Exporter Heterogeneity and Price Discrimination: A Quantitative View*

Jae Wook Jung†
KIEP

Ina Simonovska‡
University of California, Davis and NBER

Ariel Weinberger§
University of Oklahoma

August 2018

Abstract

We quantify a general equilibrium model of international trade and pricing-to-market that features firm-level heterogeneity and consumers with non-homothetic preferences—generalized CES (GCES). We demonstrate theoretically that, relative to existing frameworks, the GCES model exhibits features of the data that are essential to conduct quantitative analysis. The framework can reconcile the documented price dispersion across firms and markets, while maintaining consistency with cross-sectional observations on firm productivity, markups, and sales. We estimate the model's parameters to match bilateral trade flows across 66 countries as well as moments from the markup and sales distributions of Chilean firms. The model reconciles both micro and macro facts quantitatively, and yields trade elasticity estimates that are in line with the existing literature. Hence, we conclude that the GCES model constitutes a plausible and parsimonious quantitative workhorse framework that can be used to analyze gains from trade.

JEL Classification: F12, F14, F17, F61

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

---

*We are indebted to Andres Rodriguez-Clare and two anonymous referees for extremely insightful comments that were critical to the analysis in this paper. We have benefited from discussions with Svetlana Demidova, Hylke Vandenbussche, and Olga Timoshenko. We thank seminar participants at the Dallas Federal Reserve and conference participants at Midwest Trade 2015, Southern Economic Association Meetings 2016, Rocky Mountain Empirical Trade 2016, and Society of Economic Dynamics Meetings 2017.

†jwjung@kiep.go.kr
‡inasimonovska@ucdavis.edu
§ariel.weinberger@ou.edu

1 Introduction

The empirical international trade literature has documented several stylized facts that shed light on the role that heterogeneous firms play in shaping prices and flows of goods across borders. Bernard et al. (2003) and Bernard et al. (2012) document that exporters constitute a minority of producing plants but they enjoy a large productivity and size advantage over non-exporters. 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. Finally, Arkolakis (2010) and Mrazova et al. (2015) highlight that the size distribution of firms includes a Pareto right tail and the existence of many small exporters. Reconciling these facts proves to be important for the quantification of gains from trade as derived in Arkolakis et al. (2017).

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 (GCES) 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. Within a market, firms charge variable markups, but the markup of the most productive firms is bounded above by the familiar Dixit-Stiglitz markup. 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 an advantage in sales and measured value added 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. (2017) (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 66 countries to uncover all, but two of, the parameters necessary to simulate moments at the firm level. The two key remaining

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

parameters constitute the Pareto shape parameter and the utility curvature parameter. We identify these parameters by matching moments from the sales and markup distributions for the universe of Chilean domestic and exporting firms.

The fact that the model falls within the non-homothetic class of models studied in ACDR is important because the same key model parameters that drive welfare are those that ensure the quantitative fit of the model to cross-sectional data. Yet, the nature of our quantitative exercise is different than ACDR's. While ACDR compare welfare gains from trade between homothetic and non-homothetic models for a given *identical* trade elasticity and given aggregate trade data, we parameterize our non-homothetic model so as to reconcile aggregate trade observations as well as sales and markup distributions observed in the data. The ACDR welfare results rely on two structural parameters: the shape parameter of the productivity distribution and the revenue-weighted average elasticity of markups with respect to marginal costs. Since the sales and markup distributions contain information about these two parameters, we argue that our model's ability to match moments from these distributions is essential for quantifying welfare and running counterfactuals. We do not calculate the welfare gains from trade – the formula is derived in ACDR – but instead we estimate the parameters necessary to compute the gains from a counterfactual reduction in the domestic consumption share and we examine the implications of these estimates for a number of micro- and macro-level moments in cross-country data.[2]

We find that the GCES model performs very well with respect to both micro and macro data; hence, it 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. We provide an algorithm to take this model to the data that requires minimal computational costs. A natural question, however, is whether existing non-homothetic frameworks that share key qualitative predictions with the GCES model represent equally plausible alternatives. In particular, Melitz and Ottaviano (2008) (MO), Behrens et al. (2014) (BMMS), and Simonovska (2015) (SIM) offer frameworks that share the supply-side structure of the GCES model but rely on different utility functions.[3] Therefore, when framed in a multi-country general equilibrium setting, these models are qualitatively in line with the price discrimination 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 prices in all the models are proportional to the one generated by the GCES.[4] Hence, for

[2]Dhingra and Morrow (2018) and Weinberger (2017) examine the welfare consequences of variable markups and incomplete pass-through in similar frameworks.

[3]Feenstra and Weinstein (2017) also examine firm-level price variation in a monopolistically competitive world that features heterogeneous firms; however, the authors rely on translog preferences, which are homothetic.

[4]In the MO model, the parameter denoted by η must be set to zero. For completeness, we also examine

a given Pareto shape parameter value, all models—existing as well as GCES—yield identical productivity cutoffs and macro aggregates.

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 moments in the sales and markup distributions, unlike the flexible GCES model. We demonstrate theoretically that, in general equilibrium with free entry and under the assumption that firm productivity is Pareto distributed, a single parameter, the Pareto shape parameter, governs the sales and markup distributions in the three existing non-homothetic models. In those models, markups are unbounded which limits the sales of superstar firms. Alternatively, in the GCES, the markups of the largest firms approach a constant determined by the demand curvature parameter. Consequently, the existing models cannot reconcile observed behavior of the largest firms and yield different predictions about price dispersion and ultimately quantitative measures of welfare than the GCES model. In particular, when the models attempt to match the heterogeneity in sales and markup data, the resulting Pareto shape parameter and revenue-weighted markup elasticity are mismeasured, which highlights the need for a flexible parameterization such as the GCES.[5]

With the above discussion in mind, a natural question arises: why focus the analysis on these particular moments in the data? First, the cross-section of sales and markups have constituted core moments traditionally examined by the quantitative trade literature of micro-level heterogeneity (see Bernard et al. (2003), Arkolakis (2010), and Behrens et al. (2014) among others). Second, a unification of the two distributions has proven elusive. CES frameworks are successful on the sales side, but do not generate price discrimination. In this paper, we demonstrate that existing non-homothetic frameworks that are consistent with firm pricing behavior cannot generate the very large sales of the most productive firms seen in data.[6] Finally, as argued above, moments from these distributions directly speak to the key parameters that govern welfare in a non-homothetic model (see ACDR).[7] Nonetheless, for robustness: i) we evaluate the performance of the GCES along a number of other dimensions in the data; ii) in order to provide meaningful comparisons to the literature that estimates trade elasticities, we explore an alternative estimation that relies on micro-level price data similar in spirit to Simonovska

MO's general non-separable case qualitatively and quantitatively.

[5]Feenstra (2018) 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 find that incorporating a bounded Pareto into all the models examined in this paper further restricts the predicted sales heterogeneity.

[6]On this front, Mrazova et al. (2015) also attempt to bridge markup and sales distributions, though their paper does not explore a quantitative exercise to evaluate the performance along other dimensions.

[7]Our estimation procedure identifies the parameters necessary to compute *local welfare gains*. We discuss an extension that could recover additional parameters needed to compute global gains. It should be noted, however, that ACDR compute both for the GCES model and find that they are quantitatively very similar.

and Waugh (2014b), but modified accordingly to account for non-homothetic preferences.

Focusing on additional moments in the data for a cross-section of firms, we evaluate several predictions of the model. Using parameters estimated to match moments from the sales and markup distributions, the revenue-weighted pass-through of costs to markups is between 0.4 and 0.48, within the range documented by De Loecker and Warzynski (2012) and De Loecker et al. (2016). Further, the model yields average markups between 13 and 26 percent – consistent with the markups we find in Chile. Focusing on exporters, as documented by Simonovska (2015), there is a positive relationship between prices of tradable goods and destination per-capita income, although our estimated elasticity of price with respect to income is higher than the one reported in the data.[8] Relatedly, the minority of firms are exporters and the majority of these exporters sell mostly domestically, although we do not match the percent of firms that export due to our simple gravity-based trade-barrier estimation strategy which relies on bilateral trade data only. We outline estimation strategies that could match the number of exporters, but require additional data that we do not have access to. Finally, we compare our benchmark estimation results that target distributional moments to an alternative strategy that relies only on micro-level price data to match the trade elasticity. This yields an estimate for the Pareto shape parameter that is similar to our benchmark as well as to results found in the trade-elasticity literature, notably in Simonovska and Waugh (2014b).

Overall, the GCES model fits the data better than existing non-homothetic frameworks. In the GCES model, sales follow a Pareto distribution in the right tail, and there exist a large number of very small (in fact infinitesimally so) firms as in the data. We also outline shortcomings that apply to the use of a Pareto distribution in the quantitative analysis. For a given Pareto shape parameter, the model matches one important moment from the sales distribution–the ratio of average domestic sales of exporters relative to non-exporters. To simultaneously target multiple moments of the empirical sales distribution, we conduct an over-identified estimation that further targets the ratio of sales for firms in the upper percentiles relative to firms in lower percentiles. The strategy yields a larger productivity dispersion and a higher substitution parameter relative to the exactly-identified benchmark. It under-predicts the size advantage of the firms in the 90th relative to the 10th percentile, but over-predicts the size advantage of exporters over non-exporters.[9] The short-comings exemplified by this exercise motivate an avenue for future research: to explore non-homothetic demand along with more flexible productivity distributions.

We contribute to a large and important literature that emphasizes the role that heteroge-

[8]It is difficult to compare the estimates quantitatively because we compute the average elasticity for all Chilean exporters while Simonovska (2015) documents an estimate for one large Spanish exporter (Mango). In our model, large firms price discriminate less, which is consistent with Mango's lower magnitude.

[9]Still, a test for over-identifying restrictions does not reject the null that the model is valid.

neous 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, sell more and enjoy higher measured productivity.[10] 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.

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) 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. In addition, recent studies such as Bas et al. (2017), Sager and Timoshenko (2017), and Fernandes et al. (2017) examine more flexible productivity distributions to rationalize observed export sales moments. The unifying feature of these models 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. In contrast, the GCES model rationalizes the existence of small exporters via non-homothetic preferences, which also yield predictions about the cross-section of prices that are in line with data.

We organize the remainder of the paper as follows. We outline the supply side of the economy in Section 2. In that section, we also outline the GCES model and go on to derive important predictions in line with the current heterogeneous-firm literature. We compare these to the existing non-homothetic models in Section 3. We describe the solution algorithm for the GCES model and we generate moments for estimation in Section 4. In Section 5 we quantify

[10]Notably, deBlas and Russ (2015) and Edmond et al. (2015) generalize the BEJK model and derive richer predictions about aggregate pass-through and welfare.

the GCES model and analyze the fit to micro and macro data. We conclude in Section 6.

2 Generalized CES Model

In this section we outline the generalized CES (GCES) model. We begin by describing the market structure, which the model shares with existing models that we summarize in Section 3.

2.1 Framework

The environment is static. Goods are differentiated by the producers' identity and by 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 destination country. In every country, i, there exists a pool of potential entrants who pay a one time cost, $f_e > 0$, which is common across countries. Since labor is the only factor of production, the fixed cost is paid in units of labor. Each entrant gets 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 productivity distribution is Pareto with shape parameter θ and lower bound b_i, which governs average productivity in i. We denote J_i as the measure of entrants in market i. Only a subset of these entrants have positive demand in 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. We assume that each consumer has a unit labor endowment which, when supplied (inelastically) to the local labor market, earns a wage rate of w_i; hence, per-capita income in country i necessarily equals w_i.

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:

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

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 3 we explore the role that utility parameterizations as in the SIM, (separable) MO, and BMMS models play in governing cutoff costs as well as mark-up and sales distributions of firms. In Appendix F we repeat the analysis for the non-separable/flexible MO model. Below, we outline the GCES model.

2.2 Consumer and Firm Problems

The GCES 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) (SIM). Relative to the SIM specification, the GCES 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 constant, and $\sigma \geq 1$ governs curvature. Notice that a monotonic transformation of the utility function is possible such that $U(0) = 0$.[11] Finally, the sign of $\bar{q}$ is important. If it is equal to 0, our utility function becomes the ubiquitous CES and preferences are homothetic. A negative $\bar{q}$ implies that the demand elasticity increases with sales, and therefore more productive firms have lower markups. Conversely, our assumption of $\bar{q} > 0$ will imply that more productive (and larger) firms have larger markups.[12]

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

[11]For $q^c(\omega) = 0$, $u = \left(\bar{q}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$. Since these are all parameters, a monotonic transformation such that $u(0) = 0$ is possible by just subtracting from the utility function the value of those parameters. For example, we could examine the following utility function: $U^c = \left(\int_{\omega\in\Omega}(q^c(\omega) + \bar{q})^{\frac{\sigma-1}{\sigma}} - \bar{q}^{\frac{\sigma-1}{\sigma}}\right) d\omega$.

[12]This is supported by much of the empirical literature (e.g. De Loecker and Warzynski (2012), Berman et al. (2012), and others), and consistent with Marshall's "Second Law of Demand."

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:

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

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

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.

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 (2003) model, which also has implications for the shape of the sales distribution of the largest firms. 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 GCES 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. As shown in the next subsection, sales of the largest firms reflect those of a model with CES demand. In contrast, firms will not demand infinite labor in the $\sigma = 1$ model. There can be many such firms— which bounds above differences in sizes between exporters and non-exporters. In Section 3 we discuss why this limitation, which is shared by other existing non-homothetic models, is an important aspect of our analysis.

