International Trade with Indirect Additivity*

Paolo Bertoletti† — University of Pavia

Federico Etro‡ — Ca' Foscari University of Venice

Ina Simonovska§ — UC Davis, NBER

April 26, 2017

Abstract

We develop a general equilibrium model of trade that features "indirectly additive" preferences and heterogeneous firms. Monopolistic competition generates markups that are increasing in firm productivity and in destination country per-capita income, but independent from destination population, as documented empirically. The gains from trade liberalization are lower than in models based on CES preferences, and the difference is governed by the average pass-through. When we calibrate the model so as to match observed pricing-to-market in micro-data, it generates welfare gains that are substantially lower than those predicted by commonly-employed frameworks.

JEL Classification: D11, D43, L11, F12

Keywords: Pricing to market, Melitz model, Monopolistic competition, Non-homotheticity, Indirect additivity, Gravity, Welfare gains from trade

---

*We thank Daron Acemoglu, Pedro Bento, Matteo Cervellati, Arnaud Costinot, Paolo Epifani, Robert Feenstra, Emanuele Forlani, Antonio Lijoi, Elena Molho, Walter Steingress, especially Andres Rodriguez-Clare and David Weinstein, and seminar participants at NBER ITI 2015 Winter Meeting, ESEM 2016, UCSD, UCLA, UC Davis, Bayreuth University, Boston College, Brown University, University of Bologna, University of Dusseldorf, University of Essex, University of Milan Bicocca, University of Rennes, and the University of Minnesota International Economics Workshop for their comments and suggestions. Luca Macedoni provided outstanding research assistance. Ina Simonovska acknowledges financial support from the Hellman Fellowship and the Institute of Social Sciences at UC Davis.

†Dept. of Economics and Management, University of Pavia, Via San Felice, 5, I-27100 Pavia, Italy. Email: *paolo.bertoletti@unipv.it.*

‡Dept. of Economics, University of Venice Ca' Foscari, Sestiere Cannaregio, 30121, Fond.ta S.Giobbe 873, Venice, Italy. Email: *federico.etro@unive.it.*

§Dept. of Economics, University of California, Davis, One Shields Avenue, 95616, Davis, California. Email: *inasimonovska@ucdavis.edu.*

1 Introduction

Gains from intra-industry trade derive largely from the consumption of new and cheaper imported varieties (Broda and Weinstein, 2006). These gains have been the focus of international trade theory under monopolistic competition at least since Krugman (1980) and the subsequent large literature based on CES preferences. Summarizing this literature, Arkolakis *et al.* (2012, ACR) have shown that: i) the gains from trade liberalization can be simply captured by a formula featuring only the change in the domestic expenditure share and a "trade elasticity", namely the elasticity of the relative imports with respect to variable trade costs; and ii) for heterogeneous firm models *à la* Melitz (2003) and Chaney (2008) with a Pareto distribution of productivity, the trade elasticity depends only on the shape parameter of the distribution. Arkolakis *et al.* (2015, ACDR) have proved that, when demands feature a choke price and there are no fixed costs, the same welfare formula applies to some prominent examples of homothetic preferences and have shown that the welfare gains from trade are only marginally different under non-homothetic directly additive preferences *à la* Krugman (1979). These surprising results appear to suggest that not only the supply side dimension, but also consumer preferences play a limited role in shaping the gains from trade.

In this paper, we introduce in the literature on multi-country trade with heterogeneous firms a class of preferences that generates variable demand elasticities across firms (encompassing models with isoelastic, linear and other direct demand functions), and we show that it is crucial in shaping pricing and trade patterns as well as the gains from trade. We aim to quantify these effects and to this end we test a specific functional form. Our preferences are indirectly additive (IA), which means that they are represented by an indirect utility which is additive in prices (Houthakker, 1960). This class includes CES preferences as the only case in common with the classes of directly additive and homothetic preferences. In addition, it contains an entire family of well-behaved non-homothetic preferences with the unique property that the demand function of each good has an elasticity that depends on its own price and on income but not on other prices, and can be described empirically by a standard multinomial logit model.[1] Our assumptions on the supply side are standard. Each variety is produced by a firm after paying an entry cost with productivity drawn from a known distribution. Monopolistic competition reigns.

We initially analyze the equilibrium in autarky for general IA preferences and cost distributions. Firms adopt markups that can be variable in firms' marginal costs (incomplete

[1]Bertoletti and Etro (2017) consider trade under IA only between two countries and with identical firms. Among recent general equilibrium trade models with non-homothetic preferences and heterogeneous producers see Fieler (2011), Behrens *et al.* (2014), Feentra and Romalis (2014), Simonovska (2015), and ACDR *inter alia*. For a recent general equilibrium dynamic model with non-homothetic preferences see Etro (2016).

pass-thorugh) and in the income of consumers (pricing to market), but that are always independent of the size of the market. Therefore, opening up to costless trade induces gains from variety that are qualitatively *à la* Krugman (1980). However, except for the CES case, the equilibrium is inefficient: too many goods are consumed (relative to the mass of created goods) and low-cost firms produce too little. In the case of costly trade between identical countries, trade liberalization induces selection effects *à la* Melitz (2003) as long as production involves fixed costs. Moreover, the model generates incomplete pass-through of trade-cost reductions on the prices of imported goods and no impact on domestic prices, which limits the gains from trade liberalization.

