (10) and let $j = i$. Then, solve this for the marginal exporter, i.e., set $c_{ii} = \tilde{c}^x_{ii}$ into expression (10). It yields the desired value implicitly: $(1-\sigma)\tilde{t}_{ii} + \sigma\frac{\tilde{c}^x_{ii}}{\bar{c}_i} = \tilde{t}^{\sigma+1}_{ii}$. Notice that $\tilde{t}_{ii} = f(\frac{\tilde{c}^x_{ii}}{\bar{c}_i}, \sigma)$; namely the variable of interest is a function of country i's domestic cost cutoff and cost cutoff to export as well as σ.

Integrating over the logged measured productivity of all non-exporters yields the average non-exporter logged productivity:

$$\begin{aligned} MP^{NX}_i =& \frac{\theta}{\sigma^\theta(1-\xi_i^\theta)}\int_{\tilde{t}_{ii}}^1 \left[\log\left(t_{ii}^{1-\sigma}(s) - t_{ii}(s)\right) - \log\left(\left[t_{ii}^{-\sigma}(s) - 1\right]\left[\frac{t_{ii}^{\sigma+1}(s) + (\sigma-1)t_{ii}(s)}{\sigma}\right]\right) + \log(w_i)\right] \\ & \left[t_{ii}^{\sigma+1}(s) + (\sigma-1)t_{ii}(s)\right]^{\theta-1}\left[(\sigma+1)t_{ii}^{\sigma}(s) + \sigma - 1\right] dt_{ii}(s) \\ =& \frac{\theta}{\sigma^\theta(1-\xi_i^\theta)}\left[\log(w_i)\left(\frac{\sigma^\theta}{\theta} - \frac{\tilde{t}_{ii}\left(\sigma - 1 + (\tilde{t}_{ii})^\sigma\right)^\theta}{\theta}\right) + \beta^{NX}_{MP,i}(\sigma,\theta,\tilde{t}_{ii})\right] \end{aligned}$$

where $\xi_i \equiv \frac{\tilde{c}^x_{ii}}{\bar{c}_i}$ and $\beta^{NX}_{MP,i}(\sigma,\theta,\tilde{t}_{ii})$ is a constant term given $(\sigma,\theta,\tilde{t}_{ii})$.[13] See Appendix A.4 for the closed form solution of this constant.

To compute the equivalent moment for exporters, we rely on a similar change of variables. For any destination v, we introduce a mapping between $t_{iv}(s)$ and $t_{ii}(s)$ so as to be able to integrate over $t_{ii}(s)$ in every export destination v as we did in the case of non-exporters above. In equation (10), set $c_{iv} = c_{ii}\tau_{iv}$, then $(1-\sigma)t_{iv}(s) - t_{iv}^{\sigma+1}(s) = \left[(1-\sigma)t_{ii}(s) - t_{ii}^{\sigma+1}(s)\right]\frac{\tau_{iv}\bar{c}_i}{\bar{c}_v}$, and this allows us to solve for $t_{ii}(s)$. Therefore, we use the notation $h_{iv}(t_{ii}(s))$ to refer to the implicit solution of $h_{iv}(t_{ii}(s)) = t_{iv}(s)$. Integrating over the logged measured productivity of

[13] The above integration requires the conditional distribution of t_{ii}. The following relates the cost distribution to the pdf of t_{ii}: $d\mu_{ii}(c) = \bar{c}_i\left(\frac{\bar{c}_i}{\hat{c}_{ii}}\right)^\theta \frac{\theta}{\sigma^\theta}\left[(\sigma+1)t_{ii}^\sigma + (\sigma-1)\right]\left[t_{ii}^{\sigma+1} + (\sigma-1)t_{ii}\right]^{\theta-1} dt_{ii}$.

all exporters yields the average exporter logged productivity:

$$
\begin{aligned}
MP_i^{EXP} = &\frac{\theta}{(\xi_i\sigma)^\theta}\int_0^{\tilde{t}_{ii}}\Bigg[\log\left\{\sum_v \delta_{iv}(s)\chi_{iv}L_v\left(h_{iv}(t_{ii}(s))^{1-\sigma}-h_{iv}(t_{ii}(s))\right)\right\}\\
&-\log\left\{\sum_v \delta_{iv}(s)\chi_{iv}L_v\left(h_{iv}(t_{ii}(s))^{-\sigma}-1\right)\frac{h_{iv}(t_{ii}(s))^{\sigma+1}+(\sigma-1)h_{iv}(t_{ii}(s))}{\sigma}\right\}\Bigg]\\
&\left[t_{ii}(s)^{\sigma+1}+(\sigma-1)t_{ii}(s)\right]^{\theta-1}\left[(\sigma+1)t_{ii}(s)^{\sigma}+\sigma-1\right]dt_{ii}(s)+\frac{\theta}{(\xi_i\sigma)^\theta}\frac{\left(\tilde{t}_{ii}(\sigma-1+\tilde{t}_{ii}^{\sigma})\right)^\theta}{\theta}\log(w_i)\\
= &\frac{\theta}{(\xi_i\sigma)^\theta}\beta_{MP,i}^{EX}\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)+\frac{1}{(\xi_i\sigma)^\theta}\left(\tilde{t}_{ii}(\sigma-1+\tilde{t}_{ii}^{\sigma})\right)^\theta\log(w_i)
\end{aligned}
$$

where $\chi_{ij}=\frac{\bar{c}_j}{\bar{c}_i}$ and $\beta_{MP,i}^{EX}\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)$ is a constant for given $\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)$.[14]

Finally, the productivity moment is defined as:

$$
\begin{aligned}
M_{i,prod}^m &= MP_i^{EXP}-MP_i^{NX}\\
&=\frac{\theta}{(\xi_i\sigma)^\theta}\beta_{MP,i}^{EX}\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)-\frac{\theta}{\sigma^\theta(1-\xi_i^\theta)}\beta_{MP,i}^{NX}(\sigma,\theta,\tilde{t}_{ii})\\
&+\frac{\left(\tilde{t}_{ii}(\sigma-1+\tilde{t}_{ii}^{\sigma})\right)^\theta}{(\xi_i\sigma)^\theta(1-\xi_i^\theta)}\log(w_i)-\frac{1}{1-\xi_i^\theta}\log(w_i)
\end{aligned}
\tag{21}
$$

Sales Advantage. The second moment of interest is the domestic sales advantage of exporters. BEJK report this statistic as the ratio between the average domestic sales of exporters to non-exporters. Using equations (9) and (10), and integrating, yields the corresponding statistic in the model:[15]

$$
M_{i,sales}^m=\frac{1-\xi_i^\theta}{\xi_i^\theta}\frac{\int_0^{\tilde{t}_{ii}}\left[t_{ii}(s)^{1-\sigma}-t_{ii}(s)\right]\left[(\sigma+1)t_{ii}(s)^\sigma+(\sigma-1)\right]\left[t_{ii}(s)^{\sigma+1}+(\sigma-1)t_{ii}(s)\right]^{\theta-1}dt_{ii}(s)}{\int_{\tilde{t}_{ii}}^1\left[t_{ii}(s)^{1-\sigma}-t_{ii}(s)\right]\left[(\sigma+1)t_{ii}(s)^\sigma+(\sigma-1)\right]\left[t_{ii}(s)^{\sigma+1}+(\sigma-1)t_{ii}(s)\right]^{\theta-1}dt_{ii}(s)}.
\tag{22}
$$

Notice that this expression also reduces to a constant characterized in terms of parameters $(\sigma,\theta,\tilde{t}_{ii})$.

Having derived the two moments of interest from the model, we exactly identify $\{\theta,\sigma\}$ by perfectly matching the two corresponding moments in US data: $M_{US,sales}^m=M_{US,sales}^d$ and $M_{US,prod}^m=M_{US,prod}^d$, where the superscript d denotes data.

It is worth pointing out that neither of our two moments of interest depend on $\bar{q}$ or f_e as

[14] Although there is no closed form solution for $\beta_{MP,i}^{EX}\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)$ it can be solved numerically. The main takeaway here is that it depends only on $\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)$.

[15] For an explicit closed-form expression of this moment see Appendix A.4.

can be seen from the moment equations derived above. This observation, together with the observation made above that cost cutoffs (modulo proportionality constant) are independent of these parameters, implies that we need not identify these two parameters in order to study the moments of interest in this paper. It is for this reason that $\bar{q}, f_e$ were not included in the set of parameters Θ defined at the outset of this section.

3.1.2 Summary

In the quantitative section below, we demonstrate that the model can jointly reconcile the sales and measured productivity advantage of exporters observed in the data. To evaluate the importance of this flexibility, we then engage in two sets of exercises. First, we derive aggregate pricing predictions from the model and compare them to data. Second, we derive the theoretical welfare gains from trade from the model and we quantify them. Before proceeding to the quantitative exercises, however, we discuss how the model relates to the existing literature.

4 Existing Variable Markup Models

In order to evaluate the contribution of the generalized CES model to the literature, we discuss key predictions from existing theories: Simonovska (2015) (SIM), Behrens et al. (2014) (BMMS), and the separable-preference case of Melitz and Ottaviano (2008) (MO).[16] We derive the predictions within the general framework developed at the outset of the paper for comparability. When framed as such, the models fit within the ACDR class of a strong CES import demand system with a choke price. Along with the generalized CES framework, the three existing models that we examine feature separable preferences, identical log-linear gravity and market clearing equations, as well as identical choke prices (modulo proportionality constant). Hence, the models' welfare gains can be quantified by invoking theoretical results from ACDR.

In general equilibrium, the existing models, alongside the generalized one, yield non-constant demand elasticity along two dimensions. First, in a given destination, firms face heterogeneous demand elasticities and therefore charge different markups depending on their productivity. Second, for a given firm, there is markup variation across destination markets because a firm faces different consumer willingness to pay. Therefore, these models produce an empirically-relevant non-degenerate markup distribution and predict that firms engage in price discrimination across markets with different characteristics. The models' prediction of price discrimination is key because it represents a plausible explanation for the observation that tradable consumption goods are more expensive in countries with higher per-capita income levels, as documented

[16] This corresponds to MO preferences with $\eta = 0$. For completeness, we quantify the more general case where $\eta > 0$ in Appendix E.

in Alessandria and Kaboski (2011).

The distinctive feature of all three models is that preferences are non-homothetic. In particular, the marginal utility from consuming a good is bounded from above so that a consumer does not have positive demand for all goods. There exists a reservation price at which firms see zero demand. The reservation price is higher in richer markets, reflecting higher willingness to pay. Demand elasticities directly reflect this reservation price and drive firms' pricing behavior. In richer countries, firms face lower demand elasticities, which allows them to raise their price-cost margin.

The prediction related to cross-country variation in prices, however, could also be derived from respective counterparts to the models where firms are monopolistically competitive and homogeneous in productivity.[17] Firm productivity heterogeneity, therefore, enriches the pricing predictions of these models in that it allows for a non-degenerate mark-up distribution across firms within a country. In particular, the models predict that more productive firms enjoy higher mark-ups relative to less productive ones, as documented in De Loecker and Warzynski (2012). Mark-up differences, however, directly translate into differences in value added, and consequently measured productivity, and sales across firms. It is these last two predictions of these models that the existing literature has largely ignored.[18]

Below we demonstrate that none of the existing models can simultaneously account for the cross-sectional variation in firm sales and productivity. We focus on moments from the distribution of sales and measured productivity per worker for US exporters and non-exporters for a particular reason: these moments have been provided by BEJK and Bernard et al. (2012), hence making it possible for us to use them as calibration targets in our quantitative analysis. First, assuming that firm productivity is Pareto distributed, we show that, the distribution of firm sales in all the models is characterized by the Pareto shape parameter alone. Second, we show that the same holds true for the distribution of mark-ups. These two facts imply that the distribution of measured productivity is also governed by the key Pareto shape parameter. Hence, the three models can, at best, match moments from the distribution of firm measured productivity or the distribution of firm sales, but they lack a degree of freedom to be able to reconcile both. We quantify the implications of this limitation in Section 5.

[17]See Bertoletti and Etro (2014) for a characterization of a wide class of monopolistically competitive models with homogeneous firms, which includes the three utility parameterizations used in this paper.

[18]Recently, ACDR point to these moments, but the authors do not evaluate the models' quantitative fit to data. In addition, BMMS examine the measured-productivity moment in their model.

4.1 Simonovska (2015)—SIM

Assume each country is populated by identical consumers of measure L, whose utility is:

$$U^c = \int_{\omega\in\Omega} \log(q^c(\omega) + \bar{q})d\omega \tag{23}$$

where $q^c(\omega)$ is individual consumption of good ω, Ω is a compact set, and $\bar{q} > 0$ is a country-aspecific constant. Notice that this is a special case of the general model in Section 2, which occurs when $\sigma \to 1$.