2.3 Equilibrium

To derive aggregate predictions, we follow the insights by Arkolakis et al. (2017) (ACDR) and we define a new variable at the firm level: $t_{ij} \equiv \frac{p_{ij}}{\bar{c}_j}$. This strategy will yield very tractable aggregate equilibrium objects, which will greatly simplify the analysis. 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):

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

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

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

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

where β_1, β_2, β_P and $\beta_{\sigma P}$ are constants that depend on θ and σ, and $\pi_{i\upsilon}(c)$ denotes the profit of a firm with cost draw c from country i in destination υ and is characterized in expression (38) in Appendix A. 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 independent of the destination country, 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:

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

$$w_i L_i = \sum_{\upsilon=1}^{I} T_{\upsilon i} \tag{16}$$

Specifically, plugging (11) and (12) into (15) and (16) yields the measure of entrants: $J_i = \beta_E L_i / f_e$, where β_E 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:

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

2.4 Model Predictions

2.4.1 Sales Distribution

A key feature of the GCES model that differentiates it from existing frameworks with directly additive non-homothetic preferences is its prediction about the shape of the distribution of firm sales. There is no closed form solution of the sales distribution, but an important aspect is that the sales of the largest firms asymptote to the sales in the model that features CES preferences. This feature yields a sales distribution with a Pareto right tail, consistent with empirical observations made in Mrazova et al. (2015).

To examine the behavior of the very large firms, combine Equations (8) and (9) to write sales as $r_{ij}\left(c_{ij}\right) = \left(\frac{x_{ij}(c_{ij})}{L_j \bar{q}} + 1\right)^{-1/\sigma} \bar{c}_j x_{ij}\left(c_{ij}\right)$. [13] Then, in the limit sales are CES:

$$\begin{aligned} \lim_{x\to\infty} p_{ij}\left(c_{ij}\right) x_{ij}\left(c_{ij}\right) &= \lim_{x\to\infty} \bar{c}_j x_{ij}\left(c_{ij}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{1}{x_{ij}\left(c_{ij}\right)} + \frac{1}{L_j \bar{q}} \right)^{-1/\sigma} \\ &\propto x_{ij}\left(c_{ij}\right)^{\frac{\sigma-1}{\sigma}} \end{aligned} \tag{21}$$

Thus, the distribution of sales will be Pareto in the tail as long as productivity is distributed

[13] $r_{ij}\left(c_{ij}\right) = m_{ij}\left(c_{ij}\right) m_{ij}\left(c_{ij}\right)^{-1} \left(\frac{x_{ij}(c_{ij})}{L_j \bar{q}} + 1\right)^{-1/\sigma} \bar{c}_j x_{ij}\left(c_{ij}\right)$ using the fact that $\frac{c}{\bar{c}} = m^{-1} \left(\frac{x}{L\bar{q}} + 1\right)^{-1/\sigma}$.

Pareto (Mrazova et al. (2015)).[14] We will show in the next section that existing directly additive non-homothetic models cannot reconcile this behavior.

The GCES model also has the attractive feature of replicating the existence of small firms found in census data, a fact established by Arkolakis (2010).[15] Although this feature is an improvement relative to the Melitz-Chaney framework, it is also a feature of other non-homothetic models since those also rely on a bounded marginal utility. For that reason we do not stress this point in the quantitative analysis but do stress that it should be a part of quantitative exercises that rely on assumptions about the firm sales distribution.

2.4.2 Distribution of Markups

Next, we investigate the distribution of markups in the GCES model.[16] The markup of a firm from i in destination j relative to the average markup in j is expressed as: $\tilde{m}\left(v_{ij}\right) = \frac{m_{ij}(c_{ij})}{\bar{m}_{ij}(\bar{c}_{ij})}$, where $v_{ij} \equiv c_{ij}/\bar{c}_j$ and $\bar{m}_{ij}(\bar{c}_{ij}) \equiv \frac{\int_0^{\bar{c}_{ij}} m_{ij} d\mu_{ij}(c_{ij})}{\mu(\bar{c}_j)}$. Given the definition of $d\mu_{ij}(c_{ij})$, the latter becomes:

$$\bar{m}_{ij}(\bar{c}_{ij}) = \frac{\theta}{\sigma^{\theta-1}} \int_0^1 \left[t_{ij}^{\sigma+1} + (\sigma-1)t_{ij}\right]^{\theta-2} \left[(\sigma+1)t_{ij}^{\sigma+1} + (\sigma-1)t_{ij}\right] dt_{ij}. \tag{22}$$

The mean markup reduces to a constant that is a function of two main parameters, θ and σ, which we rewrite as $\bar{m}_{ij}(\bar{c}_{ij}) = \beta_M(\theta, \sigma)$.

We use the distribution for normalized markups, $Pr[\tilde{M} \geq \tilde{m} | \tilde{M} \geq \tilde{m}^{min}] = 1 - F(\tilde{m}) = (v_{ij})^{\theta}$, where $\tilde{m}^{min}$ is the markup of the firm with a cost equal to the cutoff and F is the distribution of $\tilde{m}$, to get a closed form solution of the distribution of normalized markups:

$$F^{GCES}(\tilde{m}) = 1 - \left[\left(\frac{1}{\tilde{m}}\right)^{\theta} \left(\frac{\beta_M(\theta,\sigma)}{(1-\sigma)\beta_M(\theta,\sigma) + \sigma\tilde{m}^{-1}}\right)^{\frac{-\theta}{\sigma}} \left(\frac{1}{\beta_M(\theta,\sigma)}\right)^{\theta}\right] \tag{23}$$

2.4.3 Pass-through Elasticity

An advantage of the model analyzed in this paper is that it fits within the class of models studied by Arkolakis et al. (2017) (ACDR). We can therefore refer to their (local) welfare results which show that gains from a reduction in domestic expenditure share are governed by the Pareto shape parameter and the weighted average markup elasticity with respect to marginal cost (or

[14] This is consistent with our findings in the Chilean firm-level data that we describe below.

[15] That framework is successful at explaining the behavior of small and large firms as has been documented in firm-level datasets. Our model allows us to capture the observed sales heterogeneity, but the two frameworks differ in their predictions regarding the distribution of markups.

[16] Markups are defined as the price to cost ratio: $m_{ij} \equiv \frac{p_{ij}}{c_{ij}}$.

one minus the price-cost pass through elasticity).[17] With variable markups, a reduction in trade costs allows foreign firms to raise the markups due to incomplete pass-through and this lowers the gains from trade relative to the case where markups are constant. The magnitude of this latter effect depends on a revenue-weighted average markup elasticity with respect to marginal costs.

To solve for this elasticity in the GCES model, we write markups as the implicit solution to the following function:

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

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 country. Following ACDR, we compute the expenditure-weighted average of the markup elasticity:

$$\rho = \frac{\int_0^1 \left[\frac{\sigma}{(\sigma+1)+(\sigma-1)m^{-\sigma}v^{-\sigma}}\right] mv\left[m^{-\sigma}v^{-\sigma}-1\right] v^{\theta-1} dv}{\int_0^1 mv\left[m^{-\sigma}v^{-\sigma}-1\right] v^{\theta-1} dv} = \frac{\int_0^1 \left[\frac{\sigma-\sigma m(v)+m(v)}{\sigma-\sigma m(v)+m(v)+1}\right] r(v) d\mu(v)}{\int_0^1 r(v) d\mu(v)}. \quad (24)$$

The predictions imply that there is a relationship between the model's implications for the sales and the markup distribution. The behavior of the very large firms allows us to capture the finite number of "superstar" firms observed in various datasets. This is possible due to the upper bound on markups, which implies a smaller pass-through elasticity in the GCES model compared to the other directly additive models where the largest firms have infinite markups. The reason why these predictions are important is that they will contain information about the Pareto shape parameter and the revenue-weighted markup elasticity, the two parameters that govern the local welfare gains from a reduction in the domestic consumption share (Arkolakis et al. (2017)). In the quantification, we take moments from these two distributions to identify the Pareto shape parameter and revenue-weighted trade elasticity.

2.4.4 Income Per Capita and Prices

A key prediction of the model is that prices of tradable goods are higher in richer destinations, as documented by Alessandria and Kaboski (2011) . In fact, 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)). 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

[17] 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.

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.

Different firms, however, price discriminate to different degrees. In order to compare the quantitative performance of the model to existing frameworks, it is useful to derive moments from the distribution of price-income elasticities. We compute the average price-income elasticity of all exporters from source country i, which we denote by ϵ^*_{p-w}. This is all relegated to Appendix A.3 because these expressions are very cumbersome. In summary, for a given σ and θ, the average can be computed using simulated firms as described in Section 4. In Section 5 we use the calibrated parameters and we compute the average elasticity of tradable goods' prices with income. We compare this result with the findings in Simonovska (2015) who uses prices of one company (Mango) that exports to a variety of destinations.[18]

2.5 Discussion

The predictions above suggest that our framework can be used to reconcile important facts about firm heterogeneity. With respect to firm sales, there is a left tail in the distribution, while the right tail reflects the success that CES frameworks have had in replicating the behavior of the largest firms. Our model can also generate markups that increase with firm size, existence of incomplete pass-through, and firm price discrimination on the basis of consumer income. ACDR show that reconciling these facts is crucial: the size of the welfare gains from trade depend on the productivity dispersion, the ability of firms to pass through cost advantages onto markups, and the sales weights of firms. Therefore, to quantify the gains from trade, the structural parameters should be disciplined in a way that the model matches moments from the sales and markup distributions. This motivates the moments from the data that we target in our estimation in Section 4.

3 Existing Non-homothetic Models

In order to evaluate the contribution of the GCES model to the literature, we discuss key predictions from widely-used models that feature non-homothetic preferences: Simonovska (2015) (SIM), Behrens et al. (2014) (BMMS), and the separable-preference case of Melitz and Ottaviano (2008) (MO).[19] We derive predictions within the general framework developed at the

[18]In the quantitative section, we document different moments for Chilean firms. However, we cannot calculate this elasticity for Chilean exporters since we do not have the necessary customs data.

[19]This corresponds to the quadratic utility in MO with $\eta = 0$. For completeness, we solve the general case where $\eta > 0$ in Appendix F.

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 finite choke price. Hence, the predicted welfare gains can also 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. 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). Markup differences directly translate into differences in value added, measured productivity, and sales across firms.[20]

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 in Alessandria and Kaboski (2011). The distinctive feature of these models is that preferences are non-homothetic. In particular, 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.

In the remainder of this section, we derive the predictions regarding the mark-up and sales distribution in the SIM, MO, and BMMS models and we demonstrate the limitations of these frameworks along this particular dimension. To do that, we first present the utility specification for each model while keeping the same supply environment as in Section 2. In each case, we solve for the cost cutoffs, firm prices, markups and sales in Appendices B-D. In Appendix G, we explore quantitative properties of the three separable non-homothetic models as well as of the non-separable MO model with $\eta > 0$.

Simonovska (2015)—SIM Assume each country is populated by identical consumers of measure L, whose utility is:

$$U^c = \int_{\omega\in\Omega} \log(q^c(\omega) + \bar{q})d\omega \tag{25}$$

[20]Recently, ACDR point to these moments, but the authors perform a separate quantification exercise. In addition, BMMS examine the measured-productivity moment in their model.

where $q^c(\omega)$ is individual consumption of good ω, Ω is a compact set, and $\bar{q} > 0$ is a constant common across countries. Notice that this is a special case of the GCES model in Section 2, which occurs when $\sigma \to 1$. Below we show that this particular parameterization yields very different predictions regarding the sales behavior of large firms compared to the GCES model where $\sigma > 1$.

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.

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." We will follow Behrens et al. (2014) in order to derive closed form solutions of aggregates that involve the Lambert function (shown in Footnote 22 below).

3.1 Sales and Markup Distributions

In each of the three models above, the cost cutoff is proportional to the one for the GCES model summarized in expression (20), differentiated only by the parameters in the first part of Equation (20). We derive the cost cutoffs for the three models, respectively, in Appendices B-D. For a given cutoff, we compare the sales and markup distributions of each existing model to the GCES explored in Section 2.4.

In order to suppress country-specific scale differences, for any model M, we examine the (normalized) sales of a firm from country i in destination j, relative to the average sales there,

$\bar{r}_{ij}^M = \int_0^{\bar{c}_j^M} r_{ij}^M(c) d\mu_{ij}(c)/\mu(\bar{c}_j^M)$:[21]

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

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

$$\tilde{r}^{SIM}\left(\frac{c_{ij}}{\bar{c}_j^{SIM}}\right) = (1+2\theta)\left(1-\left(\frac{c_{ij}}{\bar{c}_j^{SIM}}\right)^{\frac{1}{2}}\right) \qquad \text{if } 0 \leq \frac{c_{ij}}{\bar{c}_j^{SIM}} \leq 1 \tag{26}$$

$$\tilde{r}^{MO}\left(\frac{c_{ij}}{\bar{c}_j^{MO}}\right) = \frac{\theta+2}{2}\left(1-\left(\frac{c_{ij}}{\bar{c}_j^{MO}}\right)^{2}\right) \qquad \text{if } 0 \leq \frac{c_{ij}}{\bar{c}_j^{MO}} \leq 1 \tag{27}$$

$$\tilde{r}^{BMMS}\left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}\right) = \frac{1}{\kappa_1(\theta)}\left[e^{W\left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}e\right)-1} - \left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}\right)\right] \quad \text{if } 0 \leq \frac{c_{ij}}{\bar{c}_j^{BMMS}} \leq 1, \tag{28}$$

where $\kappa_1(\theta) = \theta e^{-(\theta+1)} \int_0^1 (z^{-1}-z)(ze^z)^\theta e^z dz$ is a positive constant that depends only on θ.[22] Notice that in all three cases, sales are increasing, concave in firm productivity (inverse of cost), and bounded from above.[23] Furthermore, the normalized firm-level sales are once again neither source nor destination specific since cost draws relative to cutoffs, $c_{ij}/\bar{c}_j^M$, 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 $Pr[\tilde{R} \geq \tilde{r} | \tilde{R} \geq \tilde{r}^{min}] = 1 - F^M(\tilde{r}) = \left(\frac{c_{ij}}{\bar{c}_j^M}\right)^\theta$, where $\tilde{r}^{min}$ is the normalized sales of the cutoff firm and F^M is the distribution of $\tilde{r}$ in model M. Therefore, for each model we can derive the distribution of normalized sales as:

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

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

$$F^{BMMS}(\tilde{r}) = 1 - \left[(\tilde{r}\kappa_1(\theta))^\theta \left(\frac{1}{W\left([1-F^{BMMS}(\tilde{r})]^{1/\theta} e\right)} - 1\right)^{-\theta}\right],$$

[21] $r_{ij}^M(c)$ are model-specific firm revenues. For the SIM, MO, and BMMS models they are given respectively by equations (48), (54), and (59) in Appendices B-D.