To make headway in a multi-country framework with costly trade, we abstract from fixed production costs as in Melitz and Ottaviano (2008) and ACDR, and we adopt a Pareto distribution of productivity, which sets our model in the general gravity framework of ACR and Allen *et al.* (2014). The model predicts that firms extract higher mark-ups from richer destinations, but that they do not set different mark-ups in countries of different population size. These predictions are in line with the empirical results obtained by Simonovska (2015) from cross-country price data of identical products sold via the Internet. In particular, controlling for the cost to deliver products to a destination, the author finds that a typical monopolistically-competitive apparel producer charges higher prices for identical goods in richer destinations, but does not find evidence that prices vary with the population size of the market. Dingel (2015) obtains similar results using data on unit values for individual US producers across many destinations. Notice that traditional models of monopolistic competition cannot account for prices increasing in destination income when they are based on quasilinear preferences (Melitz and Ottaviano, 2008) or homothetic preferences (Feenstra, 2014), and generate prices that are decreasing in destination population when they are based on directly additive or homothetic preferences (for instance, see Behrens *et al.*, 2014, Simonovska, 2015, and ACDR).

The model generates further firm-level predictions that are consistent with data. First, more productive firms enjoy higher mark-ups, in line with the evidence in De Loecker and Warzynski (2012). Second, when trade costs are large, exporters are more productive and represent a minority of the active firms, as documented in Bernard *et al.* (2003), yet they may sell tiny amounts per export market, as documented in Eaton *et al.* (2011). New implications emerge for the margins of trade. The extensive margin is increasing in destination per-capita income, neutral in destination population (as is natural without fixed costs)[2] and falling in the trade cost to the destination. Hence, the model predicts that the extensive margin is falling in the distance to the destination and potentially rising in overall GDP of the destination country

[2]However, directly additive or homothetic preferences imply that the extensive margin is decreasing in the population of the importing country (see ACDR).

(the product of per-capita income and population), which is in line with Bernard *et al.* (2007). Finally, the intensive margin of trade is increasing in a destination's overall GDP and decreasing in the destination's per-capita income, which is in line with exploratory findings by Eaton *et al.* (2011) for several exporting countries across their export destinations.

As in Melitz (2003) and ACR, trade liberalization reallocates production across exporting and non-exporting firms and across countries. The two key differences are that reductions in trade costs are only partially shifted into lower prices due to incomplete pass-through, and that consumers purchase new foreign goods but keep buying the same domestic goods at the same prices, though in smaller quantities. We obtain a global quantitative measure of the welfare gains from trade liberalization that differs from the formula derived for CES (ACR) and other homothetic preferences (ACDR). In particular, the gains from trade are reduced by a coefficient that corresponds to the sales-weighted average pass-through, which in our model is also one minus the sales-weighted average elasticity of price with respect to income. When demand is very elastic, a high pass-through generates high gains from trade liberalization because lower trade costs are largely shifted into lower prices of imports (without consequences on the domestic prices), and these are exploited by consumers purchasing new imported varieties. In contrast, when demand is rather inelastic, pass-through is low and the welfare gains are limited. Notice that any model based on CES preferences, as in Melitz (2003) and ACR, produces complete pass-through for all firms. As shown by ACDR, a similar effect also arises in models based on other homothetic preferences since two effects of trade liberalization balance each other out: on the one hand, inframarginal exporting firms tend to increase their markups due to incomplete pass-through, and on the other hand, a selection effect on the set of domestic firms tends to reduce the markups of all firms.[3]

To obtain further predictions and to quantify the welfare gains from trade liberalization, we introduce a specification of preferences which generates demand functions nesting the special cases of perfectly elastic, linear and perfectly inelastic demands. This yields closed form solutions for firm-level and aggregate variables as well as for the welfare gains, and delivers additional predictions in line with the data. First, trade liberalization increases sales more for smaller firms, as documented by Eaton *et al.* (2008) and Arkolakis (2016). Together with the fact that trade liberalization induces new entry of foreign varieties, this implies that adjustments on the extensive margin are critical in understanding the welfare gains from trade (Broda and Weinstein, 2006; Kehoe and Ruhl, 2013). Second, the degree of cost pass-through is falling in firm productivity, as documented by Berman *et al.* (2014). This implies that larger firms change prices less with changes in costs and more with changes in income, and it is precisely

[3]ACDR obtain marginally smaller gains for directly additive non-homothetic preferences because the selection effect only partially countervails the incomplete pass-through effect.

this behavior that directly impacts the welfare gains.