To solve the model, we follow the same steps as for the general model above. The resulting price, mark-up, and sales are:

$$p_{ij}(c_{ij}) = (c_{ij}\bar{c}_j)^{\frac{1}{2}} \tag{24}$$

$$m_{ij}(c_{ij}) = \left(\frac{\bar{c}_j}{c_{ij}}\right)^{\frac{1}{2}} \tag{25}$$

$$r_{ij}(c_{ij}) = L_j\bar{q}\left(\bar{c}_j - (\bar{c}_j c_{ij})^{1/2}\right). \tag{26}$$

To solve the model we calculate total sales and pin down the maximum cost allowed in order to serve market j using the cost of the firm with zero demand:

$$T_{ij} = \frac{\bar{q}L_i f_e^{-1}}{(\theta+1)(1+2\theta)} L_j \hat{c}_{ij}^{-\theta}\bar{c}_j^{\theta+1}$$

$$\bar{c}_j^{SIM} \equiv \bar{c}_j = \left[f_e\beta_J^{SIM}\bar{q}^{-1}\right]^{\frac{1}{\theta+1}} \left[\frac{w_j}{\sum_{v=1}^{I}\frac{L_v b_v^\theta}{(\tau_{vj}w_v)^\theta}}\right]^{\frac{1}{\theta+1}}$$

where $\beta_J^{SIM} = \frac{(1+2\theta)(\theta+1)^2}{\theta}$. All else equal, a higher income raises the cutoff cost and prices/markups. Clearly, the cost cutoff (modulo proportionality constant) is identical to the one for the generalized model in expression (20).

4.2 Melitz and Ottaviano (2008) with $\eta = 0$—MO

The framework builds on Melitz and Ottaviano (2008) and, in particular, on the extension to general equilibrium that is outlined in the Web Appendix of Simonovska (2015). Additionally, we let $\eta = 0$, which makes the preference class separable. Assume that each country is populated by identical consumers of measure L, whose utility function is:

$$U^c = \int_{\omega\in\Omega} q^c(\omega)d\omega - \frac{1}{2}\gamma\int_{\omega\in\Omega}\left(q^c(\omega)\right)^2 d\omega,$$

where $\gamma > 0$ governs the degree of product differentiation between the varieties.
The resulting price, mark-up, and sales are:

$$p_{ij}(c_{ij}) = \frac{1}{2}\left(c_{ij} + \bar{c}_j\right) \tag{27}$$

$$m_{ij}(c_{ij}) = \frac{1}{2}\left(\frac{c_{ij} + \bar{c}_j}{c_{ij}}\right) \tag{28}$$

$$r_{ij}\left(c_{ij}\right) = \frac{L_j}{4\bar{c}_j\gamma}\left(\bar{c}_j^2 - c_{ij}^2\right). \tag{29}$$

We refer to the Web Appendix of Simonovska (2015) for derivations of the positive relationship between destination per-capita income and prices/markups.

To solve the model we calculate total sales and the cost cutoffs, which are:

$$T_{ij} = \frac{L_i f_e^{-1} L_j \bar{c}_j^{\theta+1}}{\hat{c}_{ij}^{\theta} 2\gamma(\theta+2)(\theta+1)}$$

$$\bar{c}_j^{MO} \equiv \bar{c}_j = \left[f_e \beta_J^{MO} 2\gamma\right]^{\frac{1}{\theta+1}} \left[\frac{w_j}{\sum_{v=1}^{I} \frac{L_v b_v^{\theta}}{(\tau_{vj} w_v)^{\theta}}}\right]^{\frac{1}{\theta+1}}$$

where $\beta_j^{MO} = (\theta+2)(\theta+1)$. Clearly, the cost cutoff (modulo proportionality constant) is identical to the one for the generalized model from expression (20).

4.3 Behrens et al. (2014)—BMMS

Behrens et al. (2014) assume that each country is populated by identical consumers of measure L, whose utility function is:

$$U^c = \int_{\omega \in \Omega} \left[1 - e^{-\alpha q^c(\omega)}\right] d\omega$$

where $\alpha > 0$ governs the degree of "love of variety."

The resulting price, mark-up, and sales are:

$$p_{ij}(c_{ij}) = \frac{c_{ij}}{W\left(\frac{c_{ij}}{\bar{c}_j}e\right)} \tag{30}$$

$$m_{ij}(c_{ij}) = \frac{1}{W\left(\frac{c_{ij}}{\bar{c}_j}e\right)} \tag{31}$$

$$r_{ij}\left(c_{ij}\right) = \frac{L_j}{\alpha}c_{ij}\frac{1 - W\left(\frac{c_{ij}}{\bar{c}_j}e\right)}{W\left(\frac{c_{ij}}{\bar{c}_j}e\right)}, \tag{32}$$

where W denotes the Lambert W function with argument $\left(\frac{c_{ij}}{\bar{c}_j}e\right)$.[19] As in the generalized CES model, per-capita income is positively associated with prices and mark-ups (see Behrens et al. (2014) for derivations).

We follow Behrens et al. (2014) in order to derive closed form solutions of aggregates that involve the Lambert function. We define the following change of variables:

$$z \equiv W\left(\frac{c_{ii}}{\bar{c}_i}e\right) \Rightarrow \frac{c_{ii}}{\bar{c}_i}e = ze^z.$$

Then $c_{ij} = \bar{c}_j z_{ij} e^{z_{ij}-1}$ and $dc_{ij} = \left(e^{z_{ij}-1}\bar{c}_j\right)(1 + z_{ij})dz_{ij}$ with integration bounds between 0 and 1. Incorporating this change of variable and solving for the free entry measure of entrants we obtain total sales and cutoffs:

$$T_{ij} = \frac{L_j}{\alpha}L_i f_e^{-1}\kappa_1(\theta)\left(\bar{c}_j\right)^{\theta+1}\hat{c}_{ij}^{-\theta}, \text{ where } \kappa_1(\theta) = \theta e^{-(\theta+1)}\int_0^1 (z^{-1} - z - 1 + z^2)(ze^z)^\theta e^z dz$$

$$\bar{c}_j^{BMMS} \equiv \bar{c}_j = \left[\frac{\alpha f_e}{\kappa_1(\theta)}\right]^{\frac{1}{\theta+1}}\left[\frac{w_j}{\sum_{v=1}^{I}\frac{L_v b_v^\theta}{(\tau_{vj}w_v)^\theta}}\right]^{\frac{1}{\theta+1}}$$

Clearly, the cost cutoff (modulo proportionality constant) is identical to the one for the generalized model from expression (20).

4.4 Firm Sales and Markup Distribution

In this section, we show that the sales and markup distributions in the three models outlined above depend only on the parameter θ (the dispersion parameter of the Pareto productivity distribution), which makes it impossible for any of the three models to jointly match empirical moments from the sales and productivity distributions.

[19] The Lambert function is defined as $\varphi = W(\varphi)e^{W(\varphi)}$.

We follow Arkolakis (2010) and Eaton et al. (2011) and examine the (normalized) sales of a firm from country i in destination j, relative to the average sales there, $\bar{r}_{ij} = \int_0^{\bar{c}_j} r_{ij}(c) d\mu_{ij}(c)$:[20]

$$\tilde{r}\left(\frac{c_{ij}}{\bar{c}_j}\right) = \frac{r_{ij}(c_{ij})}{\bar{r}_{ij}(\bar{c}_j)}$$

The normalized firm-level sales for the three models are:

$$\tilde{r}^{SIM}\left(\frac{c_{ij}}{\bar{c}_j^{SIM}}\right) = (1+2\theta)\left(1-\left(\frac{c_{ij}}{\bar{c}_j^{SIM}}\right)^{\frac{1}{2}}\right) \qquad \text{if } 0 \leq \frac{c_{ij}}{\bar{c}_j^{SIM}} \leq 1 \tag{33}$$

$$\tilde{r}^{MO}\left(\frac{c_{ij}}{\bar{c}_j^{MO}}\right) = \frac{\theta+2}{2}\left(1-\left(\frac{c_{ij}}{\bar{c}_j^{MO}}\right)^{2}\right) \qquad \text{if } 0 \leq \frac{c_{ij}}{\bar{c}_j^{MO}} \leq 1 \tag{34}$$

$$\tilde{r}^{BMMS}\left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}\right) = \frac{1}{\kappa_2(\theta)}\left[e^{W\left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}e\right)-1} - \left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}\right)\right] \qquad \text{if } 0 \leq \frac{c_{ij}}{\bar{c}_j^{BMMS}} \leq 1, \tag{35}$$

where $\kappa_2(\theta) = \theta e^{-(\theta+1)} \int_0^1 (z^{-1} - z)(ze^z)^\theta e^z dz$ is a positive constant that depends only on θ. Notice that in all three cases, sales are increasing, concave in firm productivity (inverse of cost), and bounded from above.[21] Furthermore, the normalized firm-level sales are once again neither source nor destination specific since cost draws relative to cutoffs, $c_{ij}/\bar{c}_j$, lie between 0 and 1. For this reason there are no country subscripts on the normalized firm sales.

Using these expressions for normalized sales, we derive the distributions of (normalized) firm sales following the steps from Eaton et al. (2011). The distribution is:

$$Pr[\tilde{R}_{\geq}\tilde{r}|\tilde{R}_{\geq}\tilde{r}^{min}] = 1 - F(\tilde{r}) = \left(\frac{c_{ij}}{\bar{c}_j}\right)^\theta, \tag{36}$$

where $\tilde{r}^{min} = \tilde{r}(\bar{c}_j/\bar{c}_j)$ and F is the distribution of $\tilde{r}$. Therefore, for each model we can derive the distribution of normalized sales as:

$$F^{SIM}(\tilde{r}) = 1 - \left[1 - \frac{\tilde{r}}{2\theta+1}\right]^{2\theta}$$

$$F^{MO}(\tilde{r}) = 1 - \left[1 - \frac{2\tilde{r}}{\theta+2}\right]^{\frac{\theta}{2}}$$

$$F^{BMMS}(\tilde{r}) = 1 - \left[(\tilde{r}\kappa_2(\theta))^\theta \left(\frac{1}{W\left([1-F^{BMMS}(\tilde{r})]^{1/\theta}e\right)} - 1\right)^{-\theta}\right],$$

[20] $r_{ij}(c)$ are model-specific firm revenues. For the SIM, MO, and BMMS models they are given by (26), (29), and (32) respectively.

[21] In the three cases, $c_{ij} = \bar{c}_j$ yields zero sales and $c_{ij} \to 0$ yields positive but finite sales.

where the last expression constitutes an implicit function. Notice that these distributions are a function solely of the parameter θ. Although there is no closed form solution for the BMMS distribution, it is evident that it is implicitly defined as a function of θ.

A similar result applies for the markup distributions. Using the definitions for firm-level markups given above for all three models, the average markup in each model is:

$$\begin{aligned}
\bar{m}^{SIM} &= \frac{\theta}{\theta - 0.5} \\
\bar{m}^{MO} &= \frac{\theta - 0.5}{\theta - 1} \\
\bar{m}^{BMMS} &= \kappa_2(\theta)
\end{aligned}$$

Let $\tilde{m}\left(\frac{c_{ij}}{\bar{c}_j}\right)$ be the markup relative to the mean markup. Its distribution in each model is:

$$\begin{aligned}
\Gamma^{SIM}(\tilde{m}) &= 1 - \left(\frac{\theta - 0.5}{\theta\tilde{m}}\right)^{-2\theta} \\
\Gamma^{MO}(\tilde{m}) &= 1 - \left[\frac{2\theta - 1}{\theta - 1}\tilde{m} - 1\right]^{-\theta} \\
\Gamma^{BMMS}(\tilde{m}) &= 1 - \left[e^{W\left([1-\Gamma^{BMMS}(\tilde{m})]^{1/\theta}e\right)-1}\left(\frac{1}{\tilde{m}}\right)\left(\frac{1}{\kappa_2(\theta)}\right)\right]^{\theta},
\end{aligned}$$

where once again the only parameter that governs the distributions is θ.

The two sets of distributions derived above demonstrate the limitations of the existing models: they are not flexible enough to jointly reconcile moments from the distribution of firm sales and measured productivity. Due to space constraints, we relegate derivations of the two moments that characterize the exporter sales and measured productivity advantage for each model to Appendices B-D.

4.5 Linear Demand with Non-Separable Preferences

Before proceeding to the quantitative analysis, we discuss the non-separable version of the MO model, namely the case in which $\eta > 0$. We outline the model, derive moments so as to identify the model's parameters, and report quantitative results in Appendix E. We show that the model is able to jointly match the sales and the measured productivity advantage of exporters in US data. The reason behind this flexibility is the following: while the model generates identical functional forms for the distributions of firm sales and mark-ups as its restricted counterpart, the general model yields different cost cutoff expressions. In fact, the cost cutoffs are no longer solely characterized by "macro variables" and θ as was the case for all the other models that we examine in this paper (including the generalized CES model). In the non-separable MO model, $\eta > 0$ affects the relative cutoffs across destinations, thus yielding a flexible number of exporters and non-exporters. The implication is the following: a value of θ and a value of η can reconcile the measured productivity advantage of exporters as well as the sales advantage by adjusting relative cutoffs rather than the curvature of the sales distribution.

While this added flexibility of the model is a desirable feature in theory, we show that, in order for the model to match the moments in US data, it would require an unreasonably high fraction of exporters—in fact, it would predict that exporters are in the majority, which is not in line with data. In addition, the model would predict an extremely high level of price discrimination and an extremely low variance in sales, whose distribution is still governed by the Pareto shape parameter. As we demonstrate below, the separable non-homothetic models quantitatively outperform this non-separable framework along all dimensions in the data.

5 Quantitative Analysis

In this section, we report the quantitative fit of the four separable models and we discuss the welfare implications.

5.1 Data

We use data on 71 countries for the year 2004. We construct bilateral trade shares, $\lambda_{ij} = \frac{X_{ij}}{Y_j}$, using nominal trade flows X_{ij} and gross output, adjusted for trade imbalances Y_j[22], from UNIDO. To estimate parameters from the gravity equation, we use country-pair gravity variables from CEPII. The Penn World Table (PWT) 8.0 provides our variable L, population.