[22] To understand $\kappa_1(\theta)$, define the following change of variables: $z \equiv W\left(\frac{c_{ii}}{\bar{c}_i^{BMMS}}e\right) \Rightarrow \frac{c_{ii}}{\bar{c}_i^{BMMS}}e = ze^z$. W denotes the Lambert W function with argument $\left(\frac{c_{ij}}{\bar{c}_j^{BMMS}}e\right)$ and $\bar{c}_j^{BMMS}$ is the cost cutoff for this model, which we derive in detail in Appendix D.

[23] In each model, $c_{ij} = \bar{c}_j^M$ 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. We demonstrate next that the same is true for the markup distributions in these models.

As we did for sales, let $\tilde{m}\left(\frac{c_{ij}}{\bar{c}_j^M}\right)$ be the markup relative to the mean markup, $\bar{m}_{ij}^M = \int_0^{\bar{c}_j^M} m_{ij}^M(c)d\mu_{ij}(c)/\mu(\bar{c}_j^M)$. The closed form solutions for the mean markups are shown in Appendix E. We point out here that in all three cases markups are unbounded. This is related to the point about bounded sales: since markups approach infinity as the cost approaches zero, then the sales of these zero cost firms are finite. Finally, for each model we can derive the distribution of normalized markups:

$$\Gamma^{SIM}(\tilde{m}) = 1 - \left(\frac{\theta - 0.5}{\theta\tilde{m}}\right)^{-2\theta}$$
$$\Gamma^{MO}(\tilde{m}) = 1 - \left[\frac{2\theta - 1}{\theta - 1}\tilde{m} - 1\right]^{-\theta}$$
$$\Gamma^{BMMS}(\tilde{m}) = 1 - \left[e^{W\left([1-\Gamma^{BMMS}(\tilde{m})]^{1/\theta}e\right)-1}\left(\frac{1}{\tilde{m}}\right)\left(\frac{1}{\kappa_1(\theta)}\right)\right]^{\theta}.$$

As is apparent, the distributions of markups are also governed by θ in all three models. The two sets of distributions derived above demonstrate the limitations of the existing models. An important advantage of the GCES model is the flexibility to jointly reconcile the sales and markup behavior in the data.[24] Limiting the sales distribution implies that the sales-weighted average of the elasticity of markups with respect to costs (ρ) in the restrictive cases must be pinned down by θ as well (see Appendix E). Therefore, the lack of flexibility *has implications for the estimation of the trade elasticity and the revenue-weighted markup elasticity with respect to costs*, which are necessary to compute the welfare gains from trade presented in Arkolakis et al. (2017).[25]

4 Solution Algorithm

This section describes our solution algorithm to conduct a quantitative analysis of the GCES model. A similar exercise can be done for the existing models in the literature, but given

[24]Recently Bertoletti et al. (2018) present a heterogeneous-firm model of trade where consumer preferences are indirectly additive (IA)—a distinct class of non-homothetic preferences from the directly additive class studied here. The IA framework also has the flexibility to reconcile the markup and sales advantage of exporters over non-exporters. However, it does not match well the behavior of the largest firms. A desirable feature of GCES is that sales asymptote to CES for the largest firms, thus matching the data well.

[25]We discuss the non-separable version of the MO model, namely the case in which $\eta > 0$, in Appendix F.

that only the GCES has the flexibility to match moments from both the markup and the sales distribution, we restrict the analysis to this model.[26]

In order to derive quantitative cross-sectional and aggregate predictions, we need to solve the model via simulation. Relative to the existing literature, we aim to match key moments on firm sales and markups, which we document using Chilean firm-level data. Although we use the gravity framework and workforce data to compute wages and cost cutoffs for every country (relative to a numeraire), the moments in the model that pin down the remaining parameters are created using simulated firms. In other words, for a guess of the parameters, we simulate firm-level outcomes and attempt to reproduce moments from the sales and markup distributions in the data. The continuum of firms in the model is discretized, which means that the number of simulated draws has to be large enough so as to best approximate the entire continuum. 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.

For all countries $i = 1, ..., I$, we derive predictions for all firms 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. Using (20), given wages, trade costs, and labor force, we compute all cutoffs *relative to a numeraire*. We proceed as follows. We draw 3,000,000 realizations 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 $c_{ii} = s^{\frac{1}{\theta}} \bar{c}_i$. We thus have 3 million firms in each country, with only the cutoffs being specific to each country. 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. Notice that the number of exporters and the number of destinations those exporters can reach vary across origins.

4.1 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 labor L_i relative to a numeraire country. Similarly, we compute wages and cost cutoffs relative to a numeraire country.

[26] In the Online Appendix, we conduct a full quantitative analysis of all these models along with the GCES. Appendices B-D do describe how to quantify the other models and Appendix G discusses the general conclusions.

To reduce the computational burden, we leverage equation (17) to solve for all endogenous objects for any guess of the parameters. 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: $\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = \Phi_j - \Phi_i - \theta \log(\tau_{ij})$, with $\Phi_i \equiv \log(L_i(w_i/b_i)^{-\theta})$ $\forall i$ recovered from country-specific fixed effects.[27] 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: $\log\left(\frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{jj}}\right) = \hat{\Phi}_j - \hat{\Phi}_i - \theta \log(\hat{\tau}_{ij})$. To recover the wages, we substitute estimated trade shares into (18) and we use data for L_i.[28] 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), where again these are relative to a numeraire. 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, Φ_i, only.*

Given estimates of Φ_i, data on labor force, and estimated wages, we could back out the technology parameters b_i with the definition of $\Phi_i = \log(L_i(w_i/b_i)^{-\theta})$. For each guess of θ, the technology parameters would satisfy $\hat{b}_i = \left(\frac{\exp(\hat{\Phi}_i)}{L_i}\right)^{1/\theta} \hat{w}_i$, 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

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

[28] 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.

moments that relate to Chilean exporters and non-exporters, the technology parameters are not necessary. Moreover, given the predictions of the model that we are interested in, the technology 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 labor force data. This allows us to also characterize cost cutoffs relative to a benchmark country, which renders the scale of the cutoffs irrelevant. From (20), notice that:

$$\Psi_j = \frac{\bar{c}_j}{\bar{c}_i} = \left[\frac{\hat{w}_j / \hat{w}_i}{\frac{\sum_{v=1}^{I} \exp(\hat{\Phi}_v)(\hat{\tau}_{vi})^{-\theta}}{\sum_{v=1}^{I} \exp(\hat{\Phi}_v)(\hat{\tau}_{vj})^{-\theta}}} \right]^{\frac{1}{\theta+1}} . \tag{29}$$

The denominator is recovered completely from gravity, while wages are recovered from (18). Then, $\bar{c}_j = \Psi_j \bar{c}_{Chile}$.[29] The third and final step will be to simulate firms, and use these firms to create moments that identify σ and θ. Although each moment could be constructed for any of the countries that appear in the bilateral trade data, we only have access to firm-level Chilean data, so we choose to compare the moments in the model to moments for Chilean firms. Thus, the moments from the sales and markup distribution that are detailed below are computed only for $i = Chile$.

4.1.1 Identifying θ and σ

To identify these two parameters, we begin with an exactly-identified calibration exercise and we engage in over-identified exercises for robustness. We focus on moments from the sales and markup distribution because they directly speak to the key parameters of interest. First, we derive the following two moments from the model that we perfectly match in the Chilean data: 1) measured markup advantage of exporters over non-exporters (in logs) and 2) domestic sales advantage of exporters over non-exporters. These are equivalent to the two key moments described in BEJK, since the markup in our model with no intermediate inputs is equal to the value added per worker. However, i in our case should be interpreted as $Chile$, as we are only trying to match the behavior of *Chilean* firms. For robustness, in an over-identified strategy, we add moments from the sales and markup distribution of firms at different percentile levels. In particular, we use the following percentile ratios: 99-90, 90-10, and 90-50.

[29] An important implication is that the parameter that characterizes the gains from variety ($\bar{q}$) will also cancel out and is essentially governed by gravity variables.

Markup Advantage. To derive the first moment, we begin by deriving value added per worker as in BEJK. 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 $\tau_{i\upsilon}$, a firm must produce $\tau_{i\upsilon} x_{i\upsilon}$ so that $x_{i\upsilon}$ units reach the consumer. With no intermediate goods, variations in markups are the same as variations in measured productivity, where 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: $\tilde{c}_{ii}^x \equiv \max_{\upsilon \neq i} \frac{\bar{c}_\upsilon}{\tau_{i\upsilon}}$. 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 $c_{ii} = \tilde{c}_{ii}^x$ into expression (10). It yields the desired value implicitly: $(1-\sigma)\tilde{t}_{ii} + \sigma \frac{\tilde{c}_{ii}^x}{\bar{c}_i} = \tilde{t}_{ii}^{\sigma+1}$. Notice that $\tilde{t}_{ii} = f(\frac{\tilde{c}_{ii}^x}{\bar{c}_i}, \sigma)$; 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 markup:

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

where $\xi_i \equiv \frac{\tilde{c}_{ii}^x}{\bar{c}_i}$ and $\beta_{MP,i}^{NX}(\sigma,\theta,\tilde{t}_{ii})$ is a constant term given $(\sigma,\theta,\tilde{t}_{ii})$.[30] 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 $t_{ii}(s)$. We integrate over the logged measured productivity of all exporters (expression relegated to the Appendix), which yields the average exporter logged markup:

$$
MP_i^{EXP} = \frac{\theta}{(\xi_i\sigma)^\theta}\beta_{MP,i}^{EX}\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)+\frac{1}{(\xi_i\sigma)^\theta}\left(\tilde{t}_{ii}(\sigma-1+\tilde{t}_{ii}^\sigma)\right)^\theta\log(w_i)
$$

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 $\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)$.[31] We point out that there is no closed form expression for MP_i^{EXP}, but we can solve it numerically using the simulated firms.

Finally, the markup moment is defined as:

$$
\begin{aligned}
M_{i,mkup}^m &= MP_i^{EXP} - MP_i^{NX} \\
&= \frac{\theta}{(\xi_i\sigma)^\theta}\beta_{MP,i}^{EX}\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)-\frac{\theta}{\sigma^\theta(1-\xi_i^\theta)}\beta_{MP,i}^{NX}(\sigma,\theta,\tilde{t}_{ii}) \\
&+ \frac{\left(\tilde{t}_{ii}(\sigma-1+\tilde{t}_{ii}^\sigma)\right)^\theta}{(\xi_i\sigma)^\theta(1-\xi_i^\theta)}\log(w_i)-\frac{1}{1-\xi_i^\theta}\log(w_i)
\end{aligned} \tag{30}
$$

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

[31]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 which requires the simulated firms. The main takeaway here is that it depends only on $\left(\{\chi_{ij}\}_{j=1}^I,\{L_j\}_{j=1}^I,\tilde{t}_{ii},\sigma,\theta\right)$.

Sales Advantage. The second moment of interest is the domestic sales advantage of exporters. Following BEJK, we derive the ratio between the average domestic sales of exporters and non-exporters. Using equations (9) and (10), and integrating, yields the corresponding statistic in the model (again, for an explicit closed-form expression of this moment see Appendix A.4):

$$M^m_{i,sales} = \frac{1-\xi_i^\theta}{\xi_i^\theta} \frac{\int_0^{\tilde{t}_{ii}} \left[t_{ii}(s)^{1-\sigma} - t_{ii}(s)\right]\left[(\sigma+1)t_{ii}(s)^{\sigma} + (\sigma-1)\right]\left[t_{ii}(s)^{\sigma+1} + (\sigma-1)t_{ii}(s)\right]^{\theta-1} dt_{ii}(s)}{\int_{\tilde{t}_{ii}}^1 \left[t_{ii}(s)^{1-\sigma} - t_{ii}(s)\right]\left[(\sigma+1)t_{ii}(s)^{\sigma} + (\sigma-1)\right]\left[t_{ii}(s)^{\sigma+1} + (\sigma-1)t_{ii}(s)\right]^{\theta-1} dt_{ii}(s)}. \tag{31}$$

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

Remaining Parameters. It is worth pointing out that neither of our two moments of interest depend on $\bar{q}$ or f_e as can be seen from the moment equations derived above. Additionally, cost cutoffs (modulo proportionality constant) are independent of these parameters, since the level of every cutoff is not necessary to derive the moments in the model. Essentially, $\bar{q}$ and f_e are subsumed into the gravity results. Hence, we need not identify these two parameters in order to study the moments of interest in this paper. It is for this reason that $\bar{q}$ and f_e were not included in the set of parameters Θ defined at the outset of this section.

The fact that $\bar{q}$ and f_e are not required for the moments we choose to target does not imply that these do not play a role in the model. In fact, $\bar{q}$ parameterizes the "love of variety" and is intuitively important for the level of welfare. However, ACDR show that the local gains from trade can be computed using only trade shares, θ, and ρ (defined in (24)). Targeting moments that are independent of the variety parameter allows us to simplify the estimation method by not parameterizing $\bar{q}$. Although this parameter would be necessary if we wanted to speak to global changes in welfare, ACDR show that the welfare results for large changes in trade costs are almost identical to the local gains. For this reason we view the simplification as an important advantage of our quantification strategy. We stress however that if we were to calibrate $\bar{q}$, and this resulted in different estimates for θ and ρ, then the welfare predictions would change as well. Next, we give one example where parameterizing $\bar{q}$ would be necessary.

The value for the utility parameter $\bar{q} > 0$ does not affect our moments of interest only because it is not country specific. In the case of country-specific utility parameters (i.e. $\bar{q}_j \neq \bar{q}_{j'}$), the cost cutoff would be given by $\bar{c}_j = \left[(\beta_{\sigma P} - \beta_P)\beta_E f_e^{-1}\right]^{-\frac{1}{\theta+1}} \left[\sum_v \frac{L_v b_v^\theta}{(\tau_{vj} w_v)^\theta}\right]^{-\frac{1}{\theta+1}} (w_j/\bar{q}_j)^{\frac{1}{\theta+1}}$. In this case, cutoffs (relative to a numeraire) no longer depend on gravity-estimated parameters alone as can be seen by the previous expression. In this case, we would need I more moments to identify all the parameters. In the next section we show that I such informative moments are the percentages of firms that export in *each* country. However, such data are not available, which makes the exercise infeasible with many countries.

4.1.2 Exact- and Over-Identified Techniques

Having derived the sales and markup advantage of exporters over non-exporters in the model, the exactly identified strategy involved estimating θ and σ by perfectly matching the two corresponding moments in Chilean data: $M^m_{CHL,sales} = M^d_{CHL,sales}$ and $M^m_{CHL,mkup} = M^d_{CHL,mkup}$, where the superscript d denotes data.

For robustness, we engage in over-identified estimation. In particular, we add more moments from the distribution of sales and markups. In both the data and the model, we compute the 99th, 90th, 50th, and 10th percentile of the two variables, and we construct 3 ratios: 99-90, 90-10, and 90-50, which adds six moments to the two moments above. We identify θ and σ using a GMM procedure with an optimal weighting matrix, which we discuss in detail in Appendix H.

4.2 Alternative Identification: Trade Elasticity

It is important to discuss the difference between our quantitative exercise and the exercise in Arkolakis et al. (2017). The authors use disaggregate trade data to identify σ and $\bar{q}$ by using changes in demand (through changes in tariffs) to explain changes in prices of traded goods. They compute the gains from trade for *a given trade elasticity* by assuming $\theta = 5$. An interpretation of their result is: "by fixing θ (the trade elasticity) to be the same across homothetic and non-homothetic models, how much of the difference in welfare comes from non-homothetic demand?" The main advantage of that method is it allows for a comparison across the two types of frameworks. Our exercise above is different since we pin down θ using moments from firm distributions.[32] The downside to this exercise is that our estimate of the trade elasticity may differ substantially from standard estimates in the literature, which would result in substantially different welfare gain estimates. For this reason, we consider an alternative identification strategy for θ, which is more in the spirit of the existing literature that estimates trade elasticities within gravity frameworks using price variation, but follows the structural estimation strategy described in the previous subsection.

We motivate the new set of moments in the alternative identification by deriving the expression for the trade elasticity that is identical in all additive models (of which the GCES is

[32] Notice that our method does not require us to pin down $\bar{q}$ as long as it is positive and equal across countries. This is an advantage since the parameter is not needed to identify θ, σ, nor ρ, so that local welfare gains from trade can still be computed. Furthermore, the assumption that it is constant across countries is taken in ACDR as well.

an example). Following similar arguments as in Simonovska and Waugh (2014a), relative trade shares can be written using relative cost cutoffs, and then using Equations (13) and (14), the following estimating equation arises:

$$\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) = -\theta \log \tau_{ij} + \log\left(\frac{w_i}{w_j}\right) + (\theta+1)\log\left(\frac{P_j}{P_i}\right) + (\theta+1)\log\left(\frac{N_i}{N_j}\right). \tag{32}$$

As is the case in ACDR, the trade elasticity is given by θ, which they set equal to 5. Instead, in this exercise, we use (32) to motivate a structural estimation of the parameters in our GCES model.

The strategy follows closely Simonovska and Waugh (2014a) and Simonovska and Waugh (2014b), and we refer to those papers for greater detail. Notice that the trade elasticity is identified through variation in trade frictions that is independent of trade flows plus the other aggregate statistics in (32). The gravity equation only allows one to recover trade frictions scaled by θ. The authors therefore rely on moments from price data that are informative about $\log \tau_{ij}$, such as the maximum price difference across identical goods between countries (Eaton and Kortum, 2002). However, those papers concentrate on homothetic models, while in this paper we require the trade elasticity of a non-homothetic model which yields different predictions about markups. A closed-form solution for prices is not available in our model, but in Appendix A.5 we derive price gaps under some simplifying assumptions. We argue that even in the non-homothetic model, similar price moments are informative about the trade elasticity, with the additional prediction that the price gaps covary with wages and trade shares. Therefore, if we estimate our model to match moments from the maximum price difference across goods between countries, and the covariance of these price differences with relative wages and bilateral trade shares, then θ reconciles the observed cross-country variation in tradable goods prices with the trade elasticity.

The empirical literature has established a strong link between price differences and income differences, but trade share differences have not been documented to drive price differentials (see discussion in Simonovska (2015)). Consequently, with these insights in mind, we choose three moments to infer information about trade costs and separate them from the parameter of interest, θ. The three moments are simulated in the GCES model and compared to their counterparts in the data. The first moment is the maximal price gap, adjusted for fixed effects: $d_{ni} = \log \tilde{\tau}_{ni} + \zeta_i + \zeta_n$ where $\log \tilde{\tau}_{ni} = \max_{l \in \tilde{N}} \{\log p_n(l) - \log p_i(l)\}$ and $\tilde{N}$ is the number of micro-level prices in the sample.[33] Notice that our estimation procedure detailed above allows us to simulate a matrix that includes the prices charged by all firms, in each of the 66 origin

[33] Fixed effects are incorporated by controlling for the average logged prices over all available goods in each country. In the simulation, we can record prices of all available goods in each country. We set $\tilde{N}$ to 110 to mimic the Economist Intelligence Unit data we try to match.

countries, to both home consumers and foreign consumers. Given trade costs and cost cutoffs in each destination, we keep track of the simulated firms that sell a "common good", defined as a firm from one origin that sells to at least a certain number of destinations. Of all the "common goods", we randomly choose 110 goods. The matrix of prices of the 110 goods in each destination is used to construct the model-implied maximal price gap d^m_{ni}. Similarly, we construct a second moment: the price gap at the 85th percentile of prices, also adjusted for fixed effects for countries i and n. These first two moments are used in Simonovska and Waugh (2014b), but simulated assuming alternative homothetic models. The third moment is the covariance of the maximal price gap with relative per-capita incomes in the two destinations. This is a new moment that is not used in the previous literature and captures the distinct prediction of the non-homothetic model that prices covary strongly with per-capita income. We construct this moment by computing the model-implied wages in each country (from Equation (18)), and then calculate the covariance of relative wages between each pair of countries and their maximal price gap as recorded in the simulation.[34]

The solution algorithm is identical to the one described earlier. First, we make a guess of the parameters, then we solve for trade costs scaled by θ, implied trade shares, wages, and cutoffs from market clearing. Armed with the necessary parameters and cost cutoffs, we simulate firms and compute their domestic price and export price to each destination (given they can sell below the destination cutoff). Having recorded the set of goods that are common to a certain number of destinations – defining a set of "common" goods that is as close as possible to the procedure used by our data source – we construct model-implied price gaps defined above.[35] Finally, to compute the data counterparts of the price moments, we require micro-level prices of identical goods sold in every destination. The Economist Intelligence Unit (EIU) database reports prices for 110 "commonly-available" goods for our sample of countries. This allows us to compute the price difference for the same good across destination pairs, which we use to construct the moments in the data.[36] Since the strategy is over-identified, we once again use a GMM procedure with an optimal weighting matrix. Finally, for robustness, we explore fixing σ and using the first of the three moments to exactly identify θ.

[34]Equation (45) in the Appendix makes it clear that the covariance of the maximal price gap and relative wages is informative about the trade costs. For completeness, we experimented with the covariance of the maximal price gap with logged trade shares and logged distance as well, with similar results.

[35]A good is defined as a firm-origin observation. We randomly choose 110 of the common goods that are simulated in the model, but produce 100 matrices for each simulation and take the average price gaps.

[36]Notice that a "common" good is not necessarily purchased by *all* countries. In the EIU data, there are some missing country-good observations. In our simulation, we check the results under different criteria for what constitutes a "common good".

5 Quantitative Results

In this section, we describe the data, report the estimation results, and judge the quantitative fit of the GCES model.

5.1 Data

There are two sets of data used in the quantitative analysis: firm-level data of the Chilean manufacturing sector to construct sales and markup distributions, and macro-trade data to construct the cutoff costs in Equation (20).

Firm-level data The Chilean data allows us to measure sales and markups for a cross-section of manufacturing firms. This census of manufacturing firms has been used extensively in the trade literature, including in Weinberger (2017). It is provided by Encuesta Nacional Industrial Anual (ENIA, National Industrial Survey) and collected by the National Institute of Statistics (INE). It covers a census of manufacturing firms, ISIC (rev. 3) classification 15-37, with more than 10 workers. There are approximately 5,000 firm level observations per year, although we restrict the sample to the year 2004 because we are interested in a cross section of firms. Each firm provides detailed economic data such as total sales, number of workers, wages, value of fixed capital, expenditures on intermediate inputs, etc. Importantly, firms also report the value of their total sales that are exports. We can therefore construct domestic sales for each firm and also markups using value added and compensation to labor.[37] This is the basis for the construction of the moments introduced in Section 4.1.

Data for trade shares We use data on 66 countries for the year 2004.[38] We construct bilateral trade shares, $\lambda_{ij} = \frac{X_{ij}}{Y_j}$, using nominal trade flows X_{ij}, and gross output (Y_j) adjusted for trade imbalances, all taken from UNIDO.[39] 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, the total workforce.

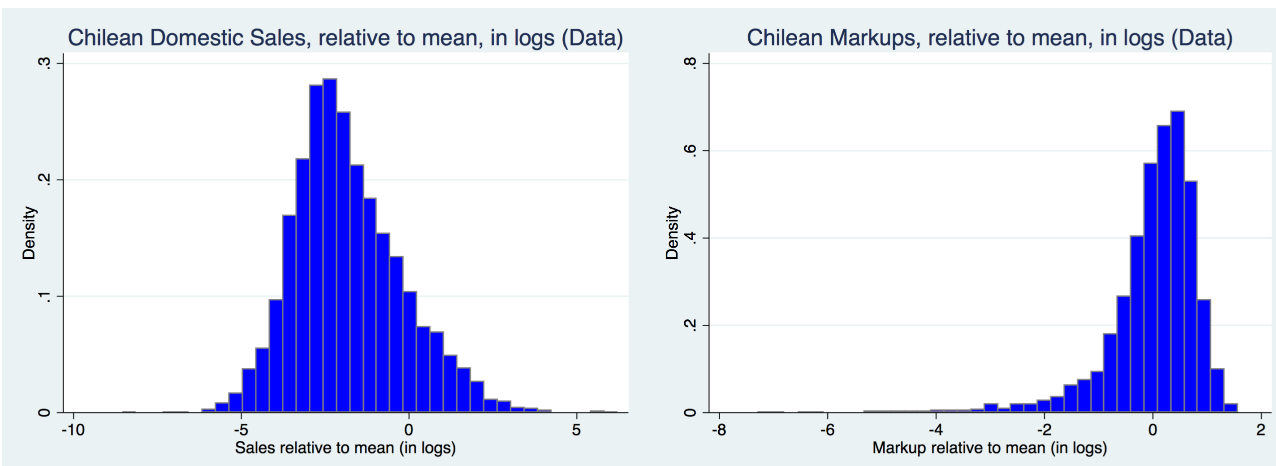
[37]To get at a markup measure that is as close as possible to the model, in the data, we subtract capital expenditure from value added and divide by firm labor compensation. A separate option is to construct markups using the method in De Loecker and Warzynski (2012). In this case markup differences are slightly more compressed.

[38]This is based on data availability in the EIU database, which is a sample of 71 countries. We drop 5 small countries where the predicted trade costs are so large we cannot simulate enough exporters to compute a meaningful maximal price gap. This is because even with 3 million draws, there are too few firms that export to multiple destinations.

[39]For each country, this corresponds to $Y_j = GO_j - X_j + M_j$. This is gross output minus manufacturing exports (to the rest of the world) plus manufacturing imports (in the sample).

Sales and Markups Moments In this section, we present the moments that we use in the estimation. For the exactly-identified case, we use the markup advantage and domestic sales advantage of exporters relative to non-exporters as described in section 4. In 2004, we find that exporters had markups that are 28% higher, and their domestic sales are 4.48 times those of non-exporters. Computing domestic sales in the data is very easy because the value of exports is reported. However, the construction of the markup advantage moment in the data is slightly different, but aims to reproduce the model as much as possible. Wages in the model are not firm-specific, but in the data bigger firms pay higher wages. Therefore, we multiply the firm value added in the data by the ratio of average wages to the firm wage (where the wage is total compensation of employees divided by number of employees). We then divide by the number of employees to create a firm specific $\frac{va_i(s)}{emp_i(s)}$ as in the model.

Figure 1: Chilean Sales and Markup Distribution



The plots report pdfs of log domestic sales and markups relative to the mean in the Chilean data. We use the cross-section of firms in 2004 and eliminate the Basic Metal industry as in Weinberger (2017).

Figure 1 plots the distribution of domestic sales and markups of the 2004 Chilean census of firms. Notice that the sales distribution appears Pareto in the right tail, as has been established in other firm-level datasets. As discussed in Sager and Timoshenko (2017), there is also a fat *left* tail. This is consistent with our model, which will generate Pareto tails on *both* ends. The largest firms behave as with CES demand, while the generalized CES can also produce the left side of the distribution that a CES model with fixed costs cannot.[40] The markup distribution has a more negative skew, but our model will not generate this distribution of markups since it does not generate the left tail. However, we will be able to match the fat tail at the right of

[40]See Arkolakis (2010) for a discussion and a solution to this problem in the CES framework.

the distribution.[41] Next, we use moments from these distributions to identify the parameters in our model.

5.2 Estimates

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 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. In the working paper version, Jung et al. (2015), results from which can be found in the Online Appendix, we verify that a demand curve that is linear or log-linear cannot match the sales moment unless there is a large increase in the trade elasticity. The Pareto shape parameter determines the variability in firm productivity, so a lower θ (higher variance) raises the markup advantage of more productive firms.