We also rely on two moments that characterize firm behavior: US exporter sales advantage in the domestic market, which is 4.8 according to BEJK, and US exporter measured productivity advantage (in logs), which is 33 percent according to BEJK. These two moments serve to identify the two parameters that govern the size and productivity distribution of firms (θ, σ).

5.2 Results

The quantitative results for the generalized CES model verify our claim that adding curvature to the demand function (relative to the SIM benchmark), or allowing σ to be greater than one, is necessary to jointly match the two firm-level moments of interest. Notice that σ affects the size distribution of firms by varying the substitution across goods. As σ increases, the market

[22] For each country, this corresponds to gross manufacturing production minus manufacturing exports (in the sample) plus manufacturing imports (in the sample) to j.

power of each firm is smaller, and more productive firms gain sales relative to less productive firms as consumers do not value variety as much. The Pareto shape parameter determines the variability in firm productivity, so a lower θ (more variance) raises the measured productivity advantage of more productive firms.

Table 1: Moments and Parameters

Model	Data/Targets	σ	θ	Simulated Moment
Generalized CES	$M_{sales} = 4.80$, $M_{prod} = 0.33$	1.41	1.92	$M_{sales} = 4.80$, $M_{prod} = 0.33$
SIM (MP)	$M_{prod} = 0.33$	1	2.11	$M_{sales} = 4.16$, $M_{prod} = 0.33$
SIM (SA)	$M_{sales} = 4.80$	1	3.33	$M_{sales} = 4.80$, $M_{prod} = 0.20$
MO ($\eta = 0$) (MP)	$M_{prod} = 0.33$	–	2.44	$M_{sales} = 3.04$, $M_{prod} = 0.33$
MO ($\eta = 0$) (SA)	$M_{sales} = 4.80$	–	6.46	$M_{sales} = 4.80$, $M_{prod} = 0.10$
BMMS (MP)	$M_{prod} = 0.33$	–	2.37	$M_{sales} = 3.47$, $M_{prod} = 0.33$
BMMS (SA)	$M_{sales} = 4.80$	–	5.09	$M_{sales} = 4.80$, $M_{prod} = 0.13$

Table 1 displays the calibrated parameter values for all the separable models. For the existing models, we calibrate θ to match either the measured productivity advantage (MP) or the sales advantage of exporters (SA). We demonstrated theoretically that it is impossible to match both moments with the same value for θ. In the second row of the table, we successfully match only the productivity advantage in the SIM model with a θ close to the general model. In the third row, for a higher value of the Pareto shape parameter, the sales advantage is attainable but at the cost of a very low productivity advantage.

The separable MO and BMMS models behave qualitatively similarly to the SIM model, but there are notable quantitative differences.[23] Having reconciled the productivity advantage, the two models struggle more to generate a dispersion in sales. Notice that, when the SIM model matches the measured productivity dispersion, it generates a sales dispersion that falls only somewhat short of the moment observed in the data. This is further confirmed by the fact that the generalized CES model requires a value for σ that is only 40% higher than unity to match the sales advantage in the data, conditional on matching the measured productivity with a value for the Pareto shape parameter of 1.92, which is only slightly below the 2.11 value that the SIM model requires.

[23]The value of the Pareto parameter that we obtain is lower than in Behrens et al. (2014) who calibrate it to be 8.5. Our strategies are different in terms of the sample of countries used, wages (they use labor income share for gross output), and, most importantly, trade cost specification: Behrens et al. (2014) posit that trade costs are symmetric.

5.3 Model Predictions and Fit to Data

Given that the general model is consistent with the cross-sectional observations on firm productivity and sales, we also test how the model fits other aspects of the data. Throughout, we maintain a comparison to the fit of the restricted models under the two values for θ reported above. We examine the following moments in the data: i) percentage of US firms that export, ii) the export intensity of exporting firms, iii) the standard deviation in log domestic sales, iv) the elasticity of prices of tradables with respect to per capita income, and v) the average markup.

5.3.1 Exporting Firms

The first prediction that we examine is the fraction of firms that export. Bernard et al. (2012) report that in 2002, 18% of American manufacturing firms exported, and BEJK report this number to be 21% in 1992 data. In our model, there are two determinants of export entry: i) iceberg trade barriers raise the cost to serve foreign markets and ii) destination specific aggregates determine the cutoff cost to sell in each country. The fraction of firms that export is given by the following equation:

$$M^m_{i,ex} = \frac{(\tilde{c}^x_{ii})^\theta}{(\bar{c}_i)^\theta}, \tag{37}$$

where the numerator is the familiar cost cutoff for firms from country i that serve at least one export destination j. We are interested in the relative cutoffs of the US and its easiest exporting destination, which corresponds to Canada in the simulated models.

Table 2: US Exporters, % of Total

Model	Data	σ	θ	% of Firms that Export
Generalized CES	18-21%	1.41	1.92	41.4%
SIM (MP)	18-21%	1	2.11	42.7%
SIM (SA)	18-21%	1	3.33	49.2%
MO ($\eta = 0$) (MP)	18-21%	–	2.44	44.8%
MO ($\eta = 0$) (SA)	18-21%	–	6.46	57.1%
BMMS (MP)	18-21%	–	2.37	44.4%
BMMS (SA)	18-21%	–	5.09	54.5%

Table 2 presents the results for the four models along with the data. The generalized CES model and the restricted models, under the MP parameterization, predict a very similar fraction of exporters, which is below 50%—hence, exporters are in the minority as documented in BEJK.[24] The small difference in the values of the last column are solely attributed to the (relatively) small difference in the Pareto shape parameter across the calibrated models. However, when the restricted models fit the sales advantage moment, they severely overpredict the fraction of exporters, which is problematic.

5.3.2 Export Intensity

BEJK report that even the small fraction of firms that do export sell mostly at home. To evaluate the models' predictions along this dimension, for US exporters, we also compute the fraction of total firm sales that are exported and call this the export intensity:

$$EXINT_i(s) = \frac{\sum_{v\neq i} \delta_{iv}(s)\bar{c}_v L_v \bar{q}(t_{iv}^{1-\sigma}(s) - t_{iv}(s))}{\sum_{v=1}^{I} \delta_{iv}(s)\bar{c}_v L_v \bar{q}(t_{iv}^{1-\sigma}(s) - t_{iv}(s))}$$

with firms indexed by s. Then, as in BEJK, we measure the percentage of exporters that fall into certain ranges of export intensity.

Table 3: % of Exporting Plants Conditional on Export Intensity

Exp. Intensity (%)	Data (%)	General CES	SIM (MP)	MO ($\eta = 0$) (MP)	BMMS (MP)
0-10	66	88.3	85.9	79.5	82.2
10-20	16	11.4	13.7	20.0	17.3
20-30	7.7	0.1	0.2	0.1	0.2
30-40	4.4	0.1	0.1	0.1	0.1
40-50	2.4	0.1	0.1	0.1	0.1
50-60	1.5	0	0	0.1	0.1
60-100	2.8	0	0	0.1	0

In Table 3, we report the moments in the data and the models. We condition on only exporting firms and split export intensity into deciles in order to measure the percentage of firms that fall within a certain range of export intensity. For example, the first row shows that in the generalized CES model 88.3% of exporting firms have export revenue that is less than 10% of their total revenue. In the simulated generalized CES model, there are many very small exporting firms, more-so than in the data. Part of this is due to the fact that trade costs are large when θ is only 1.92. This increases the difference between the domestic cutoff cost and the cutoff for exporting. Still, we are able to pick up the small number of exporters that have

[24] All four models overpredict the fraction of exporters. Similarly, BEJK find that their calibrated model overpredicts the fraction of exporters—the authors obtain a value of 51%.

a very large export intensity in the lower rows. 0.1% of the simulated exporters have an export intensity between 40-50%, and none greater than 50%.

Reducing σ can also reduce the fraction of exporters with very low export intensity because with less substitution there is a lower sales advantage for the very productive firms. When $\sigma = 1$, the value in the first row of the fourth column decreases, although by little. The separable MO and BMMS models show little bit better predictions for the portion of exporters with higher export intensity. All the separable models that match the SA moment yield similar numbers but exhibit higher weight on lower export intensities.

5.3.3 Variability of Domestic Sales

We compute the standard deviation of the log of normalized domestic sales. Log (normalized) domestic sales are firm domestic sales relative to total domestic sales: $\log\left(\frac{r_{ii}(s)}{T_{ii}}\right)$ (as given by (9) and (12)). T_{ii} is constant (as shown in Section 2), so the standard deviation of log domestic sales relative to total domestic sales is equivalent to the standard deviation of log domestic sales, but the normalization allows us to compute the desired statistics without having to calibrate additional parameters from the models.

For the generalized CES model, the value amounts to 1.26, which falls somewhat short of the statistic in the US in 1992 of 1.67, as reported in BEJK. Similarly, BEJK struggle to match this statistic using the model that they develop. As in BEJK, higher values of σ raise the sales variance, but they also raise the sales advantage of exporters relative to non-exporters. To the extent that we discipline the model along the second dimension, we fall short along the first.

In the SIM model, the statistics are comparable and amount to 1.20 (MP) and 1.22 (SA). The values are considerably lower in the remaining separable models. The standard deviations of log domestic sales in the MO model are 1.04 (MP) and 1.17 (SA), whereas in BMMS they are 1.10 (MP) and 1.19 (SA).

5.3.4 Prices and Income

An important aspect of the data that the models analyzed in this paper attempt to explain is price dispersion across markets. A major feature of variable markup models is that they can explain a large portion of the variation in the prices of identical tradable goods (see the discussion in Simonovska (2015)). In particular, as argued in previous sections, the models yield a positive relationship between prices and per-capita income of destinations. Below, we quantify this relationship and we compare the findings to those that we obtain from two commonly-employed price datasets.

To begin, we investigate the relationship between price and income in the data using prices reported by the Economist Intelligence Unit (EIU) and the International Comparison Program (ICP). The EIU reports prices of 110 goods sold in all the countries in our sample. The EIU surveys the prices of individual goods across various cities in two types of retail stores: mid-priced, or branded stores, and supermarkets, or chain stores. The dataset contains the nominal prices of goods and services, reported in local currency, as well as nominal exchange rates relative to the US dollar, which are recorded at the time of the survey. While in the majority of the countries, price surveys are conducted in a single major city, in 17 of the 71 countries multiple cities are surveyed. For these countries, we use the price data from the city which provided the maximum coverage of goods. In most instances, the location that satisfied this requirement was the largest city in the country.

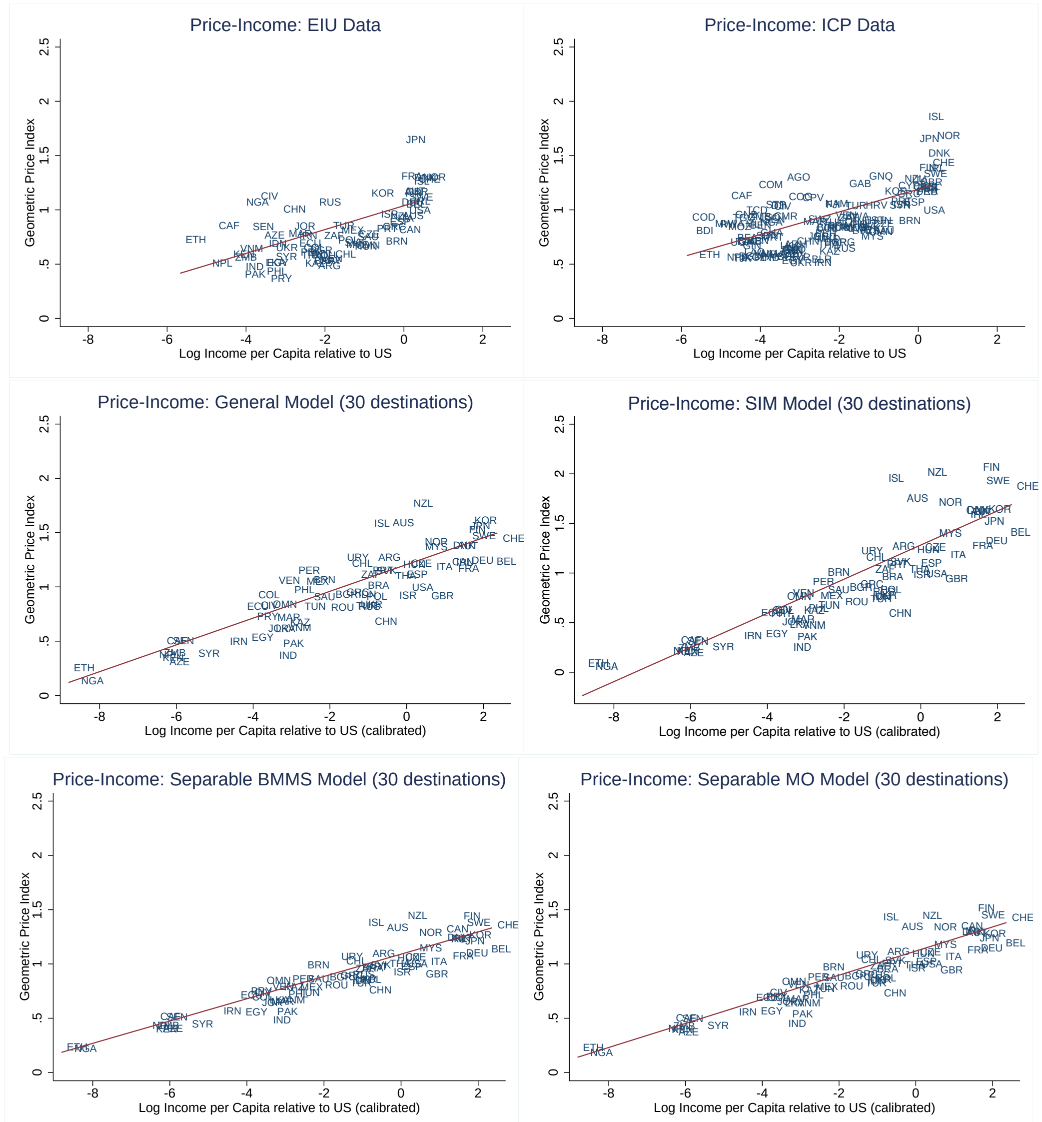
For comparison, we also examine prices from the ICP. The ICP collects price data on goods with identical characteristics across retail locations in 123 countries during the 2003-2005 period. The basic-heading level represents a narrowly-defined group of goods for which expenditure data are available. The data set contains a total of 129 basic headings, which include goods and services. We employ a subsample of 62 tradable categories in order to maintain consistency with the models' assumption that all products are potentially tradable.[25] For both datasets, we construct price indices by taking a geometric average across goods within a country, where all individual good prices are normalized relative to the US.

We follow a similar strategy to generate price indices from the models. However, before constructing price indices, we need to simulate individual good prices. We proceed by following a simulation and sampling methodology introduced by Simonovska and Waugh (2014a), which aims to replicate steps taken to construct the ICP database. In particular, we construct a set of "common" goods and then we draw 100 random samples of 110 products from the set. We compute relative prices and geometric average price indices for each sample and we plot the mean index across all 100 samples for each country. What remains is to discuss the definition of a "common" good. We define a good to be "common" if it appears in at least 30 destinations—nearly half the destinations used in our analysis. Since each good is produced by a single firm, this rule implies that we consider firms that serve at least 30 destinations. The motivation to follow this rule is that Eaton et al. (2004) report that only 1.5% of exporters serve more than 50 destinations but many exporters (20%) serve at least 10 destinations. We choose a value in the middle that would still include a significant number of exporters.

Figure 1 illustrates the geometric price index, plotted against per capita income, in the data and in the four simulated models, when the Pareto shape parameter is calibrated to match the measured productivity moment. In the EIU data (with 71 countries), the slope of the best fit

[25] For more detailed information about the ICP data, see the discussions in Simonovska and Waugh (2014a) and Deaton and Heston (2010).

Figure 1: Geometric Mean of Relative Prices versus Income per Capita



line is a statistically significant 0.11 (standard deviation, 0.015). For robustness, in the top right plot, we show results obtained from the ICP database, which samples a different set of goods than the EIU one and covers a broader set of countries (123 in total). The price-income relationship is almost identical to the one in the EIU data, with the slope of the line of best fit being equal to 0.10 (standard deviation, 0.011). We simulate the models and compute price indices for the set of countries covered in the EIU database. The generalized CES model yields a (statistically significant) slope of 0.12 (standard deviation, 0.009), while its more restricted counterpart yields a coefficient of 0.17 (standard deviation, 0.011), under the calibration that targets measured-productivity. Similarly, the BMMS model yields a slope of 0.10 (standard deviation, 0.006), while the separable MO model yields a slope of 0.11 (standard deviation, 0.006).[26] Hence, the models behave similarly along this dimension and are quantitatively in line with the data.

5.3.5 Average Markups

The average markup is another important price statistic emphasized by the empirical trade and macro literature. Jaimovich and Floetotto (2008) conduct a survey of the literature and document that the average markup found using value added data is in the range of 20-40%. We compute the average markup on domestic sales for domestic firms in our simulation as $\int_0^{\bar{c}_i} \frac{p_{ii}(c_{ii})}{c_{ii}} d\mu_{ii}(c)$. Although we use only domestic firms[27], the average is not source-specific so this number would not change if foreign firms were included.

In the generalized CES model, the average markup of producers selling domestically amounts to 31%, which lies in the middle of the range provided by Jaimovich and Floetotto (2008). In the restricted counterpart, the average markups are 31% and 18% for the MP and SA calibrations, respectively. Hence, when the restricted model's parameters are calibrated to match the measured-productivity advantage of exporters, the model yields a comparable average mark-up to the generalized counterpart. As in De Loecker and Warzynski (2012), exporters have higher markups than non-exporters. The relative markups of exporters to non-exporters amount to 5.7 in the generalized CES model and in the restricted model they amount to roughly 6.4 between the two calibration strategies.

In the remaining two models, the average level of mark-ups is more sensitive to the calibration of choice. In the BMMS model, the mark-ups are 32% (MP) and 12% (SA), respectively. In the separable MO model that matches the productivity advantage moment the mark-up is 35%, however the mark-up drops to a mere 9% in the alternative calibration. Hence, the three separable models yield average mark-ups that are in line with data when their parameters are calibrated to match the measured-productivity moments, but they predict significantly lower mark-ups under the alternative calibration strategy. The sensitivity of the average mark-up to the calibration strategy of choice appears to be lowest for the SIM model.

5.4 Welfare

Having demonstrated the quantitative fit of the models, we turn to evaluate the predicted welfare gains from trade. We view this exercise as quantifying the cost of tractability: namely,

[26] When the models' parameters are calibrated to match the sales advantage moment, the price elasticities are 0.13 (0.005), 0.06 (0.002), and 0.06 (0.002) for the SIM, MO, and BMMS models, respectively, where standard deviations are in parentheses.

[27] Domestic firms are the basis for the markup estimates in Basu and Fernald (1997), Hall (1988), Roeger (1995), Norrbin (1993) (US), and Martins et al. (1996) (OECD).

the mismeasurement in welfare that would occur by relying on a model that lacks the flexibility to reconcile salient features of the data.

An advantage of all the models analyzed in this paper, including the generalized CES model with an unbounded Pareto productivity distribution, is that they fit within the class of models studied by Arkolakis et al. (2015) (ACDR). We can therefore refer to their welfare results for a large change in trade costs and provide a quantitative assessment using our parameter estimates. An intuitive statistic that we will report is the consumer welfare gain with current domestic expenditure shares relative to a country in autarky. We interpret this statistic as representing the welfare gains of moving away from autarky and toward the trade equilibrium in year 2004. We follow ACDR and compute the following welfare gains from trade:

$$\Delta W_j = \left(\frac{1}{\lambda_{jj}}\right)^{\left(1-\frac{\rho}{1+\theta}\right)/\theta} - 1, \tag{38}$$

where λ_{jj} is the domestic expenditure share in the data. ρ stands for the weighted average markup elasticity with respect to marginal cost (or one minus the price-cost pass through elasticity).[28]

Therefore, what remains is to characterize ρ in the models. In the generalized CES model, markups are the implicit solution to the following function:

$$H(m_{ij}, c_{ij}) = (1-\sigma)m_{ij} + \sigma - m_{ij}^{\sigma+1} v_{ij}^{\sigma},$$

where $v_{ij} \equiv c_{ij}/\bar{c}_j$ and m_{ij} is the mark-up. The implicit function theorem implies that $\frac{dm_{ij}}{dv_{ij}} = \frac{\sigma m_{ij}^{\sigma+1} v_{ij}^{\sigma-1}}{(1-\sigma)-(\sigma+1)m_{ij}^{\sigma} v_{ij}^{\sigma}}$. The Pareto assumption guarantees that markup elasticities do not depend on the source. Following ACDR, we compute ρ as the expenditure weighted average of the markup elasticity. Let $i = j = US$ since we use moments from US data. The expenditure weight for each firm is its domestic sales relative to total domestic sales: $\frac{r_{ii}(v)}{\int_0^1 r_{ii}(v) d\mu_{ii}(v) dv}$. We then define the average markup elasticity:

$$\rho = \frac{\int_0^1 \left[\frac{\sigma}{(\sigma+1)+(\sigma-1)m^{-\sigma}v^{-\sigma}}\right] mv \left[m^{-\sigma}v^{-\sigma} - 1\right] v^{\theta-1} dv}{\int_0^1 mv \left[m^{-\sigma}v^{-\sigma} - 1\right] v^{\theta-1} dv} \tag{39}$$

Since the markup elasticity is not source-specific, we can calculate the average weighted elasticity using only domestic producers (taking the average markups of their *domestic* sales). In Appendix F, we apply a similar procedure to compute the corresponding elasticity in the

[28] We can set their β parameter, which represents the difference between the total price elasticity and cross-price elasticity, equal to 0 in the additively separable cases.

remaining models.

Table 4: Welfare Parameters and Gains from Trade

Model	θ	ρ	$(1-\frac{\rho}{1+\theta})$	ΔW_j in US (relative to autarky)	Average ΔW_j in all countries (relative to autarky)
Generalized CES	1.92	0.41	0.86	4.0%	16%
SIM (MP)	2.11	0.5	0.84	3.5%	14%
MO ($\eta=0$, MP)	2.44	0.65	0.81	2.9%	11.5%
BMMS (MP)	2.37	1	0.70	2.6%	10%

Welfare results are calculated using expression (38). For the average across all countries we drop 6 countries with predicted gains of over 100%: Azerbaijan, Central African Republic, Ethiopia, Nigeria, Senegal, and Zambia, so that they do not over-influence the result.

Table 4 displays the results for ρ and θ in the general and restricted models, and the implications for welfare gains when a country moves from autarky to its current domestic expenditure share. We examine the parameterization of the restricted models that relies on an estimate of the Pareto shape parameter from measured-productivity data. As is apparent from the welfare formula above, welfare gains are falling in θ, thus this parameterization of the models yields the largest possible welfare gains.

Notice that in the welfare formula only the domestic share is country-specific. We report the gains from trade for the United States and an average across countries in our sample. The generalized model has higher welfare gains for two reasons: i) the markup elasticity is lower, and ii) the Pareto shape parameter is lower. In a move from autarky to the current domestic expenditure share, it is easy to verify that welfare strictly decreases with the average markup elasticity and the Pareto shape parameter.[29]

Given the current thought experiment, the consumer welfare gains in the generalized CES model are 4.0% when the share of consumption that is imported rises from autarky to the predicted trade share in the United States. The SIM model predicts a 3.5% welfare change in the United States. This suggests that, for the US, the restricted model when $\sigma = 1$ would underestimate the welfare gains by 14%. The separable MO and BMMS models yield even lower welfare gains, with the difference from the generalized model's prediction amounting to 54% in the case of the BMMS framework.

We also report the average welfare gains for all the countries in our sample in the last column.

[29]The latter requires that $\rho(1+2\theta) < 1+2\theta+\theta^2$. With the parameter restriction of a positive θ and $\rho < 1$, this always holds.

The average welfare gains across the sample range from 16% to 10%—a 60% difference.[30] Therefore, the results suggest that the welfare gains from trade are potentially underestimated by at least 14-54% for the US and at least 60% for a set of 65 countries in the world.

6 Conclusion

In this paper, we quantify a class of commonly-employed general equilibrium models of international trade and pricing-to-market that feature firm-level heterogeneity and consumers with non-homothetic preferences. We evaluate the quantitative performance of the models along dimensions in the data that relate to firm-level productivity and sales dispersion as well as cross-country price dispersion, and we discuss the implications for welfare analysis. We view the work as offering insights to the literature regarding the trade-off between precisely quantifying the pro-competitive welfare gains from trade and exploiting highly-desirable tractability features of different models. Admittedly, we evaluate the performance of existing non-homothetic models along a limited number of dimensions. While we believe that, by focusing on moments that do not necessarily relate to prices, we put the models to a difficult test, we leave it for future work to examine the performance of these models along other dimensions as well as to develop viable alternatives that offer further insights into the sources of the welfare gains from trade.

References

Alessandria, G. and J. P. Kaboski (2011): "Pricing-to-Market and the Failure of Absolute PPP," *American Economic Journal: Macroeconomics*, 3, 91–127.

Arkolakis, C. (2010): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 118, 1151 – 1199.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodrigues-Clare (2015): "The Elusive Pro-Competitive Effects of Trade," Working paper.

Basu, S. and J. G. Fernald (1997): "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy.*

Behrens, K., G. Mion, Y. Murata, and J. Sudekum (2014): "Trade, Wages, and Productivity," *International Economic Review*, 55, 1305–1348.

[30] Our calibrated domestic expenditure share is 91.6% in the US, and the average in the whole sample (eliminating 6 outliers) is 77%. For the average across all countries we drop 6 countries with predicted gains of over 100%: Azerbaijan, Central African Republic, Ethiopia, Nigeria, Senegal, and Zambia, so that they do not over-influence the result.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," *The American Economic Review*, 93, 1268–1290.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2012): "The Empirics of Firm Heterogeneity and International Trade," *Annual Review of Economics*, 4, 283–313.

Bertoletti, P. and F. Etro (2014): "A General Theory of Endogenous Market Structures," DEM Working Papers Series 081, University of Pavia, Department of Economics and Management.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98 (4), 1701–121.

De Loecker, J. and F. Warzynski (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102, 2437–71.