Panel A of Table 1 displays the estimation results and is split between the estimation that uses firm-level moments (sales and markups from Chilean firm data) and price moments (micro price data and trade shares). In the former case we can match the exporter advantage moments exactly with $\theta = 3.45$ and $\sigma = 1.26$. Once we attempt to match the tails of the sales and markup distributions in the over-identified case, θ decreases and σ increases to match the large dispersion in the firm-level data. Panel B compares the data to the model moments and there is evidence that we cannot fully match the large dispersion. However, the J-statistic computed from the optimal weighting matrix is 3.37 (p-value=0.76), which is not close to rejecting the null that the over-identifying restrictions are valid at 95% confidence level.[42] This is important as it strengthens our claim that the GCES utility with Pareto productivity is a plausible model to match the sales and markup heterogeneity displayed in firm-level data.[43]

The results from targeting firm moments differ from ACDR, who set θ to 5 and find a σ of 2.88. Compare this to our results: if θ were to increase to 5, it would lower the productivity dispersion and raise the sales advantage. However, since we target different moments from ACDR, it is difficult to compare parameter estimates. We view the second set of results, which uses micro-level price data to match the trade elasticity, as an important robustness check. Due to the role of the trade elasticity in computing the welfare gains from trade, we believe that it

[41]Mrazova et al. (2015) have shown that CREMR demand with a log-normal distribution of productivity will generate a "Log-Normal-Odds" distribution of markups. However, this would also generate a Log-Normal distribution of sales.

[42]The over-identified case is done using an optimal weighting matrix (Gourieroux and Monfort (1997)).

[43]We note that the restrictive non-homothetic models would perform extremely poorly in this exercise, as they cannot match the right tail of the sales distribution.

Table 1: Estimates of θ and σ

Panel A: Parameter Estimates

	Firm Moments		Price Moments	
	Exact ID	Over ID	Exact ID ($\sigma = 2$)	Over ID
θ	3.45 (.001)	2.15 (.008)	2.74 (.000)	3.08 (.000)
σ	1.26 (.001)	1.57 (.002)	-	3.98 (.000)
J-stat	-	3.37	-	0.07

Panel B: Data Targets and Simulation Results

	Firm Moments				Price Moments		
	Data	Exact ID	Over ID		Data	Exact ID ($\sigma = 2$)	Over ID
Sales Adv	4.48	4.48	5.70	Max Price Gap	1.6	1.6	1.62
Markup Adv	0.28	0.28	0.38	85th Price Gap	0.62	-	0.35
90-10 Sales	57.62	-	22.22	Cov(gap,wages)	0.12	-	0.09
90-10 Markups	1.51	-	0.46				
90-50 Sales	12.18	-	3.38				
90-50 Markups	0.94	-	0.33				
99-90 Sales	7.24	-	2.19				
99-90 Markups	1.14	-	0.08				

Panel A reports standard errors in parenthesis. Standard errors computed from bootstrap samples (100 samples). For details on the estimation see Section 4. Data in Panel B is based on authors' estimates. Firm moments are computed using the 2004 census of Chilean firms (provided by INE). Price moments are based on micro price reports of the Economist Intelligence Unit (EIU).

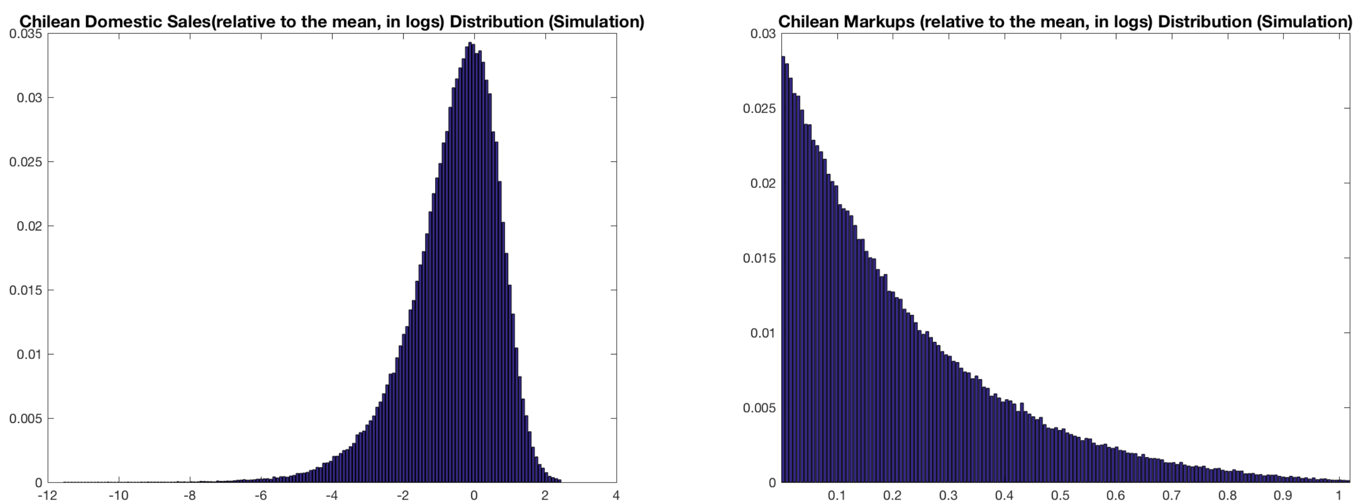
is important to show that our results are robust to matching the trade-elasticity moment. This exercise also provides an extension to the work in ACDR, because the authors take this value as given. In this robustness exercise, we follow Section 4, which describes the three moments used in the over-identified estimation. The result that $\theta = 3.08$ is consistent with both the results from our benchmark firm-level estimations and the findings in Simonovska and Waugh (2014b). The authors find a trade elasticity of 3.15-3.32 in a model with variable markups, albeit with CES preferences. The substitution parameter that we estimate in the robustness exercise is however higher than the estimate from our firm-level results (now $\sigma = 3.98$). For this reason we also check an exactly identified case where we set $\sigma = 2$ and we match the average maximal price gap in the data. In this case, θ is slightly lower (2.74) but still in line with Simonovska and Waugh (2014b).

Finally, as a separate cross-check of the general validity of our results, we use our parameter estimates to simulate US firms and we compare their behavior to US data as reported in Bernard et al. (2003). Using the parameter estimates from the benchmark exactly-identified procedure, the model generates an exporter advantage of sales of 5.4 (about 10% higher than that found in BEJK), and an exporter measured productivity advantage of 0.19 (BEJK report this to be between 0.15 and 0.33, with the smaller number being within 4-digit industries). Notice that the data moments in Chile and the U.S. are fairly similar to each other, and in fact the sales distribution of firms is consistent across most datasets that we are aware of.

5.3 Model's Fit to the Data

This section uses the estimation results to simulate the model and compare model predictions to the data. Section 2.4 derived theoretical results relating to the sales and markup distributions, which have been recently highlighted in Arkolakis et al. (2017). Evaluating the performance with observations from the data is an important contribution to the literature as the quantitative fit of monopolistic competition models with heterogeneous firms and non-homothetic preferences is a direction that has thus far not been explored. We also confirm the consistency of the current model with the existing literature in the prediction of a positive correlation between income per capita and prices of tradable goods.

Figure 2: Calibrated Model Sales and Markup Distribution



Firms are simulated given the calibration results of the over-identified case where we match firm level moments. For given parameter values, domestic sales and markups can be computed for each firm. The plots report pdfs of log sales and markups relative to the mean.

Figure 2 plots the distributions of log sales and markups for the calibrated economy (firm moments with the over-identified procedure). Comparing these to the firm data in Figure 1, we match the fat tails on both sides of the sales distribution, but with less heterogeneity (fewer of the very big firms). As noted earlier, this model will not generate the left tail of the markup distribution seen in the data. Nonetheless, the generalized CES model performs very well along the sales-distribution dimension while generating variable markups, which makes it a desirable framework for trade-policy analysis.

Table 2: Model Predictions

	Data	Firm (Exact)	Firm (Over)	Price (Exact)	Price (Over)
Markup Elas. (ρ)	0.30-0.65 (DW and DGKP)	0.48	0.40	0.35	0.06
Avg Price-Income Elast.	0.14 (SIM 2015)	0.68	0.35	0.30	0.03
% of Exporters	21% (Chile)	4%	4%	4%	4%
	21% (U.S.)	54%	47%	50%	52%
	- (average)	7%	6%	6%	6%
Average Markup	0.18-0.3 (Chile)	0.16	0.26	0.19	0.14
SD of Log Sales	1.7 (Chile)	1.25	1.29	1.34	1.68

Table 2 displays the predictions for various moments in the data that are generated using our estimated model. The first row is the sales-weighted markup elasticity, which we have argued is important because of its place in the gains from trade calculations. In the benchmark estimation this value is between 0.40 and 0.48. For reference, ACDR find a value of 0.36 when they fix the trade elasticity at 5 and calibrate the model to trade data. This parameter has been difficult to pin down in the empirical literature, with a large range depending on the methodology. Two recent papers have relevant cross-sectional results: De Loecker and Warzynski (2012) (DW) find a markup elasticity with respect to productivity of 0.3, while De Loecker et al. (2016) (DGKP) find a *price* elasticity with respect to costs of 0.35 (which corresponds to 0.65 for our estimate). In the alternative estimation we find a smaller markup elasticity which is due to the high substitution parameter. With a $\sigma = 2$, $\rho = 0.35$, but the markup elasticity is only 0.06 when σ rises to 3.98 in order to match the price gap moments.

The second row is our measure of price discrimination: the elasticity of price with respect to income for an average exporter. There is clear price discrimination in this model as this statistic is positive. Although the value is above that found by Simonovska (2015), it is a difficult comparison because that study uses only one exporter (Mango) to compute that statistic.[44] Due to the bound on markups, larger exporters will price discriminate less in our model, and it is expected that a large exporter like Mango would have a price-income elasticity well below the average. The price moment results show that this statistic can be reduced by raising the substitution across varieties as firms have less market power.[45]

[44]We are not aware of another study that computes a similar statistic.

[45]In the Online Appendix, we compare the price-income relationship with what is found in micro price data

The next two moments, the fraction of exporters and the average markup, are straightforward to compute. The average markup is very reasonable, in the range of 13-26% for all specifications. Using Chilean firm data, we find average markups of 18-30% depending on the method used.[46]

The fraction of exporters in the model varies depending on the origin country, as large estimated trade costs will reduce the fraction of exporters. The percent of firms in a country j that export is given by the following equation:

$$\%\ exporters_j = \frac{\tilde{N}^x_{jj}}{N_{jj}} = \left(\frac{\max_{v \neq j} \frac{\bar{c}_v}{\tau_{jv}}}{\bar{c}_j} \right)^{\theta} \tag{33}$$

where the numerator displays the cutoff cost to the easiest export destination of a firm whose origin is country j. In the data, 21% of Chilean firms export, a number similar to what has been found for the U.S. In the model, only 4% of Chilean firms export, but 47-54% of US firms export. It is much easier (a combination of trade costs and demand in the destination) for a U.S. firm to sell in Canada (the easiest destination) than for a Chilean firm to sell in Argentina (Chile's easiest destination). Across all 66 countries in the model, the mean percentage of exporters is 6-7%, with a large variance. This result is driven by the trade costs, which can vary widely when θ is low. When θ is estimated to match moments from the firm distribution, trade costs and cutoffs are generated entirely from gravity (for a given θ). The tight link can be broken if we allow the utility parameter $\bar{q}$ to be country-specific. In this case, we can potentially match each country's fraction of exporters. We are not able to target this moment because it requires the fraction of exporters for every country in our gravity framework, but we hope that the future literature extends the GCES model for quantitative analysis both across sectors and countries.[47]

Finally, 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)). For the GCES model, the value amounts to 1.25, which falls somewhat short of the statistic in Chile of 1.7. However, for a large σ this can be matched

reported by the Economist Intelligence Unit (EIU) and the International Comparison Program (ICP). It is possible to simulate prices in our model for each country, then show the relationship of an average price level and country income. The relationship of prices with destination income is extremely close to what is found in the data. In fact all of the existing non-homothetic models that we summarize in this paper perform reasonably well along this dimension, which is reassuring as it is one of the main goals of this class of models. However, the usefulness of the generalized CES is in the flexibility to get at the sales and markup distributions.

[46]We computed them using price-cost margins from sales and input cost data, and also using the DLW procedure. Markups are computed by sector, and we compute the average using sector fixed effects.

[47]This is a moment matched more closely by the estimation procedure of Behrens et al. (2014) using trade between the United States and Canadian provinces.

almost exactly (last column).[48] Relying on the micro price data requires less price dispersion and higher sales dispersion. The tradeoff of having a large substitution parameter is that the exporter advantage in sales is much bigger than in the data.

The results point to a tradeoff between matching moments from different parts of the price and sales distributions. For a range of the trade elasticity which is stable across our specifications (between 2.15 and 3.45), the ability to match moments from the sales distribution hinges on huge variations in the substitution parameter. A relatively small value (although greater than 1) is necessary to be consistent with moments that relate to exporters versus non-exporters, but a much higher value is needed to match the very fat right tail in sales or the price gaps observed in the micro price data (at least in the EIU sample). We note that we can match either of these facts but not both simultaneously. This hints at the fact pointed out by Sager and Timoshenko (2017) that to match the very fat left and right tails of the sales distribution we need something more flexible than Pareto. They also connect their findings to gains from trade as the trade elasticity depends on the ability to match the sales moments. However, that study does not incorporate the cross-sectional markup distribution as it relies on CES preferences. In our model, it is not only the trade elasticity that hinges on these moments, but also the sales-weighted markup elasticity.

An alternative extension is to incorporate additional sources of price variation into the model. Fan et al. (2017) offer one such extension of the present model, where firms not only price discriminate across destinations, but they also offer products of different qualities. We believe that reconciling these aspects of the sales and price distributions is an important avenue for future research.