Deaton, A. and A. Heston (2010): "Understanding PPPs and PPP-Based National Accounts," *American Economic Journal: Macroeconomics*, 2, 1–35.

deBlas, B. and K. N. Russ (2015): "Understanding Markups in the Open Economy," *American Economic Journal: Macroeconomics*, 7, 157–80.

Dhingra, S. and J. Morrow (2012): "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," *Working Paper*.

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

Eaton, J., S. Kortum, and F. Kramarz (2004): "Dissecting Trade: Firms, Industries, and Export Destinations," *The American Economic Review, Papers and Proceedings*, 94 (2), 150–154.

——— (2011): "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 79, 1453–1498.

Edmond, C., V. Midrigan, and D. Xu (2014): "Competition, Markups and the Gains from International Trade," *American Economic Review*, forthcoming.

Feenstra, R. C. (2014): "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," NBER Working Papers 19833, National Bureau of Economic Research, Inc.

Feenstra, R. C. and D. E. Weinstein (2010): "Globalization, Markups and U.S. Welfare," NBER Working Papers 15749, National Bureau of Economic Research, Inc.

Hall, R. E. (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96, 921–47.

Jaimovich, N. and M. Floetotto (2008): "Firm dynamics, Markup variations, and the Business Cycle," *Journal of Monetary Economics*, 55, 1238–1252.

Martins, J., S. Scarpetta, and D. Pilat (1996): "Mark-up Pricing, Market Structure, and the Business Cycle," *OECD Economic Studies*, 27, 72–105.

Melitz, M. and G. Ottaviano (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75 (1), 295–316.

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

Norrbin, S. C. (1993): "The Relation between Price and Marginal Cost in U.S. Industry: A Contradiction," *Journal of Political Economy*, 101, 1149–64.

Roeger, W. (1995): "Can Imperfect Competition Explain the Difference between Primal and Dual Productivity Measures? Estimates for U.S. Manufacturing," *Journal of Political Economy*, 103, pp. 316–330.

Simonovska, I. (2015): "Income Differences and Prices of Tradables: Insights From An Online Retailer," *Review of Economic Studies*, forthcoming.

Simonovska, I. and M. E. Waugh (2014a): "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 92, 34–50.

——— (2014b): "Trade Models, Trade Elasticities, and the Gains from Trade," NBER Working Papers 20495, National Bureau of Economic Research, Inc.

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

Weinberger, A. (2014): "Markups and Misallocation with Trade and Heterogeneous Firms," Working papers, UC Davis.

Appendix

A Generalized CES Model

A.1 Consumer and Firm Problems

Given the utility function and the budget constraint ($\int_{\omega\in\Omega} p(\omega)q^c(\omega)d\omega \leq w+\pi$), the first order conditions are:

$$\left(\int_{\omega\in\Omega}(q^c(\omega)+\bar{q})^{\frac{\sigma-1}{\sigma}}d\omega\right)^{\frac{\sigma}{\sigma-1}-1}(q^c(\omega)+\bar{q})^{-\frac{1}{\sigma}}=\lambda p(\omega),$$

where λ is the Lagrange multiplier. We use the FOCs and integrate over all ω to get the total demand, equation (4), in the main text. The FOC's of (5) yield:

$$(1-\sigma)p_{ij}^{-\sigma}L_j\frac{w_j+\bar{q}P_j}{P_{j\sigma}^{1-\sigma}}-L_j\bar{q}+\sigma c_{ij}p_{ij}^{-\sigma-1}L_j\frac{w_j+\bar{q}P_j}{P_{j\sigma}^{1-\sigma}}=0 \tag{40}$$

This is the expression that is used to obtain the implicit equation for prices (7) in the main text (substitute (6) into these FOC's).

The following provides details on the comparative statics of prices and markups. Define the following implicit function that attains zero at the optimal price:

$$F(p_{ij},c_{ij})=(1-\sigma)p_{ij}+\sigma c_{ij}-p_{ij}^{\sigma+1}(\bar{c}_j)^{-\sigma} \tag{41}$$

By the implicit function theorem, $dp_{ij}/dc_{ij}>0$. Hence, high cost firms charge higher prices. However, price rises by less than proportional with cost. To see this, define the mark-up as $m_{ij}=p_{ij}/c_{ij}$. Dividing the both sides of expression (41) by $\left(m_{ij}^{\sigma+1}c_{ij}\right)$ yields the implicit function for mark-ups and costs as:

$$G(m_{ij},c_{ij})=(1-\sigma)m_{ij}^{-\sigma}+\sigma m_{ij}^{-(\sigma+1)}-c_{ij}^{\sigma}(\bar{c}_j)^{-\sigma} \tag{42}$$

By the implicit function theorem,

$$\frac{dm_{ij}}{dc_{ij}}=\frac{c_{ij}^{\sigma-1}(\bar{c}_j)^{-\sigma}}{(\sigma-1)m_{ij}^{-\sigma-1}-(\sigma+1)m_{ij}^{-\sigma-2}} \tag{43}$$

with $dm_{ij}/dc_{ij}<0$ iff $m_{ij}<(\sigma+1)/(\sigma-1)$, which is larger than the Dixit Stiglitz mark-up. Notice that if $dm_{ij}/dc_{ij}<0$, then, as $c_{ij}\to 0$, m_{ij} must approach an upper bound. To find this upper bound, set $G(m_{ij}, 0) = 0$. Clearly, the solution to this system is $m_{ij} = \sigma/(\sigma - 1)$—the Dixit Stiglitz mark-up—which falls below the necessary bound. Hence, the necessary restriction always holds and mark-ups fall with costs and converge to the D-S mark-up as costs fall to zero. Hence the range for mark-ups is $[1, \sigma/(\sigma - 1)]$.

Finally, we do not report the equation for profits in the main text. It is given by:

$$\begin{aligned}\pi_{ij}(c_{ij}) &= (p_{ij}(c_{ij}) - c_{ij})\, x_{ij}(c_{ij}) \\ &= L_j \bar{q} \left((\bar{c}_j)^{\sigma} (p_{ij}(c_{ij}))^{1-\sigma} - p_{ij}(c_{ij}) - (\bar{c}_j)^{\sigma} c_{ij} (p_{ij}(c_{ij}))^{-\sigma} + c_{ij} \right)\end{aligned} \tag{44}$$

A.2 Equilibrium

We use the change of variable described in the text to derive the aggregate predictions. Here we give more details of the work done to arrive at equations (11)-(14). Given the change of variables, we rewrite profits and sales but normalize both by $\bar{c}_j$:

$$\begin{aligned}\tilde{T}_{ij} &= \frac{T_{ij}}{\bar{c}_j} = J_i \bar{q} L_j \int_0^{\bar{c}_j} [t_{ij}^{1-\sigma} - t_{ij}] d\mu_{ij}(c) \\ T_{ij} &= J_i \bar{c}_j \left(\frac{\bar{c}_j}{\hat{c}_{ij}} \right)^{\theta} \frac{L_j \bar{q} \theta}{\sigma^{\theta}} \int_0^1 \left[t_{ij}^{1-\sigma} - t_{ij} \right] \left[(\sigma+1) t_{ij}^{\sigma} + (\sigma - 1) \right] \left[t_{ij}^{\sigma+1} + (\sigma-1) t_{ij} \right]^{\theta-1} dt_{ij}\end{aligned}$$

Notice that we use $d\mu_{ij}(c) = \theta \frac{c_{ij}^{\theta-1}}{\hat{c}_{ij}^{\theta}} dc_{ij}$, where $c_{ij} = \frac{\bar{c}_j}{\sigma}(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij})$ and $dc_{ij} = \frac{\bar{c}_j}{\sigma}((\sigma + 1)t_{ij}^{\sigma} + (\sigma - 1))dt_{ij}$. Similarly, for profits:

$$\begin{aligned}\tilde{\pi}_{ij} &= \frac{\pi_{ij}}{\bar{c}_j} = \bar{q} L_j \int_0^{\bar{c}_j} \frac{1}{\sigma} [t_{ij}^{1-\sigma} + t_{ij}^{\sigma+1} - 2t_{ij}] d\mu_{ij}(c) \\ &= \sum_{j=1}^{I} \bar{c}_j \left(\frac{\bar{c}_j}{\hat{c}_{ij}} \right)^{\theta} \frac{\bar{q} L_j \theta}{\sigma^{\theta+1}} \int_0^1 \Big[\sigma t_{ij} \left(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij} \right)^{\theta-1} + t_{ij}^{-\sigma} \left(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij} \right)^{\theta} \\ &+ \sigma t_{ij}^{2\sigma+1} \left(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij} \right)^{\theta-1} + t_{ij}^{\sigma} \left(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij} \right)^{\theta} - 2\sigma t_{ij}^{\sigma+1} \left(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij} \right)^{\theta+1} \\ &- 2 \left(t_{ij}^{\sigma+1} + (\sigma-1)t_{ij} \right)^{\theta} \Big] dt_{ij}\end{aligned}$$

Since the integration range is between 0 and 1, the integrals reduce to constants. The following are the constants in the sales and profit equations used in the main text: $\beta_1 = -\frac{(\sigma-3)\,{}_2F_1\left(1-\theta, \frac{\theta+1}{\sigma}; \frac{\sigma+\theta+1}{\sigma}; \frac{1}{1-\sigma}\right)}{\theta+1} + \frac{(\sigma+1)\,{}_2F_1\left(1-\theta, \frac{2\sigma+\theta+1}{\sigma}; \frac{3\sigma+\theta+1}{\sigma}; \frac{1}{1-\sigma}\right)}{2\sigma+\theta+1} - \frac{(\sigma-1)\,{}_2F_1\left(1-\theta, \frac{-\sigma+\theta+1}{\sigma}; \frac{\theta+1}{\sigma}; \frac{1}{1-\sigma}\right)}{\sigma-\theta-1} - \frac{(\sigma+3)\,{}_2F_1\left(1-\theta, \frac{\sigma+\theta+1}{\sigma}; \frac{2\sigma+\theta+1}{\sigma}; \frac{1}{1-\sigma}\right)}{\sigma+\theta+1}$ and $\beta_2 = \frac{2\,{}_2F_1\left(1-\theta, \frac{\theta+1}{\sigma}; \frac{\sigma+\theta+1}{\sigma}; \frac{1}{1-\sigma}\right)}{\theta+1} - \frac{(\sigma-1)\,{}_2F_1\left(1-\theta, \frac{-\sigma+\theta+1}{\sigma}; \frac{\theta+1}{\sigma}; \frac{1}{1-\sigma}\right)}{\sigma-\theta-1} -$

$\frac{(\sigma+1)\,{}_2F_1\left(1-\theta,\frac{\sigma+\theta+1}{\sigma};\frac{2\sigma+\theta+1}{\sigma};\frac{1}{1-\sigma}\right)}{\sigma+\theta+1}$, where

$$\begin{aligned} F_{2,1}(\alpha,\beta;\delta;z) &= \frac{\Gamma(\delta)}{\Gamma(\beta)\Gamma(\delta-\beta)}\int_0^1 \frac{t^{\beta-1}(1-t)^{\delta-\beta-1}}{(1-tz)^{\alpha}}dt, \\ \Gamma(z) &= \int_0^1 \left[\log\left(\frac{1}{t}\right)\right]^{z-1} dt \end{aligned}$$

is the hypergeometric function with vector $(\alpha,\beta)=(1,\kappa+1)$, scalar $\delta=\kappa+2$ and element $z=-\gamma x$ evaluated at the two endpoints:[31]

The aggregate price statistics are:

$$\begin{aligned} P_j &= \sum_{\upsilon}^{I} J_{\upsilon}\bar{c}_j\left(\frac{\bar{c}_j}{\hat{c}_{\upsilon j}}\right)^{\theta}\frac{\theta}{\sigma^{\theta}}\int_0^1 \left[t_{\upsilon j}^{\sigma+1}+(\sigma-1)t_{\upsilon j}\right]^{\theta-1}\left[(\sigma+1)t_{\upsilon j}^{\sigma+1}+(\sigma-1)t_{\upsilon j}\right]dt_{\upsilon j} \\ &= \sum_{\upsilon}^{I} J_{\upsilon}\frac{\bar{c}_j^{\theta+1}b_{\upsilon}^{\theta}}{(\tau_{\upsilon j}w_{\upsilon})^{\theta}}\beta_P \qquad (45) \\ P_{j\sigma}^{1-\sigma} &= \sum_{\upsilon}^{I} J_{\upsilon}\frac{\bar{c}^{\theta+1-\sigma}}{\hat{c}_{\upsilon j}^{\theta}}\frac{\theta}{\sigma^{\theta}}\int_0^1 \left[t_{\upsilon j}^{\sigma+1}+(\sigma-1)t_{\upsilon j}\right]^{\theta-1}\left[(\sigma+1)t_{\upsilon j}+(\sigma-1)t_{\upsilon j}^{1-\sigma}\right]dt_{\upsilon j} \\ &= \sum_{\upsilon}^{I} J_{\upsilon}\frac{\bar{c}^{\theta+1-\sigma}b_{\upsilon}^{\theta}}{(\tau_{\upsilon j}w_{\upsilon})^{\theta}}\beta_{\sigma P} \qquad (46) \end{aligned}$$

where again we group constants in β_P and $\beta_{\sigma P}$.

The measure of entrants results from substituting profits and total sales into FE and IS. The expression for measure of entrants is:

$$J_i = \frac{\Gamma_1}{\Gamma_2}\frac{L_i}{f_e} \qquad (47)$$

where we group together constants. Specifically, these are: $\Gamma_1 = \frac{\theta}{\sigma^{\theta+1}}\left(\frac{1}{\sigma-1}\right)^{1-\theta}\beta_1$ and $\Gamma_2 = \frac{\theta}{\sigma^{\theta}}\left(\frac{1}{\sigma-1}\right)^{1-\theta}\beta_2$.

Lastly we get the expression for the cutoff costs (20) by substituting the aggregate price statistics into (6):

$$\bar{c}_j = \left[\frac{w_j}{(\beta_{\sigma P}-\beta_P)\bar{q}\sum_{\upsilon}^{I} J_{\upsilon}\frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j}w_{\upsilon})^{\theta}}}\right]^{\frac{1}{\theta+1}} \qquad (48)$$

[31] In Matlab, however, the Hypergeometric function, *hypergeom*(a,b,z), corresponds to the generalized Hypergeometric function where a is a vector of "upper parameters", b is vector of "lower parameters" and z is the argument. $F_{2,1}(\alpha,\beta;\delta;z)$ is the special case where $a=(\alpha,\beta)$ is a 1 by 2 matrix and $b=\delta$ is a scalar.

In the main text we substitute for J_v which results in equation (20).

A.3 Income Per-Capita and Prices

As stated in the main text we can show analytically that prices of identical goods are higher in richer destinations. We apply the implicit function theorem to (7) to verify that $dp_{ij}/d\bar{c}_j > 0$ and combine this with $d\bar{c}_j/dw_j$ which can be solved with the closed form solution for cutoff costs above:

$$\begin{aligned}
\frac{dp_{ij}}{dw_j} &= \frac{dp_{ij}}{d\bar{c}_j}\frac{d\bar{c}_j}{dw_j} \\
&= \underbrace{\left[\frac{\sigma p_{ij}^{\sigma+1}\left(\bar{c}_j\right)^{-\sigma-1}}{(\sigma+1)p_{ij}^{\sigma}\left(\bar{c}_j\right)^{-\sigma}-(1-\sigma)}\right]}_{>0}\left[\frac{d\bar{c}_j}{dw_j}\right]
\end{aligned}$$

We need to verify that the second term is positive:

$$\begin{aligned}
\frac{\partial \bar{c}_j}{\partial w_j} = \frac{1}{\theta+1}&\left[\frac{w_j}{(\beta_{\sigma P}-\beta_P)\beta_J\bar{q}f_e^{-1}\sum_{v=1}^{I}L_v\frac{b_v^\theta}{(\tau_{vj}w_v)^\theta}}\right]^{\frac{1}{\theta+1}-1}\times \\
&\left\{\frac{1}{(\beta_{\sigma P}-\beta_P)\beta_J\bar{q}f_e^{-1}\sum_{v=1}^{I}L_v\frac{b_v^\theta}{(\tau_{vj}w_v)^\theta}}-\frac{w_j(\beta_{\sigma P}-\beta_P)\beta_J\bar{q}f_e^{-1}L_j b_j^\theta(-\theta)w_j^{-\theta-1}\tau_{jj}^{-\theta}}{\left[(\beta_{\sigma P}-\beta_P)\beta_J\bar{q}f_e^{-1}\sum_{v=1}^{I}L_v\frac{b_v^\theta}{(\tau_{vj}w_v)^\theta}\right]^2}\right\} \\
&= \frac{1}{\theta+1}\left[\frac{w_j}{(\beta_{\sigma P}-\beta_P)\beta_J\bar{q}f_e^{-1}\sum_{v=1}^{I}L_v\frac{b_v^\theta}{(\tau_{vj}w_v)^\theta}}\right]^{\frac{1}{\theta+1}}\times\left\{\frac{1}{w_j}+\frac{\theta L_j b_j^\theta w_j^{-\theta-1}}{\sum_{v=1}^{I}L_v\frac{b_v^\theta}{(\tau_{vj}w_v)^\theta}}\right\} \\
&= \frac{1}{\theta+1}\bar{c}_j\frac{1+\theta\lambda_{jj}}{w_j} > 0
\end{aligned}$$

The end result is therefore that $\frac{dp_{ij}}{dw_j} > 0$.

A.4 Moments in the Generalized CES

In the main text we simplify the solutions. The full solution for the measured productivity of non-exporters requires $\beta_{MP,i}^{NX}(\sigma,\theta,\tilde{t}_{ii})$:

$$\beta_{MP,i}^{NX}(\sigma,\theta,\tilde{t}_{ii}) = \frac{1}{(\sigma-1)\theta(\sigma+\theta)}\times$$

$$\Bigg[\sigma^{\theta}\left(\frac{1}{\sigma-1}+1\right)^{-\theta}\Bigg\{(\sigma\theta\log\sigma+\theta\log\sigma+\sigma)\ {}_2F_1\left(1-\theta,\frac{\sigma+\theta}{\sigma};\frac{\theta}{\sigma}+2;\frac{1}{1-\sigma}\right)-(\sigma-1)(\sigma+\theta)\times$$

$$\left(-\log\sigma\,{}_2F_1\left(1-\theta,\frac{\theta}{\sigma};\frac{\sigma+\theta}{\sigma};\frac{1}{1-\sigma}\right)+\left(\frac{1}{\sigma-1}+1\right)^{\theta}\log\sigma\right)\Bigg\}$$

$$-\left(\tilde{t}_{ii}\left(\tilde{t}_{ii}^{\sigma}+\sigma-1\right)\right)^{\theta}\left(\frac{\tilde{t}_{ii}^{\sigma}}{\sigma-1}+1\right)^{-\theta}\Bigg\{\tilde{t}_{ii}^{\sigma}\Bigg(\sigma\theta\log\sigma+\theta\log\sigma+\sigma\Bigg)\,{}_2F_1\left(1-\theta,\frac{\sigma+\theta}{\sigma};\frac{\theta}{\sigma}+2;-\frac{\tilde{t}_{ii}^{\sigma}}{\sigma-1}\right)$$

$$-(\sigma-1)(\sigma+\theta)\Bigg((-\log\sigma)\ {}_2F_1\left(1-\theta,\frac{\theta}{\sigma};\frac{\sigma+\theta}{\sigma};-\frac{\tilde{t}_{ii}^{\sigma}}{\sigma-1}\right)+\left(\frac{\tilde{t}_{ii}^{\sigma}}{\sigma-1}+1\right)^{\theta}\log\left(\tilde{t}_{ii}^{\sigma}+\sigma-1\right)\Bigg)\Bigg\}\Bigg]$$

where ${}_2F_1$ is Gauss's hypergeometric function.

The sales moment is also not written out completely in the main text. The closed form solution is:

$$M_{i,sales}^{m} = (1-\xi_i^{\theta})\Bigg[\tilde{t}_{ii}^{2}(\tilde{t}_{ii}(\tilde{t}_{ii}^{\sigma}+\sigma-1))^{\theta-1}\left(\frac{\tilde{t}_{ii}^{\sigma}}{\sigma-1}+1\right)^{1-\theta}\Bigg\{\frac{2}{\theta+1}{}_2F_1\left(1-\theta,\frac{\theta+1}{\sigma};\frac{\sigma+\theta+1}{\sigma};\frac{\tilde{t}_{ii}^{\sigma}}{1-\sigma}\right)+$$

$$\frac{1}{(\sigma-\theta-1)(\sigma+\theta+1)}\tilde{t}_{ii}^{-\sigma}\Bigg((1-\sigma)(\sigma+\theta+1){}_2F_1\left(1-\theta,\frac{1-\sigma+\theta}{\sigma};\frac{\theta+1}{\sigma};\frac{\tilde{t}_{ii}^{\sigma}}{1-\sigma}\right)$$

$$-(\sigma+1)(\sigma-\theta-1)\tilde{t}_{ii}^{2\sigma}\,{}_2F_1\left(1-\theta,\frac{1+\sigma+\theta}{\sigma};\frac{2\sigma+\theta+1}{\sigma};\frac{\tilde{t}_{ii}^{\sigma}}{1-\sigma}\right)\Bigg)\Bigg\}\Bigg]\div$$

$$(\xi_i^{\theta})\Bigg[\sigma^{\theta-1}\left(\frac{1}{\sigma-1}+1\right)^{1-\theta}\times\Bigg\{\frac{2}{\theta+1}{}_2F_1\left(1-\theta,\frac{\theta+1}{\sigma};\frac{\sigma+\theta+1}{\sigma};\frac{1}{1-\sigma}\right)$$

$$+\frac{1}{(\sigma-\theta-1)(\sigma+\theta+1)}\Bigg((1-\sigma)(\sigma+\theta+1){}_2F_1\left(1-\theta,\frac{1-\sigma+\theta}{\sigma};\frac{\theta+1}{\sigma};\frac{1}{1-\sigma}\right)$$

$$-(\sigma+1)(\sigma-\theta-1){}_2F_1\left(1-\theta,\frac{1+\sigma+\theta}{\sigma};\frac{2\sigma+\theta+1}{\sigma};\frac{1}{1-\sigma}\right)\Bigg)\Bigg\}-$$

$$\tilde{t}_{ii}^{2}(\tilde{t}_{ii}(\tilde{t}_{ii}^{\sigma}+\sigma-1))^{\theta-1}\left(\frac{\tilde{t}_{ii}^{\sigma}}{\sigma-1}+1\right)^{1-\theta}\Bigg\{\frac{2}{\theta+1}{}_2F_1\left(1-\theta,\frac{\theta+1}{\sigma};\frac{\sigma+\theta+1}{\sigma};\frac{\tilde{t}_{ii}^{\sigma}}{1-\sigma}\right)+$$

$$\frac{1}{(\sigma-\theta-1)(\sigma+\theta+1)}\tilde{t}_{ii}^{-\sigma}\Bigg((1-\sigma)(\sigma+\theta+1){}_2F_1\left(1-\theta,\frac{1-\sigma+\theta}{\sigma};\frac{\theta+1}{\sigma};\frac{\tilde{t}_{ii}^{\sigma}}{1-\sigma}\right)$$

$$-(\sigma+1)(\sigma-\theta-1)\tilde{t}_{ii}^{2\sigma}\,{}_2F_1\left(1-\theta,\frac{1+\sigma+\theta}{\sigma};\frac{2\sigma+\theta+1}{\sigma};\frac{\tilde{t}_{ii}^{\sigma}}{1-\sigma}\right)\Bigg)\Bigg\}\Bigg]$$

B Moments in the SIM Model

The moments below can be used to calibrate the value of θ. We compute two firm-level moments as in BEJK. The first firm level moment is the productivity advantage of exporters over non-exporters. To construct it, we compute value added for each s as in BEJK, where there are no intermediate inputs:

$$va_i(s) = \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s)\bar{c}_\upsilon L_\upsilon \bar{q}\left(1 - t_{i\upsilon}(s)\right),$$

where $t_{i\upsilon}(s) = \frac{p_{i\upsilon}(s)}{\bar{c}_\upsilon}$ and $\delta_{i\upsilon}(s) = 1$ if firm s from i sells to υ. Employment of the same s is

$$emp_i(s) = \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s)\bar{c}_\upsilon L_\upsilon \bar{q}\frac{t_{i\upsilon}(s) - t_{i\upsilon}(s)^2}{w_i}$$

Measured productivity of s is the ratio of the two objects

$$mp_i(s) = \log\left(\frac{va_i(s)}{emp_i(s)}\right)$$

For exporters the average measured productivity (in logs) is:

$$MP_i^{EXP} = \frac{\theta}{(\xi_i)^\theta}\int_0^{\tilde{t}_{ii}}\left[\log\left\{\sum_{\upsilon}^{I}\delta_{i\upsilon}(s)\chi_{i\upsilon}L_\upsilon(1 - h_{i\upsilon}(t_{ii}(s)))\right\}\right.$$
$$\left. - \log\left\{\sum_{\upsilon}^{I}\delta_{i\upsilon}(s)\chi_{i\upsilon}L_\upsilon\left(h_{i\upsilon}(t_{ii}(s)) - h_{i\upsilon}(t_{ii}(s))^2\right)\right\}\right]2t_{ii}(s)^{2\theta-1}dt_{ii}(s) + \frac{1}{\xi_i^\theta}\tilde{t}_{ii}^{2\theta}\log(w_i) \qquad (49)$$

As in the general model: $\xi_i = \frac{\tilde{c}_{ii}^x}{\bar{c}_i}$, $\chi_{i\upsilon} = \frac{\bar{c}_\upsilon}{\bar{c}_i}$, $h_{i\upsilon}(t_{ii}(s))$ is the solution of $t_{i\upsilon}(s)$ corresponding $t_{ii}(s)$, and $\tilde{t}_{ii}$ is the t_{ii} of the marginal exporter in i. The average measured productivity of non-exporters (in logs) is:

$$MP_i^{NX} = \log(w_i) + \frac{1}{2\theta} + \frac{\xi_i^\theta}{1 - \xi_i^\theta}\log(\tilde{t}_{ii}) \qquad (50)$$

We compute the moment as the difference between the average logged measured productivity of exporters and non-exporters: $M_{i,prod}^{SIM} = MP_i^{EXP} - MP_i^{NX}$, for $i = US$.

The second firm level moment is the domestic sales advantage of exporters. BEJK compute this statistic as the ratio between the average domestic sales among exporters and non-exporters.