We also compute the export intensity of Chilean exporters as in Bernard et al. (2003). To evaluate the models' predictions along this dimension, for Chilean exporters, we compute the fraction of total firm sales that are exported and refer to this object as 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, we measure the percentage of exporters that fall into certain ranges of export intensity. In Table 3, we report the moments in the data and the model. 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 GCES model, 97-99% of exporting firms have export revenue that is less than 10% of their total revenue. The results are similar for each estimation exercise, where the

[48] 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.

model fails to match the fraction of firms that export a large percentage of their sales. There are a couple ways to improve the fit in this regard. First, relaxing the assumption of identical cross-country variety parameters, which would raise the predicted percent of Chilean firms that export, would improve the fit in this aspect. Second, the model does not attempt to explain the observed behavior of exporters that do not sell domestically. In the data, it is evident that there are exporters that sell either all or the majority of their sales abroad (last row of second column).[49]

Table 3: % of Exporting Plants Conditional on Export Intensity

Exp. Intensity (%)	Data (Chile, %)	Firm (Exact)	Firm (Over)	Price (Exact)	Price (Over)
0-10	42.7	97.1	97.3	97.3	98.7
10-20	12.1	0.5	0.6	0.6	1.2
20-30	7.3	0.5	0.6	0.8	0.1
30-40	5.6	0.6	0.7	0.8	0.0
40-50	4.8	0.6	0.5	0.2	0.0
50-60	6.0	0.4	0.1	0.1	0.0
60-100	21.5	0.3	0.3	0.2	0.0

In this table, 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 GCES model, 97-99% of exporting firms have export revenue that is less than 10% of their total revenue.

6 Conclusion

In this paper, we quantify a flexible general equilibrium model of international trade and pricing-to-market that features firm-level heterogeneity and consumers with non-homothetic preferences. We argue that the model's flexibility to match aspects of the firm sales and markup distributions is key because it allows one to precisely estimate the parameters necessary in order to quantify the predicted welfare gains from trade. The quantification exercise highlights the importance of different models' predictions regarding the firm sales distribution. In particular, we find that the parameter that governs the demand curvature in these models is essential to match the observed heterogeneity in sales. Future work is needed to reconcile the disconnect between mean and dispersion moments of the sales distribution, and there is already progress on this front. Work on more general productivity distributions — notably the log-normal (Bas et al., 2017; Fernandes et al., 2017) and double EMG (Sager and Timo-

[49] We are only aware of Behrens et al. (2014) as a type of model that allows for this behavior, as they incorporate internal trade costs. We are indebted to an anonymous referee for bringing to light these ways in which export intensity fails in our model.

shenko, 2017) — combined with CES preferences, allow for flexibility in sales distributions.[50] However, by construction markups are constant in these papers. We view the combination of flexibility, non-homotheticity, and tractability in our framework as an attractive feature. An avenue for future research would be to connect more flexible productivity distributions with non-homothetic preferences.

References

Adda, J. and R. W. Cooper (2003): *Dynamic Economics: Quantitative Methods and Applications*, MIT Press.

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 (2017): "The Elusive Pro-Competitive Effects of Trade," *Review of Economic Studies*, forthcoming.

Bas, M., T. Mayer, and M. Thoenig (2017): "From micro to macro: Demand, supply, and heterogeneity in the trade elasticity," *Journal of International Economics*, 108, 1–19.

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

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

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., F. Etro, and I. Simonovska (2018): "International Trade with Indirect Additivity," *American Economic Journal: Microeconomics*, 10, 1–59.

[50]However, the log-normal would not improve upon the tradeoff of average versus largest firms in our model. The double EMG offers a potential solution.

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

De Loecker, J., P. Goldberg, A. Khandelwal, and N. Pavcnik (2016): "Prices, Markups and Trade Reform," *Econometrica*, 84, 445–510.

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

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 (2018): "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," *Journal of Political Economy*, forthcoming.

Dominicy, Y. and D. Veredas (2013): "The Method of Simulated Quantiles," *Journal of Econometrics*, 172, 235–247.

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

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

Edmond, C., V. Midrigan, and D. Xu (2015): "Competition, Markups and the Gains from International Trade," *American Economic Review*, 105, 3183–3221.

Fan, H., Y. A. Lee, S. Xu, and S. R. Yeaple (2017): "Quality, Variable Markups, and Welfare: A Quantitative General Equilibrium Analysis of Export Prices," *Pennsylvania State University, unpublished manuscript.*

Feenstra, R. C. (2018): "Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," *Journal of International Economics*, 110, 16–27.

Feenstra, R. C. and D. E. Weinstein (2017): "Globalization, Markups and U.S. Welfare," *Journal of Political Economy*, 125, 1040–1074.

Fernandes, A., P. Klenow, S. Meleschuk, M. D. Pierola, and A. Rodriguez-Clare (2017): "The intensive Margin in Trade," Working paper.

Gourieroux, C. and A. Monfort (1997): *Simulation-based Econometric Methods*, OUP Catalogue, Oxford University Press.

Jung, J.-W., I. Simonovska, and A. Weinberger (2015): "Exporter Heterogeneity and Price Discrimination: A Quantitative View," NBER Working Papers 21408.

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.

Mrazova, M., P. J. Neary, and M. Parenti (2015): "Technology, Demand, and the Size Distribution of Firms," Tech. rep.

Sager, E. and O. Timoshenko (2017): "The EMG Distribution and Trade Elasticities," Working paper.

Simonovska, I. (2015): "Income Differences and Prices of Tradables: Insights From An Online Retailer," *Review of Economic Studies*, 82, 1612–1656.

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. (2017): "Markups and Misallocation with Evidence from an Exchange Rate Appreciation," FRB Dallas Globalization and Monetary Policy Institute Working Paper 302.

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:

$$(1-\sigma)p_{ij}^{-\sigma}L_j\frac{w_j+\bar{q}P_j}{P_{j\sigma}^{1-\sigma}}-L_j\bar{q}+\sigma c_{ij}p_{ij}^{-\sigma-1}L_j\frac{w_j+\bar{q}P_j}{P_{j\sigma}^{1-\sigma}}=0 \tag{34}$$

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:

$$F(p_{ij},c_{ij})=(1-\sigma)p_{ij}+\sigma c_{ij}-p_{ij}^{\sigma+1}(\bar{c}_j)^{-\sigma} \tag{35}$$

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 both sides of expression (35) by $\left(m_{ij}^{\sigma+1}c_{ij}\right)$ yields the implicit function for mark-ups and costs as:

$$G(m_{ij},c_{ij})=(1-\sigma)m_{ij}^{-\sigma}+\sigma m_{ij}^{-(\sigma+1)}-c_{ij}^{\sigma}(\bar{c}_j)^{-\sigma} \tag{36}$$

By the implicit function theorem,

$$\frac{dm_{ij}}{dc_{ij}}=\frac{c_{ij}^{\sigma-1}(\bar{c}_j)^{-\sigma}}{(\sigma-1)m_{ij}^{-\sigma-1}-(\sigma+1)m_{ij}^{-\sigma-2}} \tag{37}$$

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. Therefore, the range for mark-ups is $[1, \sigma/(\sigma - 1)]$.

Finally, profits are given by:

$$\begin{aligned} \pi_{ij}(c_{ij}) &= (p_{ij}(c_{ij}) - c_{ij})\, x_{ij}(c_{ij}) \\ &= L_j \bar{q} \left((\bar{c}_j)^\sigma (p_{ij}(c_{ij}))^{1-\sigma} - p_{ij}(c_{ij}) - (\bar{c}_j)^\sigma c_{ij} (p_{ij}(c_{ij}))^{-\sigma} + c_{ij} \right) \end{aligned} \tag{38}$$

A.2 Equilibrium

We use the change of variables 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 normalized by $\bar{c}_j$. Aggregate and normalized sales are given by the following two expressions:

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

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 $dc_{ij} = \frac{\bar{c}_j}{\sigma}((\sigma + 1) t_{ij}^\sigma + (\sigma - 1)) dt_{ij}$.

Let π_{ij} denote average profits of firms from i in destination j. Then, normalized profits are given by:

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

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

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

is the hypergeometric function with vector (α,β), scalar δ, and element z evaluated at the two endpoints. In practice, we either solve the integrals in Matlab using the simulated data, or compute a closed form solution for the constants using the *hypergeom* function in Matlab.[51]

The aggregate price statistics are:

$$P_j = \sum_{\upsilon}^{I} J_\upsilon \bar{c}_j \left(\frac{\bar{c}_j}{\hat{c}_{\upsilon j}}\right)^{\theta} \frac{\theta}{\sigma^\theta}\int_0^1 \left[t_{\upsilon j}^{\sigma+1}+(\sigma-1)t_{\upsilon j}\right]^{\theta-1}\left[(\sigma+1)t_{\upsilon j}^{\sigma+1}+(\sigma-1)t_{\upsilon j}\right]dt_{\upsilon j}$$
$$= \sum_{\upsilon}^{I} J_\upsilon \frac{\bar{c}_j^{\theta+1} b_\upsilon^\theta}{(\tau_{\upsilon j} w_\upsilon)^\theta}\beta_P \tag{39}$$

$$P_{j\sigma}^{1-\sigma} = \sum_{\upsilon}^{I} J_\upsilon \frac{\bar{c}^{\theta+1-\sigma}}{\hat{c}_{\upsilon j}^{\theta}} \frac{\theta}{\sigma^\theta}\int_0^1 \left[t_{\upsilon j}^{\sigma+1}+(\sigma-1)t_{\upsilon j}\right]^{\theta-1}\left[(\sigma+1)t_{\upsilon j}+(\sigma-1)t_{\upsilon j}^{1-\sigma}\right]dt_{\upsilon j}$$
$$= \sum_{\upsilon}^{I} J_\upsilon \frac{\bar{c}^{\theta+1-\sigma} b_\upsilon^\theta}{(\tau_{\upsilon j} w_\upsilon)^\theta}\beta_{\sigma P} \tag{40}$$

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 the equilibrium measure of entrants is:

$$J_i = \frac{\Gamma_1}{\Gamma_2}\frac{L_i}{f_e} \tag{41}$$

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 $\Gamma_2 = \frac{\theta}{\sigma^\theta}\left(\frac{1}{\sigma-1}\right)^{1-\theta}\beta_2$. Finally, in the main text we have $\beta_E = \frac{\Gamma_1}{\Gamma_2}$.

Lastly we obtain the expression for the cutoff costs (20) by substituting the aggregate price

[51] In Matlab, 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. We checked that using the *hypergeom* and solving the integrals results in the same values.

statistics into (6):

$$\bar{c}_j = \left[\frac{w_j}{(\beta_{\sigma P} - \beta_P)\bar{q} \sum_{\upsilon}^{I} J_{\upsilon} \frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j} w_{\upsilon})^{\theta}}} \right]^{\frac{1}{\theta+1}} \tag{42}$$

In the main text we substitute for J_{υ} 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 we derive using expression (20):

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

We need to verify that the second term is positive:

$$\begin{aligned}
\frac{\partial \bar{c}_j}{\partial w_j} = \frac{1}{\theta+1} & \left[\frac{w_j}{(\beta_{\sigma P} - \beta_P)\beta_E \bar{q} f_e^{-1} \sum_{\upsilon=1}^{I} L_{\upsilon} \frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j} w_{\upsilon})^{\theta}}} \right]^{\frac{1}{\theta+1}-1} \times \\
& \left\{ \frac{1}{(\beta_{\sigma P} - \beta_P)\beta_E \bar{q} f_e^{-1} \sum_{\upsilon=1}^{I} L_{\upsilon} \frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j} w_{\upsilon})^{\theta}}} - \frac{w_j (\beta_{\sigma P} - \beta_P)\beta_E \bar{q} f_e^{-1} L_j b_j^{\theta} (-\theta) w_j^{-\theta-1} \tau_{jj}^{-\theta}}{\left[(\beta_{\sigma P} - \beta_P)\beta_E \bar{q} f_e^{-1} \sum_{\upsilon=1}^{I} L_{\upsilon} \frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j} w_{\upsilon})^{\theta}}\right]^2} \right\} \\
= \frac{1}{\theta+1} & \left[\frac{w_j}{(\beta_{\sigma P} - \beta_P)\beta_E \bar{q} f_e^{-1} \sum_{\upsilon=1}^{I} L_{\upsilon} \frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j} w_{\upsilon})^{\theta}}} \right]^{\frac{1}{\theta+1}} \times \left\{ \frac{1}{w_j} + \frac{\theta L_j b_j^{\theta} w_j^{-\theta-1}}{\sum_{\upsilon=1}^{I} L_{\upsilon} \frac{b_{\upsilon}^{\theta}}{(\tau_{\upsilon j} w_{\upsilon})^{\theta}}} \right\} \\
= \frac{1}{\theta+1} & \bar{c}_j \frac{1+\theta\lambda_{jj}}{w_j} > 0
\end{aligned}$$

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

Next, we derive moments from the distribution of price-income elasticities. To do so, write the price as markup times cost and then apply the change of variables from above before differentiating with respect to income. For a producer from i selling to j:

$$\frac{\partial\left(\frac{p_{ij}}{p_{jj}}\right)}{\partial\left(\frac{w_i}{w_j}\right)}\frac{\frac{w_i}{w_j}}{\frac{p_{ij}}{p_{jj}}}=\frac{\partial ln\ \left(\frac{p_{ij}}{p_{jj}}\right)}{\partial ln\ \left(\frac{w_i}{w_j}\right)}=\frac{\partial\left(ln\ p_{ij}-ln\ p_{jj}\right)}{\partial ln\ \left(\frac{w_i}{w_j}\right)}=\frac{\partial\left(ln\ m_{ij}+ln\ c_{ij}-\left(ln\ m_{ii}+ln\ c_{ii}\right)\right)}{\partial ln\ \left(\frac{w_i}{w_j}\right)}$$

$$=\left(\frac{-\sigma}{(\sigma+1)+(\sigma-1)\,t_{ij}^{-\sigma}}\right)\left(-\frac{1+\theta\lambda_{jj}}{\theta+1}\right)-\left(\frac{-\sigma}{(\sigma+1)+(\sigma-1)\,t_{ii}^{-\sigma}}\right)\left(-\frac{\theta}{\theta+1}\lambda_{ji}\right), \tag{43}$$

where $t_{ij}\equiv p_{ij}/\bar{c}_j$, and we also used the fact that $v_{ij}\equiv c_{ij}/\bar{c}_j$, and $m_{ij}=t_{ij}/v_{ij}$. From above we have that $\frac{\partial\ln(\bar{c}_j)}{\partial\ln(w_j)}=\left(\frac{1+\theta\lambda_{jj}}{\theta+1}\right)$.

Finally, the average price income elasticity is:

$$\epsilon^*_{p-w}=-\frac{\theta}{\left(\tilde{c}^x_{ii}\sigma\right)^{\theta}}\int_0^{\tilde{t_{ii}}^x}\sum_{i\neq j}\delta_{ij}(s)\Bigg[\left(\frac{\sigma}{(\sigma+1)+(\sigma-1)\left(h_{ij}(t_{ii}(s))\right)^{-\sigma}}\right)\left(\frac{1+\theta\lambda_{jj}}{\theta+1}\right)$$
$$-\left(\frac{\sigma}{(\sigma+1)+(\sigma-1)\,t_{ii}(s)^{-\sigma}}\right)\left(\frac{\theta}{\theta+1}\lambda_{ji}\right)\Bigg]/\sum_i\delta_{ij}(s)$$
$$\left(t_{ii}(s)^{\sigma+1}+(\sigma-1)t_{ii}(s)\right)^{\theta-1}\left((\sigma+1)t_{ii}(s)^{\sigma}+(\sigma-1)\right)dt_{ii}(s)$$

where $\delta_{ij}(s)$ is an indicator that takes the value of one if firm s from i sells in market j, $h_{ij}(t_{ii}(s))$ is an implicit solution of $h_{ij}(t_{ii}(s))=t_{ij}(s)$, $\tilde{c}^x_{ii}$ is the cutoff cost of the marginal exporter s^x from source country i, and $\tilde{t}^x_{ii}$ is the corresponding $t_{ii}(s^x)$. This value is the average of the simple mean of price-income elasticity of all exporters across only export destinations and we compute it using simulated firms as described in Section 4.

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 $\beta^{NX}_{MP,i}(\sigma,\theta,\tilde{t}_{ii})$, which is given by:

$$\beta^{NX}_{MP,i}(\sigma,\theta,\tilde{t}_{ii}) = \frac{1}{(\sigma-1)\theta(\sigma+\theta)}\times$$
$$\left[\sigma^{\theta}\left(\frac{1}{\sigma-1}+1\right)^{-\theta}\left\{(\sigma\theta\log\sigma+\theta\log\sigma+\sigma)\ {}_2F_1\left(1-\theta,\frac{\sigma+\theta}{\sigma};\frac{\theta}{\sigma}+2;\frac{1}{1-\sigma}\right)-(\sigma-1)(\sigma+\theta)\times\right.\right.$$
$$\left.\left(-\log\sigma\ {}_2F_1\left(1-\theta,\frac{\theta}{\sigma};\frac{\sigma+\theta}{\sigma};\frac{1}{1-\sigma}\right)+\left(\frac{1}{\sigma-1}+1\right)^{\theta}\log\sigma\right)\right\}$$
$$-\left(\tilde{t}_{ii}\left(\tilde{t}^{\sigma}_{ii}+\sigma-1\right)\right)^{\theta}\left(\frac{\tilde{t}^{\sigma}_{ii}}{\sigma-1}+1\right)^{-\theta}\left\{\tilde{t}^{\sigma}_{ii}\left(\sigma\theta\log\sigma+\theta\log\sigma+\sigma\right){}_2F_1\left(1-\theta,\frac{\sigma+\theta}{\sigma};\frac{\theta}{\sigma}+2;-\frac{\tilde{t}^{\sigma}_{ii}}{\sigma-1}\right)\right.$$
$$\left.\left.-(\sigma-1)(\sigma+\theta)\left((-\log\sigma)\ {}_2F_1\left(1-\theta,\frac{\theta}{\sigma};\frac{\sigma+\theta}{\sigma};-\frac{\tilde{t}^{\sigma}_{ii}}{\sigma-1}\right)+\left(\frac{\tilde{t}^{\sigma}_{ii}}{\sigma-1}+1\right)^{\theta}\log\left(\tilde{t}^{\sigma}_{ii}+\sigma-1\right)\right)\right\}\right]$$

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

For exporters, we also skip over the integration in the main text and show the full solution here. 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 all exporters yields the average exporter logged productivity:

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

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

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

$$
\begin{aligned}
M^m_{i,sales} = (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]
\end{aligned}
$$

A.5 Alternative Calibration: Price Moments

Motivation of the Trade Elasticity Moments We show that the maximal price gap continues to be informative about trade costs as in Simonovska and Waugh (2014b), but we add further moments as required by the non-homothetic model to identify θ. For analytical convenience, we begin by looking at the symmetric 2-country case. This allows us to derive logged price differences with subscripts denoting country 1 and 2, respectively:

$$
\log(p_{12}(c)) - \log(p_{11}(c)) = \log(\tau_{12}) + \log\left[\frac{(\sigma-1)-p_{11}^\sigma}{(\sigma-1)-p_{12}^\sigma}\right]. \tag{44}
$$

When wages and trade shares are equal in the two countries, price differences depend on trade costs and σ.

If countries are asymmetric, cost cutoff differences enter price differences, which implies that gravity-estimated objects as well as wages also drive these price differences. A closed-form solution for the GCES model is not available in this case. However, to gain some insight into the asymmetric multi-country case, we examine the special case where $\sigma \to 1$, as in Simonovska (2015), which is an additive model that yields an identical gravity equation of trade as the GCES model. Focusing on this model, if countries i and j source the same product from country k, their logged price difference is given by:

$$\log(p_{kj}(c)) - \log(p_{kn}(c)) = \frac{2\theta+1}{2(\theta+1)}\left[\log(\tau_{kj}) - \log(\tau_{kn})\right] + \frac{1}{2(\theta+1)}\left[\log(\lambda_{kj}) - \log(\lambda_{kn})\right] + \frac{1}{2(\theta+1)}\left[\log(w_j) - \log(w_n)\right] \quad (45)$$

As is obvious in this case, implied logged trade cost differences are linear in logged price gaps, wages and trade shares. Similarly, in the GCES model, wages and trade shares in the two countries, alongside prices, contain information about trade costs; the relationship is simply non-linear. This suggests that, in order to approximate trade costs in the non-homothetic model, we should use data not only on price differences, but also trade share and wage/per-capita income differences, since part of the price variation is driven by the last two variables.

Description of EIU Data and its Model Equivalent 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.

We follow a similar strategy to generate prices of "common goods" in the model, which are then used to compute maximal price gaps. We proceed by following a simulation and sampling methodology introduced by Simonovska and Waugh (2014a), which aims to replicate steps taken to construct the EIU database. In particular, we construct a set of "common" goods and then we draw 100 random samples of 110 products from the set. 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. (2011) 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.

B Details and Moments in the SIM Model

To solve the model, we follow the same steps as for the general model. We calculate total bilateral sales and pin down the maximum cost allowed in order to serve market j using the cost of the firm with zero demand:

$$\bar{c}_j \equiv \bar{c}_j^{SIM} = \left[f_e \beta_E^{SIM} \bar{q}^{-1}\right]^{\frac{1}{\theta+1}} \left[\frac{w_j}{\sum_{\upsilon=1}^{I} \frac{L_\upsilon b_\upsilon^\theta}{(\tau_{\upsilon j} w_\upsilon)^\theta}}\right]^{\frac{1}{\theta+1}}$$

where $\beta_E^{SIM} = \frac{(1+2\theta)(\theta+1)^2}{\theta}$. 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). The resulting price, mark-up, and sales are:

$$p_{ij}(c_{ij}) = (c_{ij}\bar{c}_j)^{\frac{1}{2}} \tag{46}$$

$$m_{ij}(c_{ij}) = \left(\frac{\bar{c}_j}{c_{ij}}\right)^{\frac{1}{2}} \tag{47}$$

$$r_{ij}\left(c_{ij}\right) = L_j \bar{q}\left(\bar{c}_j - (\bar{c}_j c_{ij})^{1/2}\right). \tag{48}$$

These are used to solve for total sales:

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

The moments below can be used to estimate the value of θ. We present the two firm-level moments that we derived for the GCES model. 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:

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

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

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

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

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 the statistic is:

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

The moment equations above and the discussion in the text lead to the observation that, in the SIM model, there is only one parameter, θ, that governs both the measured productivity and sales advantage. Given the quantitative results from Section 5, where the estimate of σ is not unity, it is clear that there does not exist a value for θ such that the SIM model can match both moments in the data.

C Details and Moments in the Separable MO Model

This specification corresponds to the MO model where $\eta = 0$. To solve the model we derive the cost cutoffs, which are:

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

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

The price, mark-up, and sales are:

$$p_{ij}(c_{ij}) = \frac{1}{2}\left(c_{ij} + \bar{c}_j\right) \tag{52}$$

$$m_{ij}(c_{ij}) = \frac{1}{2}\left(\frac{c_{ij} + \bar{c}_j}{c_{ij}}\right) \tag{53}$$

$$r_{ij}(c_{ij}) = \frac{L_j}{4\bar{c}_j\gamma}\left(\bar{c}_j^2 - c_{ij}^2\right). \tag{54}$$

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

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

Then, define a new variable $v_{ij}(s) \equiv \frac{c_{ij}(s)}{\bar{c}_j}$ such that:

$$c_{ij}(s) = v_{ij}(s)\bar{c}_j \Longrightarrow dc_{ij}(s) = \bar{c}_j dv_{ij}(s) \tag{55}$$

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.

The markup advantage of exporters, equivalent to the measured productivity advantage in BEJK, requires the calculation of total value added divided by employment. The value added for each s is:

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

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

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

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:

$$\begin{aligned} MP_i^{EXP} = \frac{\theta}{(\xi_i)^\theta} \int_0^{\upsilon_{ii}^x} \Bigg[\log\left\{ \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \frac{1}{2} \frac{(\theta+2) w_\upsilon}{\sum_{j=1}^{I} L_j \left(w_j \tau_{j\upsilon}\right)^{-\theta} \bar{c}_\upsilon^\theta} \left(1 - (\tau_{i\upsilon} \upsilon_{ii}(s))^2\right) \right\} \\ - \log\left\{ \sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \frac{(\theta+2) w_\upsilon}{\sum_{j=1}^{I} L_j \left(w_j \tau_{j\upsilon}\right)^{-\theta} \bar{c}_\upsilon^\theta} \tau_{i\upsilon} \upsilon_{ii}(s)(1 - \tau_{i\upsilon}\upsilon_{ii}(s))/w_i \right\} \Bigg] \upsilon_{ii}(s)^{\theta-1} d\upsilon_{ii}(s) \end{aligned}$$

The average logged measured productivity of non-exporters is:

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

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

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

$$M_{i,sales}^{MO,\eta=0} = \left(\frac{1 - \xi_i^\theta}{\xi_i^\theta}\right) \frac{(\upsilon_{ii}^x)^\theta \left(\frac{1}{\theta} - \frac{(\upsilon_{ii}^x)^2}{\theta+2}\right)}{\frac{2}{\theta(\theta+2)} - \left((\upsilon_{ii}^x)^\theta \left(\frac{1}{\theta} - \frac{(\upsilon_{ii}^x)^2}{\theta+2}\right)\right)}. \tag{56}$$

The model shares the same limitations as the SIM model. Quantitative results can be found in the Online Appendix.

D Details and Moments in the BMMS Model

Given the Lambert function and change of variables described in the main text, $c_{ij} = \bar{c}_j z_{ij} e^{z_{ij}-1}$ 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 variables and solving for the free entry measure of entrants we obtain cost cutoffs:

$$\bar{c}_j \equiv \bar{c}_j^{BMMS} = \left[\frac{\alpha f_e}{\kappa_2(\theta)}\right]^{\frac{1}{\theta+1}} \left[\frac{w_j}{\sum_{v=1}^{I} \frac{L_v b_v^\theta}{(\tau_{vj} w_v)^\theta}}\right]^{\frac{1}{\theta+1}},$$

where $\kappa_2(\theta) = \theta e^{-(\theta+1)} \int_0^1 (z^{-1} - z - 1 + z^2)(ze^z)^\theta e^z dz$. Clearly, the cost cutoff (modulo proportionality constant) is identical to the one for the generalized model from expression (20).

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

$$p_{ij}(c_{ij}) = \frac{c_{ij}}{W\left(\frac{c_{ij}}{\bar{c}_j} e\right)} \tag{57}$$

$$m_{ij}(c_{ij}) = \frac{1}{W\left(\frac{c_{ij}}{\bar{c}_j} e\right)} \tag{58}$$

$$r_{ij}\left(c_{ij}\right) = \frac{L_j}{\alpha} c_{ij} \frac{1 - W\left(\frac{c_{ij}}{\bar{c}_j} e\right)}{W\left(\frac{c_{ij}}{\bar{c}_j} e\right)}, \tag{59}$$

As in the generalized CES model, per-capita income is positively associated with prices and mark-ups (see Behrens et al. (2014) for derivations).