In the model, for $i = US$, the statistic is

$$M_{i,sales}^{SIM} = \left(\frac{1-\xi_i^\theta}{\xi_i^\theta}\right) \frac{(\tilde{t}_{ii})^{2\theta}/\theta - 2(\tilde{t}_{ii}^{2\theta+1}/(2\theta+1))}{(1/\theta - 2/(2\theta+1) - ((\tilde{t}_{ii})^{2\theta}/\theta - 2(\tilde{t}_{ii})^{2\theta+1}/(2\theta+1)))} \tag{51}$$

In the SIM model, there is only one parameter, θ, that needs to be consistent with both the measured productivity and sales advantage moments. In Section 5 we showed that it is not possible with any value of θ to match both moments in this model.

C Moments in the Separable MO Model ($\eta = 0$)

Define a new variable $v_{ij}(s) \equiv \frac{c_{ij}(s)}{\bar{c}_j}$. Then it must be that:

$$c_{ij}(s) = v_{ij}(s)\bar{c}_j \Longrightarrow dc_{ij}(s) = \bar{c}_j dv_{ij}(s) \tag{52}$$

Since we will be integrating over firms indexed by v_{ii} we must compute the range of v over which firms export and do not export. Let the marginal exporting firm be v_{ii}^x.

We compute the same moments as in the general model, so we start with the relative advantage of exporters in terms of average logged measured productivity. As in the general model, let $\xi_i \equiv \frac{\tilde{c}_{ii}^x}{\bar{c}_i}$, where $\tilde{c}_{ii}^x$ corresponds to v_{ii}^x.

Productivity in BEJK is defined as the labor productivity, or total value added divided by employment. The value added for each s is:

$$va_i(s) = \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \frac{L_\upsilon \bar{c}_\upsilon}{4\gamma}(1 - v_{i\upsilon}(s)^2)$$

where $\delta_{i\upsilon}(s) = 1$ if firm s from i sells to υ. There are no intermediate goods in this model, so only labor is used for production. Therefore, the value added is the same as the revenue for each s.

Employment of the same s is

$$emp_i(s) = \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \frac{L_\upsilon \bar{c}_\upsilon}{(2\gamma)w_i} v_{i\upsilon}(s)(1 - v_{i\upsilon}(s))$$

Measured productivity of s is the ratio of the two objects

$$mp_i(s) = \log\left(\frac{va_i(s)}{emp_i(s)}\right)$$

For exporters the average logged measured productivity is:

$$MP_i^{EXP} = \frac{\theta}{(\xi_i)^\theta} \int_0^{v_{ii}^x} \left[\log \left\{ \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \frac{1}{2} \frac{(\theta+2) w_\upsilon}{\sum_{j=1}^{I} L_j \left(w_j \tau_{j\upsilon}\right)^{-\theta} \bar{c}_\upsilon^\theta} \left(1 - (\tau_{i\upsilon} v_{ii}(s))^2\right) \right\} \right.$$
$$\left. - \log \left\{ \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \frac{(\theta+2) w_\upsilon}{\sum_{j=1}^{I} L_j \left(w_j \tau_{j\upsilon}\right)^{-\theta} \bar{c}_\upsilon^\theta} \tau_{i\upsilon} v_{ii}(s)(1 - \tau_{i\upsilon} v_{ii}(s))/w_i \right\} \right] v_{ii}(s)^{\theta-1} dv_{ii}(s) \tag{53}$$

The average logged measured productivity of non-exporters is:

$$MP_i^{NX} = \frac{\theta}{1-\xi_i^\theta} \int_{v_{ii}^x}^{1} \left[\log\left(\frac{1}{2}\right) + \log\left((1 - v_{ii}(s)^2)\right) - \log\left(v_{ii}(s)(1 - v_{ii}(s))\right) + \log(w_i) \right]$$
$$\left(v_{ii}(s)^{\theta-1}\right) dv_{ii}(s) \tag{54}$$

Again we compute the moment as the difference between the average logged measured productivity of exporters and non-exporters: $M_{i,prod}^{MO,\eta=0} = MP_i^{EXP} - MP_i^{NX}$, for $i = US$.

The US domestic sales advantage of exporters is the ratio between the average domestic sales among exporters and non-exporters. In the model, for $i = US$, the statistic is

$$M_{i,sales}^{MO,\eta=0} = \left(\frac{1-\xi_i^\theta}{\xi_i^\theta}\right) \frac{(v_{ii}^x)^\theta \left(\frac{1}{\theta} - \frac{(v_{ii}^x)^2}{\theta+2}\right)}{\frac{2}{\theta(\theta+2)} - \left((v_{ii}^x)^\theta \left(\frac{1}{\theta} - \frac{(v_{ii}^x)^2}{\theta+2}\right)\right)}. \tag{55}$$

In this model, there is only one parameter, θ, that needs to be consistent with both the measured productivity and sales advantage moments. In Section 5 we showed that it is not possible with any value of θ to match both moments in this model.

D Moments in the BMMS Model

To compute the firm-level moments for the BMMS model, recall the same change of variables trick as in the main text:

$$z \equiv W\left(\frac{c_{ii}}{\bar{c}_i} e\right) \Rightarrow \frac{c_{ii}}{\bar{c}_i} e = z e^z.$$

Then $c_{ij} = \bar{c}_j z_{ij} e^{z_{ij}-1}$ and $dc_{ij} = \left(e^{z_{ij}-1} \bar{c}_j\right)(1 + z_{ij}) dz_{ij}$ with integration bounds between 0 and 1.

As in the general model, we compute the relative advantage of exporters in terms of average logged measured productivity and domestic sales.

Productivity in BEJK is defined as the labor productivity, or total value added divided by employment. The value added for each s is:

$$va_i(s) = \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \frac{L_\upsilon}{\alpha} \bar{c}_\upsilon (z_{i\upsilon}(s) e^{z_{i\upsilon}(s)-1}) \left[z_{i\upsilon}(s)^{-1} - 1 \right]$$

where $\delta_{i\upsilon}(s) = 1$ if firm s from i sells to υ. There are no intermediate goods in this model, so only labor is used for production. Therefore, the value added is the same as the revenue for each s.

Employment of the same s is

$$emp_i(s) = \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) \frac{L_\upsilon \bar{c}_\upsilon}{\alpha w_i} \left[1 - z_{i\upsilon}(s) \right] z_{i\upsilon}(s) e^{z_{i\upsilon}(s)-1}$$

Measured productivity of s is the ratio of the two objects:

$$mp_i(s) = \log \left(\frac{va_i(s)}{emp_i(s)} \right)$$

We map $z_{i\upsilon}(s)$ into $z_{ii}(s)$ so as to be able to integrate over $z_{ii}(s)$. $\tilde{z}_{ii} = f(\frac{\tilde{c}^x_{ii}}{\bar{c}_i})$ is once again used for the marginal exporter, where $W(\frac{c_{i\upsilon}(s)}{\bar{c}_\upsilon} e)$ is solved with costs equal to $\frac{\tilde{c}^x_{ii}}{\bar{c}_i}$. We rely on a change of variables to compute the average logged productivity of exporters. We integrate over $z_{ii}(s)$ by using $\tilde{h}_{i\upsilon}(z_{ii}(s)) = z_{i\upsilon}(s)$ as the implicit solution to $z_{i\upsilon}(s) = (z_{ii}(s)) \frac{\tau_{i\upsilon} \bar{c}_i}{\bar{c}_\upsilon}$ and solve for $z_{i\upsilon}(s)$.

For exporters the average logged measured productivity is:

$$\begin{aligned} MP_i^{EXP} = \frac{\theta}{(\xi_i)^\theta} \int_0^{\tilde{z}_{ii}} & \left[\log \left\{ \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \bar{c}_\upsilon \left(\tilde{h}_{i\upsilon}(z_{ii}(s))^{-1} - 1 \right) \left(\tilde{h}_{i\upsilon}(z_{ii}(s)) e^{\tilde{h}_{i\upsilon}(z_{ii}(s))-1} \right) \right\} \right. \\ & \left. - \log \left\{ \frac{1}{w_i} \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \bar{c}_\upsilon \left(1 - \tilde{h}_{i\upsilon}(z_{ii}(s)) \right) \left(\tilde{h}_{i\upsilon}(z_{ii}(s)) e^{\tilde{h}_{i\upsilon}(z_{ii}(s))-1} \right) \right\} \right] \\ & \left(z_{ii}(s) e^{z_{ii}(s)-1} \right)^{\theta-1} e^{z_{ii}(s)-1} (1 + z_{ii}(s)) dz_{ii}(s) \end{aligned} \tag{56}$$

where $\xi_i \equiv \frac{\tilde{c}^x_{ii}}{\bar{c}_i}$ as in the main text.

The average logged measured productivity of non-exporters is:

$$MP_i^{NX} = \frac{\theta}{1 - \xi_i^\theta} \int_{\tilde{z}_{ii}}^{1} \left[\log(w_i) - \log(z_{ii}(s)) \right] \left(z_{ii}(s) e^{z_{ii}(s)-1} \right)^{\theta-1} e^{z_{ii}(s)-1} (1 + z_{ii}(s)) dz_{ii}(s) \tag{57}$$

We then compute the moment as the difference between the average logged measured productivity of exporters and non-exporters: $M^{BMMS}_{i,prod} = MP^{EXP}_i - MP^{NX}_i$, for $i = US$.

The domestic sales advantage of exporters in the model is, for $i = US$:

$$M^{BMMS}_{i,sales} = \frac{1-\xi_i^\theta}{\xi_i^\theta} \frac{\int_0^{\tilde{z}_{ii}} (z_{ii}(s)^{-1} - z_{ii}(s))(z_{ii}(s)e^{z_{ii}(s)})^\theta e^{z_{ii}(s)} dz_{ii}(s)}{\int_{\tilde{z}_{ii}}^1 (z_{ii}(s)^{-1} - z_{ii}(s))(z_{ii}(s)e^{z_{ii}(s)})^\theta e^{z_{ii}(s)} dz_{ii}(s)} \tag{58}$$

In this model, there is only one parameter, θ, that needs to be consistent with both the measured productivity and sales advantage moments. In Section 5 we showed that it is not possible with any value of θ to match both moments in this model.

E Melitz and Ottaviano Model in General Equilibrium

E.1 Model

The framework comes from Melitz and Ottaviano (2008), with an extension to general equilibrium outlined in the Web Appendix of Simonovska (2015). Assume each country is populated by identical consumers of measure L, whose utility function is:

$$U_j^c = \sum_{i=1}^{I} \int_{\omega \in \Omega_{ij}} q_{ij}^c(\omega) d\omega - \frac{1}{2}\gamma \int_{\omega \in \Omega_{ij}} \left(q_{ij}^c(\omega)\right)^2 d\omega - \frac{1}{2}\eta \left(\int_{\omega \in \Omega_{ij}} q_{ij}^c(\omega) d\omega\right)^2,$$

with $\gamma > 0$ and $\eta > 0$. This yields the following price, markup and sales:

$$p_{ij}(c_{ij}) = \frac{1}{2}(c_{ij} + \bar{c}_j) \tag{59}$$

$$m_{ij}(c_{ij}) = \frac{1}{2}\left(\frac{c_{ij} + \bar{c}_j}{c_{ij}}\right) \tag{60}$$

$$r_{ij}(c_{ij}) = \frac{L_j(\theta+1)}{2\bar{c}_{ij}\left(2\gamma(\theta+1) + \eta N_j\right)} \left(\bar{c}_{ij}^2 - c_{ij}^2\right) \tag{61}$$

where N_j is the measure of consumed varieties in country j, $\sum_{i=1}^{I} N_{ij}$ (see 2). We refer to the Web Appendix of Simonovska (2015) for a result that the link between per-capita income and prices/markups is positive.

We define a new variable $v_{ij} \equiv \frac{c_{ij}}{\bar{c}_j}$. Then it follows that:

$$c_{ij} = v_{ij}\bar{c}_j \Longrightarrow dc_{ij} = \bar{c}_j dv_{ij} \tag{62}$$

To solve the model we need total sales and the cost cutoffs. We start with total sales:

$$T_{ij} = \frac{J_i L_j \bar{c}_j^{\theta+1}(\theta+1)}{\hat{c}_{ij}^{\theta}\left(2\gamma(\theta+1)+\eta N_j\right)(\theta+2)} \tag{63}$$

Substituting total sales into the income-spending equation: $w_i L_i = \sum_j T_{ij}$ and using balanced trade, $\sum_j T_{ij} = \sum_j T_{ji}$, yields:

$$\frac{\theta+1}{\theta+2}\tilde{c}_j^{\theta+1} - \tilde{c}_j^{\theta} w_j = \frac{w_j 2\gamma(\theta+1)\eta^{-\theta-1}}{\sum_v^I J_v b_v^{\theta}\left(w_v \tau_{vj}\right)^{-\theta}}$$

where we let $\tilde{c}_j = \frac{\bar{c}_j}{\eta}$. Then, we define cutoffs for all j relative to a numeraire country, k:

$$\frac{\frac{\theta+1}{\theta+2}\tilde{c}_j^{\theta+1} - w_j \tilde{c}_j^{\theta}}{\frac{\theta+1}{\theta+2}\tilde{c}_k(K)^{\theta+1} - w_k \tilde{c}_k(K)^{\theta}} = \frac{w_j}{w_k}\frac{\sum_v^I L_v b_v^{\theta}\left(w_v \tau_{vk}\right)^{-\theta}}{\sum_v^I L_v b_v^{\theta}\left(w_v \tau_{vj}\right)^{-\theta}}, \tag{64}$$

where $K = 2\gamma(\theta+1)\eta^{-\theta-1}$. In the normalization of the cost cutoffs we solve for $\tilde{c}_k$ given a value of K. Given $\tilde{c}_k(K)$, we find relative cutoffs. Notice that we must also calibrate K since relative cutoffs change for different values of $\tilde{c}_k(K)$.