Total sales are given by:

$$T_{ij} = \frac{L_j}{\alpha} L_i f_e^{-1} \kappa_2(\theta) \left(\bar{c}_j\right)^{\theta+1} \hat{c}_{ij}^{-\theta}, \text{ where } \kappa_2(\theta) = \theta e^{-(\theta+1)} \int_0^1 (z^{-1} - z - 1 + z^2)(ze^z)^\theta e^z dz,$$

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:

$$va_i(s) = \sum_{v=1}^{I} \delta_{iv}(s) \frac{L_v}{\alpha} \bar{c}_v (z_{iv}(s) e^{z_{iv}(s)-1}) \left[z_{iv}(s)^{-1} - 1\right]$$

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

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

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:

$$\begin{aligned} MP_i^{EXP} = \frac{\theta}{(\xi_i)^\theta} \int_0^{\tilde{z}_{ii}} \Bigg[& \log\left\{\sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \bar{c}_\upsilon \left(\tilde{h}_{i\upsilon}(z_{ii}(s))^{-1} - 1\right)\left(\tilde{h}_{i\upsilon}(z_{ii}(s)) e^{\tilde{h}_{i\upsilon}(z_{ii}(s))-1}\right)\right\} \\ & - \log\left\{\frac{1}{w_i}\sum_{\upsilon=1}^{I} \delta_{i\upsilon}(s) L_\upsilon \bar{c}_\upsilon \left(1 - \tilde{h}_{i\upsilon}(z_{ii}(s))\right)\left(\tilde{h}_{i\upsilon}(z_{ii}(s)) e^{\tilde{h}_{i\upsilon}(z_{ii}(s))-1}\right)\right\}\Bigg] \\ & \left(z_{ii}(s) e^{z_{ii}(s)-1}\right)^{\theta-1} e^{z_{ii}(s)-1}(1 + z_{ii}(s)) dz_{ii}(s) \end{aligned} \quad (60)$$

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:

$$MP_i^{NX} = \frac{\theta}{1-\xi_i^\theta} \int_{\tilde{z}_{ii}}^{1} \left[\log(w_i) - \log(z_{ii}(s))\right] \left(z_{ii}(s) e^{z_{ii}(s)-1}\right)^{\theta-1} e^{z_{ii}(s)-1}(1 + z_{ii}(s)) dz_{ii}(s) \quad (61)$$

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

The domestic sales advantage of exporters in the model is:

$$M^{BMMS}_{i,sales} = \frac{1-\xi_i^\theta}{\xi_i^\theta} \frac{\int_0^{\tilde{z}_{ii}} (z_{ii}(s)^{-1} - z_{ii}(s))(z_{ii}(s)e^{z_{ii}(s)})^\theta e^{z_{ii}(s)} dz_{ii}(s)}{\int_{\tilde{z}_{ii}}^1 (z_{ii}(s)^{-1} - z_{ii}(s))(z_{ii}(s)e^{z_{ii}(s)})^\theta e^{z_{ii}(s)} dz_{ii}(s)} \quad (62)$$

The model shares the same limitations as the SIM model. Quantitative results can be found in the Online Appendix.

E Markup Elasticities in Existing Models

The average markup in each model is:

$$\bar{m}^{SIM} = \frac{\theta}{\theta - 0.5}$$
$$\bar{m}^{MO} = \frac{\theta - 0.5}{\theta - 1}$$
$$\bar{m}^{BMMS} = \kappa_1(\theta)$$

As with the generalized CES model, we can derive the ρ parameter in ACDR. The following equations summarize the parameter for the three existing variable markup models:

$$\rho^{SIM} = 0.5 \tag{63}$$
$$\rho^{MO} = \frac{\theta + 2}{2(\theta + 1)}$$
$$\rho^{BMMS} = \frac{\int_0^1 (\frac{1}{z} - 1)(ze^z)^\theta e^z dz}{\int_0^1 (\frac{1}{z} - z)(ze^z)^\theta e^z dz},$$

where z, as defined above, is in the range $(0, 1)$.[52] Notice that the lower bound for the markup elasticity is one-half in all three models. It is constant and always equal to one-half in the SIM model. In the MO and BMMS models, the most productive firms have a price pass-through of zero since their markups go to infinity, so ρ is in the range of $(0.5, 1)$. There is a clear differentiation with the generalized CES model where $\rho \in (0, 0.5)$.

F MO Model in General Equilibrium

F.1 Linear Demand with Non-Separable Preferences

We show that this model exhibits some of the flexibility of the generalized CES model in that the two moments of interest do not rely solely on the Pareto shape parameter. While the model generates identical functional forms for the distributions of firm sales and mark-ups as its separable 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

[52] For costs that approach 0, $z \equiv W(0) = 0$ and as costs approach the cutoff, $W(e) = 1$.

and a value of η can reconcile the measured markup 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, in order for the model to match the moments in the Chilean 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.

F.2 Model

The framework comes from Melitz and Ottaviano (2008), with an extension to general equilibrium outlined in the Web Appendix of Simonovska (2015). Assume that 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:

$$p_{ij}(c_{ij}) = \frac{1}{2}\left(c_{ij} + \bar{c}_j\right) \tag{64}$$

$$m_{ij}(c_{ij}) = \frac{1}{2}\left(\frac{c_{ij} + \bar{c}_j}{c_{ij}}\right) \tag{65}$$

$$r_{ij}\left(c_{ij}\right) = \frac{L_j(\theta + 1)}{2\bar{c}_{ij}\left(2\gamma(\theta + 1) + \eta N_j\right)}\left(\bar{c}_{ij}^2 - c_{ij}^2\right) \tag{66}$$

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:

$$c_{ij} = v_{ij}\bar{c}_j \Longrightarrow dc_{ij} = \bar{c}_j dv_{ij} \tag{67}$$

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

$$T_{ij} = \frac{J_i L_j \bar{c}_j^{\theta+1}(\theta + 1)}{\hat{c}_{ij}^{\theta}\left(2\gamma(\theta + 1) + \eta N_j\right)(\theta + 2)} \tag{68}$$

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:

$$\frac{\frac{\theta+1}{\theta+2}\tilde{c}_j^{\theta+1} - w_j \tilde{c}_j^{\theta}}{\frac{\theta+1}{\theta+2}\tilde{c}_k(K)^{\theta+1} - w_k \tilde{c}_k(K)^{\theta}} = \frac{w_j}{w_k}\frac{\sum_v^I L_v b_v^{\theta}\left(w_v \tau_{vk}\right)^{-\theta}}{\sum_v^I L_v b_v^{\theta}\left(w_v \tau_{vj}\right)^{-\theta}}, \tag{69}$$

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

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

For exporters the average logged measured productivity is:

$$MP_i^{EXP} = \frac{\theta}{(\xi_i(K))^\theta}\int_0^{v_{ii}^x(K)}\left[\log\left\{\sum_{v=1}^{I}\delta_{iv}(s)L_v\frac{1}{2}\frac{w_v}{\sum_{j=1}^{I}L_j\left(w_j\tau_{jv}\right)^{-\theta}\bar{c}_v(K)^\theta}(1-(\tau_{iv}v_{ii}(s))^2)\right\}\right.$$
$$\left.-\log\left\{\sum_{v=1}^{I}\delta_{iv}(s)L_v\frac{w_v}{\sum_{j=1}^{I}L_j\left(w_j\tau_{jv}\right)^{-\theta}\bar{c}_v(K)^\theta}\tau_{iv}v_{ii}(s)\frac{(1-\tau_{iv}v_{ii}(s))}{w_i}\right\}\right]v_{ii}(s)^{\theta-1}dv_{ii}(s) \qquad (70)$$

The average logged measured productivity of non-exporters is:

$$MP_i^{NX} = \frac{\theta}{1-(\xi_i(K))^\theta}\int_{v_{ii}^x(K)}^{1}\left[\log\left(\frac{1}{2}\right)+\log\left((1-v_{ii}(s)^2)\right)-\log\left(v_{ii}(s)(1-v_{ii}(s))\right)+\log(w_i)\right]$$
$$\left(v_{ii}(s)^{\theta-1}\right)dv_{ii}(s) \qquad (71)$$

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

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

$$M_{i,sales}^{MO} = \left(\frac{1-(\xi_i(K))^\theta}{(\xi_i(K))^\theta}\right)\frac{(v_{ii}^x(K))^\theta\left(\frac{1}{\theta}-\frac{(v_{ii}^x(K))^2}{\theta+2}\right)}{\frac{2}{\theta(\theta+2)}-\left((v_{ii}^x(K))^\theta\left(\frac{1}{\theta}-\frac{(v_{ii}^x(K))^2}{\theta+2}\right)\right)} \qquad (72)$$

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.

G Quantitative Analysis of Existing Models

The above discussion suggests that the non-separable MO model, like the generalized CES one, can potentially match both the sales and the markup advantage of exporters. However, 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. Therefore K affects the average sales and markups of exporters versus non-exporters by adjusting the extensive margin but does not allow flexibility in the shape of the sales distribution.

In the Online Appendix, we conduct a quantitative analysis as in Section 5 for all the non-homothetic models described above: MO (both cases), SIM, and BMMS. For the separable models we find a θ such that either the markup advantage or the sales advantage are as in the data. For the non-separable MO model we can match the markup and sales advantage exactly for a (θ, K) pair. As discussed in the main text, the separable non-homothetic models are comparable to the GCES in terms of price discrimination but cannot jointly match the sales and markup moments. Allowing θ to match the markup advantage of exporters, all three models predict a sales advantage that is below what is seen in the data. For this reason we argue that the trade elasticity calibrated in these models will be mismeasured, leading to incorrect gains from trade estimates.

The non-separable MO model performed very poorly in all of the tests conducted in Section 5. As mentioned above, since the sales distribution is still fixed by θ, adjusting K does not allow one to reconcile the sales distribution. In fact the standard deviation of log sales is much smaller than in the GCES model. Furthermore, the average markup is extremely large — over 50%. The amount of price discrimination is much larger than the data, which together with the markup results shows that markups are too high and not in line with the data. Overall, while the model can match the two moments of interest in the data, it does so at the expense of generating highly counterfactual predictions along many dimensions.

H Over-Identified Procedure

In this section we discuss the GMM procedure with an optimal weighting matrix to identify σ and θ using six additional percentile moments from the sales and markup distributions, in addition to the two original moments used in the exact-identified estimation.[53] The eight moments are: 99-90, 90-10, and 90-50 percentile ratios of the sales and markup distributions, and sales and markup advantages of exporters over non-exporters as described in the main text.

For these moments, the orthogonality restriction is defined as

$$\mathbb{E}\left[h\left(\theta_0, \sigma_0; e\right)\right] = 0. \tag{73}$$

where $h\left(\sigma_0, \theta_0; s\right)$ is a 8×1 vector of differences between data moments and moments from

[53]This percentile moments matching estimation method is similar to the Method of Simulated Quantiles of Dominicy and Veredas (2013) that introduce quantile matching estimation, which is useful for situations where the density function does not have a closed form.

simulation for given a seed of simulation e and the true value of parameters (σ_0, θ_0). The sample average of $h(\theta_0, \sigma_0; e)$ is denoted by $g(\sigma, \theta)$ such that

$$g(\sigma, \theta) = M^d_{CHL} - \frac{1}{E}\sum_{e=1}^{E} M^m_{CHL}\left(\sigma, \theta, \tilde{t}_{ii}(e)\right), \tag{74}$$

with E the number of times we run the simulation. Note that $\tilde{t}_{ii}$ is the value of t_{ii} (domestic price relative to domestic cutoff) that corresponds to the cutoff cost that differentiates exporters from non-exporters. Then, we choose a pair of estimates of (σ, θ) that satisfy

$$\begin{pmatrix} \hat{\sigma} \\ \hat{\theta} \end{pmatrix} = \arg\min_{\sigma,\theta} g(\theta) W^{*-1} g(\theta), \tag{75}$$

where W^* is the optimal weighting matrix that produces the estimator with the smallest variance. We use the approximated optimal weighting matrix instead of W^* to reduce simulation burden following Gourieroux and Monfort (1997) and Adda and Cooper (2003).

Standard errors are computed using a parametric bootstrap technique that follows Simonovska and Waugh (2014b). First, we generate a random error term drawn from the Normal distribution with zero mean and the same variance to the estimate from the gravity regression. Then we create artificial data with the error terms added to the trade share data and re-estimate the country-specific terms $\hat{w}_i$ and $\hat{\Phi}_i$ from the artificial data. We also generate uniformly distributed random numbers between 0 and 1 to make a series of e independent from the previous random error terms and simulate costs and prices of firms for given e. Then we do the same estimation procedure with the new simulated result to get σ and θ. We repeat this procedure 100 times to compute standard errors of estimation. The standard errors allow us to compute the J-statistic from the Sargan-Hansen test for the over-identifying restrictions, reported in Table 1.

I Data

Table 4: List of Countries Used in The Analysis

Country Code	Country Names	Country Code	Country Names
ARG	ARGENTINA	KOR	KOREA
AUS	AUSTRALIA	MYS	MALAYSIA
AUT	AUSTRIA	MEX	MEXICO
BEL	BELGIUM	MOR	MOROCCO
BRA	BRAZIL	NEZ	NEW ZEALAND
BRN	BRUNEI	NOR	NORWAY
BUL	BULGARIA	OMN	OMAN
CAN	CANADA	PAK	PAKISTAN
CEF	CENTRAL AFRICAN REPUBLIC	PAR	PARAGUAY
CHL	CHILE	PER	PERU
CHN	CHINA	PHL	PHILIPPINES
COL	COLOMBIA	POL	POLAND
COT	IVORY COAST	PRT	PORTUGAL
CZE	CZECH REPUBLIC	ROM	ROMANIA
DNK	DENMARK	RUS	RUSSIA
ECU	ECUADOR	SAU	SAUDI ARABIA
EGY	EGYPT	SVK	SLOVAKIA
FIN	FINLAND	ZAF	SOUTH AFRICA
FRA	FRANCE	SPA	SPAIN
GER	GERMANY	SRI	SRI LANKA
GRC	GREECE	SWE	SWEDEN
HUN	HUNGARY	SWI	SWITZERLAND
ISL	ICELAND	SYR	SYRIA
IND	INDIA	THA	THAILAND
INDO	INDONESIA	TUN	TUNISIA
IRN	IRAN	TUR	TURKEY
IRL	IRELAND	UKR	UKRAINE
ISR	ISRAEL	UKX	UNITED KINGDOM
ITA	ITALY	USA	UNITED STATES
JAP	JAPAN	URY	URUGUAY
JOR	JORDAN	VEN	VENEZUELA
KAZ	KAZAKHSTAN	VNM	VIETNAM
KEN	KENYA	ZMB	ZAMBIA

Notes: This list represents the countries used in our analysis. There are 71 countries in the Economist Intelligence Unit (EIU) data of prices in each destination. Each of these countries are also available in the trade matrix. We drop 5 countries, and are therefore left with this list of 66 countries. The 5 dropped countries are: Azerbaijan, Ethiopia, Nigeria, Nepal, and Senegal.