Since we will be integrating over firms indexed by v_{ii} we must compute the range of v over which firms export and do not export (with, $v_{ij}(K) \equiv \frac{c_{ij}}{\bar{c}_j(K)}$). Let the marginal exporting firm be $v_{ii}^x(K)$.

We compute the same moments as in the general model, so we start with the relative advantage of exporters in terms of average logged measured productivity. As in the general model, let $\xi_i(K) \equiv \frac{\bar{c}_{ii}^x(K)}{\bar{c}_i(K)}$.

Productivity in BEJK is defined as the labor productivity, or total value added divided by employment. The value added for each s is:

$$va_i(s) = \sum_{v=1}^{I} \delta_{iv}(s)\frac{L_v(\theta+1)\bar{c}_v(K)}{2(2\gamma(\theta+1)+\eta N_v)}(1 - v_{iv}(s)^2)$$

where $\delta_{iv}(s) = 1$ if firm s from i sells to v. There are no intermediate goods in this model, so only labor is used for production. Therefore, the value added is the same as the revenue for each s.

Employment of the same s is

$$emp(s) = \sum_{v=1}^{I} \delta_{iv}(s)\frac{L_v \bar{c}_v(K)(1-\eta Q_v^c)}{(2\gamma)w_i} v_{iv}(s)(1 - v_{iv}(s))$$

where Q_j^c is the aggregate consumption over varieties defined as $Q_j = \sum_{i=1}^{I} \int_{\omega \in \Omega_{ij}} q_{ij}^c(\omega) d\omega = \frac{N_j}{2\gamma(\theta+1)+\eta N_j}$.

Measured productivity of s is the ratio of the two objects

$$mp_(s) = \log\left(\frac{va_i(s)}{emp_i(s)}\right)$$

For exporters the average logged measured productivity is:

$$MP_i^{EXP} = \frac{\theta}{(\xi_i(K))^{\theta}} \int_0^{v_{ii}^x(K)} \left[\log\left\{ \sum_{v=1}^{I} \delta_{iv}(s) L_v \frac{1}{2} \frac{w_v}{\sum_{j=1}^{I} L_j \left(w_j \tau_{jv}\right)^{-\theta} \bar{c}_v(K)^{\theta}} (1 - (\tau_{iv} v_{ii}(s))^2) \right\} \right.$$
$$\left. - \log\left\{ \sum_{v=1}^{I} \delta_{iv}(s) L_v \frac{w_v}{\sum_{j=1}^{I} L_j \left(w_j \tau_{jv}\right)^{-\theta} \bar{c}_v(K)^{\theta}} \tau_{iv} v_{ii}(s) \frac{(1 - \tau_{iv} v_{ii}(s))}{w_i} \right\} \right] v_{ii}(s)^{\theta-1} dv_{ii}(s) \quad (65)$$

The average logged measured productivity of non-exporters is:

$$MP_i^{NX} = \frac{\theta}{1 - (\xi_i(K))^{\theta}} \int_{v_{ii}^x(K)}^{1} \left[\log\left(\frac{1}{2}\right) + \log\left((1 - v_{ii}(s)^2)\right) - \log\left(v_{ii}(s)(1 - v_{ii}(s))\right) + \log(w_i) \right]$$
$$\left(v_{ii}(s)^{\theta-1}\right) dv_{ii}(s) \quad (66)$$

We compute the moment as the difference between the average logged measured productivity of exporters and non-exporters: $M_{i,prod}^{MO} = MP_i^{EXP} - MP_i^{NX}$, for $i = US$.

The second firm level moment is the domestic sales advantage of exporters, the ratio between the average domestic sales among exporters and non-exporters. In the model, the statistic is

$$M_{i,sales}^{MO} = \left(\frac{1 - (\xi_i(K))^{\theta}}{(\xi_i(K))^{\theta}}\right) \frac{(v_{ii}^x(K))^{\theta} \left(\frac{1}{\theta} - \frac{(v_{ii}^x(K))^2}{\theta+2}\right)}{\frac{2}{\theta(\theta+2)} - \left((v_{ii}^x(K))^{\theta} \left(\frac{1}{\theta} - \frac{(v_{ii}^x(K))^2}{\theta+2}\right)\right)} \quad (67)$$

In this non-separable case ($\eta > 0$), the relative cost cutoffs are not fixed by wages and gravity variables, but shift according to the parameter values in K above: namely, they depend on a combination of parameters θ, γ, η. Therefore we can calibrate θ along with K to match the two moments.

E.2 Quantitative Analysis

The above discussion suggests that this model, like the generalized CES one, can potentially match both the sales and the measured productivity advantage of exporters. It does so via a very different channel: the (normalized) sales distribution is still fixed by θ in this model, but the value of K allows for a flexible cost cutoff, which is not the case in the separable models. In this model, the relative cost cutoffs are not fixed by wages and gravity variables, but shift according to the parameter values of K.

Below, we compare the quantitative predictions of this model to the generalized CES. The calibration results are in Table 5. Again wages are computed as the implicit solution to equation (18) using trade shares predicted from gravity. We calibrate θ and K to match the two moments derived above.

Table 5: Moments and Parameters

Model	Data/Targets	σ	θ	K	Simulated Moment
Generalized CES	$M_{sales} = 4.80$, $M_{prod} = 0.33$	1.41	1.92	–	$M_{sales} = 4.80$, $M_{prod} = 0.33$
Melitz and Ottaviano (2008)	$M_{sales} = 4.80$, $M_{prod} = 0.33$	–	2.01	1.21	$M_{sales} = 4.80$, $M_{prod} = 0.33$

While the model matches the two moments, its out of sample predictions stray widely from the data and from the predictions of the calibrated non-separable MO model. First, exporters are now the majority of firms. Table 6 reports that 74% of firms export in the MO model compared to 41% in the generalized CES (and their calibrated θ are similar). Second, the non-separable MO model displays counterfactual predictions of export intensity. Table 7 compares the results for both models. In the MO model, only a minority of firms are in the 0-10% decile of export intensity.

Table 6: US Exporters, % of Total

Model	Data	σ	θ	% of Firms that Export
Generalized CES	18-21%	1.41	1.92	41.4%
Melitz and Ottaviano (2008)	18-21%	–	2.01	73.6%

Further, the standard deviation of log domestic sales is 1.0 in the model, compared to a value of 1.26 in the generalized CES model and a value of 1.67 in the 1992 US Census data as reported by BEJK.

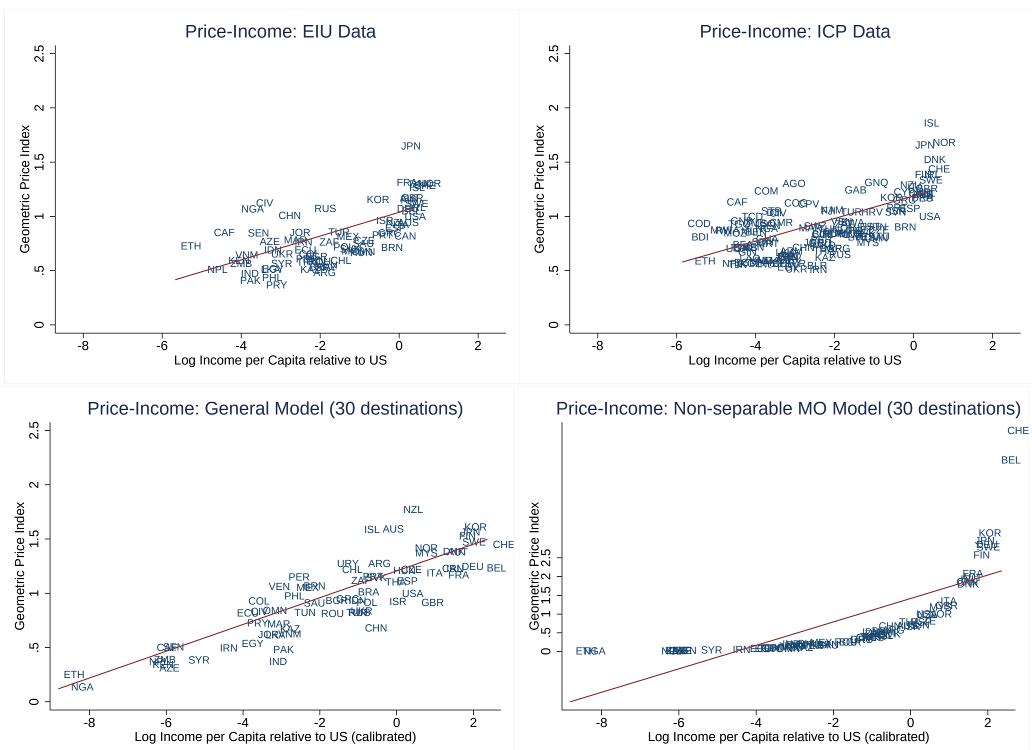
Price discrimination across countries in the model is quite high and amounts to three times more than in the data. The model yields a slope coefficient of 0.35 (standard deviation is 0.04), which is high compared to the generalized CES case (0.12 with standard deviation of 0.009) and to the data (0.11 for EIU data and 0.10 for ICP data). The fourth plot in figure 2 shows the predicted price indices in the MO model. As it can be seen, the model yields a very skewed

Table 7: % of Exporting Plants Conditional on Export Intensity

Exp. Intensity (%)	Data (%)	General CES	Melitz and Ottaviano (2008)
0-10	66	88.3	39.8
10-20	16	11.4	51.4
20-30	7.7	0.1	2.6
30-40	4.4	0.1	2.2
40-50	2.4	0.1	2.1
50-60	1.5	0	0.5
60-100	2.8	0	1.4

Conditioning on only exporting firms, the export intensity (first column) is the export revenue over total revenue, and we split firms into deciles. For example the first row shows that in the generalized CES model 88.3% of exporting firms have export revenue that is less than 10% of their total revenue. The data is also reported in Bernard et al. (2003) (using 1992 data for the US).

Figure 2: Geometric Mean of Relative Prices versus Income per Capita



Relative prices in the data are taken from EIU. In the model, we simulate firms in each calibrated model and capture firm prices to each destination. In model and data we define relative prices as the price of a good relative to the price in the US, and then take a geometric average of 110 relative prices for each country. For the simulated model, common goods are such that they are sold in at least 30 destinations. There are more than 110 common goods available in the sample so to reduce sampling error we have 100 draws of 110 random goods. For the simulation we can report the average price-income slope after having computed it for 100 draws.

distribution of prices across countries. The implication is an average markup on domestic sales of 50%, which exceeds values reported in the data.

F Welfare

In this section, we derive the welfare parameters for the remaining separable models. We recount equation (38):

$$\Delta W_j = \left(\frac{1}{\lambda_{jj}}\right)^{\left(1-\frac{\rho}{1+\theta}\right)/\theta} - 1$$

where ρ stands for the weighted average markup elasticity with respect to marginal cost (or one minus the price-cost pass through elasticity), and θ is the parameter that governs the dispersion of productivity assuming the distribution is unbounded Pareto.

ρ and θ are different across models, while trade shares are identical and come from data. Therefore, the welfare cost of autarky differs in the calibrated models due to the differences in market power.

In the SIM model, there is a closed form solution for prices: $p(c_{ij}, \bar{c}_j) = (c_{ij}\bar{c}_j)^{1/2}$. Therefore we can write the log markup as: $\log(m_{ij}) = \frac{1}{2}\log(\bar{c}_j) - \frac{1}{2}\log(c_{ij})$ and the markup elasticity becomes:

$$\rho^{SIM} = \frac{\partial \log(m_{ij})}{\partial \log(v_{ij})} = \frac{1}{2}.$$

Even without taking into consideration the firm distribution and source country, the markup elasticity is just one half.

In the separable MO model, the markup is $m_{ij} = \frac{1}{2}\left(1 + v_{ij}^{-1}\right)$ where $v_{ij} = c_{ij}/\bar{c}_j$. Then after we solve for the markup elasticity we again take the expenditure weighted average:

$$\rho^{MO,\eta=0} = \frac{\int_0^1 \frac{1}{v_{ii}(s)+1} r_{ii}(s) d\mu_{ii}(s)}{\int_0^1 r_{ii}(s) d\mu_{ii}(s)} = \frac{\theta+2}{2(\theta+1)}$$

Notice that in this model there is a *lower bound* on the markup elasticity of one-half, which corresponds to the upper bound in the generalized CES and to the constant markup elasticity in the SIM model. As θ goes to zero, the markup elasticity approaches one (percentage change in markups equals the percentage change in costs).

In the BMMS model, the domestic markup is $m_{ii} = \frac{1}{W\left(\frac{c_{ii}}{\bar{c}_i} e\right)}$ and the markup elasticity $\rho^{BMMS} = 1$, as shown in Behrens et al. (2014).[32] The authors show that the welfare gains from trade are summarized by: $d\ln W_j = -\left(\frac{1}{1+\theta}\right) d\ln\lambda_{jj}$.

[32]See p. 1345 in Behrens et al. (2014).