Since we parameterize firm productivity to be unbounded Pareto, the model shares an identical loglinear gravity equation of trade with the models examined in ACR and ACDR, where the Pareto shape parameter governs the trade elasticity. To quantify the differences in welfare gains predicted by the two classes of models, we let the trade elasticity take on the value of 5 in line with ACR, ACDR, and estimates in Caliendo and Parro (2015), and we calibrate the preference parameter which governs the degree of pricing-to-market and pass-through to match the average elasticity of price with respect to income as reported from micro data by Simonovska (2015). We find that the gains from any trade liberalization experiment are about 30% lower than the gains reported by ACR for homothetic models. Hence, the mismeasurement of welfare due to ignoring incomplete pass-through is both quantitatively and economically large, which leads us to conclude that the demand side is crucial in understanding the welfare gains from trade.[4]

We proceed as follows: In Section 2, we present the baseline setting of our model. In Section 3, we study trade between heterogenous countries. In Section 4, we quantify the mismeasurement of welfare gains from trade. We conclude in Section 5.

2 Framework

Consider a market populated by L identical agents, each one with labor endowment e. Firms can produce a variety from a set Ω at a constant marginal cost after paying a sunk entry cost $F_e > 0$. Upon entry, the "intrinsic" marginal cost c of each firm is independently and identically drawn from a distribution $G(c)$ with support $[0, \overline{c}]$ for a large, and possibly infinite, $\overline{c} > 0$. For tractability, we neglect fixed costs of production (as in Melitz and Ottaviano, 2008 and ACDR). All costs are in (efficiency) units of labor and the labor market is perfectly competitive: in this section we normalize the wage to unity so that c is marginal cost and, given zero expected profits, *per capita* income E just equals the individual labor endowment.

The indirect utility of each agent depends (exploiting homogeneity of degree zero) on the normalized prices $s(\omega) = p(\omega)/E$, for $\omega \in \Omega$, according to the following additive specification:

$$V = \int_{\Omega} v(s(\omega))d\omega, \tag{1}$$

[4]In the Appendix we also identify the model's parameters by matching well-documented firm-level moments in the literature, which provides even lower gains from trade liberalization. With increasing availability of firm- and product-level data, it would be of interest to estimate preference parameters from these data and test another unique prediction of this model, which relates the demand elasticity for a good on the good's own price and consumer income, but not on other prices.

where v is a decreasing and convex function up to a (possibly infinite) choke value a, so that aE is the maximum willingness to pay for each variety, with $v(s) = v'(s) = 0$ for all $s \geq a$. With the exception of CES preferences that it encompasses (with an infinite choke price), (1) represents a class of preferences that are neither homothetic nor directly additive (see Bertoletti and Etro, 2017). An important property of these preferences for empirical purposes is that a market demand system can be described by the multinomial logit model with income effects if and only if the representative consumer is endowed with indirectly additive preferences (Thisse and Ushchev, 2016). By Roy's identity, the individual demand for each variety ω that is actually consumed is given by:

$$x(\omega) = \frac{v'(s(\omega))}{\mu}, \tag{2}$$

where $\mu = \int_\Omega v'(s(\omega))s(\omega)d\omega = -E\left(\partial V/\partial E\right)$ depends on all prices and $\partial V/\partial E$ is the marginal utility of income. Accordingly, demand faced by a producer of variety ω is decreasing in its own price $p(\omega)$ and vanishes if this is above the choke level:

$$\widehat{p} = aE, \tag{3}$$

which depends linearly on income.[5]

2.1 Autarkic equilibrium

Let N be the measure of firms paying the entry cost: we analyze monopolistic competition among a measure $n \leq N$ of active firms producing different varieties for a given distribution of costs. The profits of a firm with marginal cost c choosing a price $p(c)$ can then be written as:

$$\pi(c) = \frac{(p(c) - c)v'\left(\frac{p(c)}{E}\right)L}{\mu}, \tag{4}$$

where μ is unaffected by a single firm. The demand function (2) has a price elasticity which is just the elasticity of $v'(s)$, namely $\theta(s) \equiv -\frac{sv''(s)}{v'(s)}$, which is variable across firms, as is the income elasticity of demand, which can be computed as $\theta(s) - \frac{\partial \ln|\mu|}{\partial \ln E}$.[6] The profit-maximizing

[5] The dependence of the choke price on income alone is a key property of IA. In other models based on homothetic or directly additive preferences, the choke price depends on the marginal utility of income, i.e., on the price distribution and on the measure of consumed varieties (see Feenstra, 2014, and ACDR). In the quasilinear model of Melitz and Ottaviano (2008), the marginal utility of income is fixed at unity, but the choke price depends on the number of varieties and on their average price.

[6] The tight connection between price and income elasticities is of course due to the additivity of preferences: see Houthakker (1960).

pricing rule is:

$$p(c) = c\left(\frac{\theta\left(p(c)/E\right)}{\theta\left(p(c)/E\right)-1}\right), \tag{5}$$

for any $c > 0$. To satisfy the conditions for the existence of a well-defined optimal price $p(c)$, we assume that $\theta(s) > 1$ and that the second-order condition $2\theta(s) > \zeta(s)$ is satisfied for all $s \in (b, a)$, where $b \equiv p(0)/E$ and $\zeta(s) \equiv -v'''(s)s/v''(s)$ is a measure of demand curvature (see Bertoletti and Etro, 2017). Notice that the optimal markup of each firm, $m(c) = (p(c) - c)/c$, is independent from the number of goods available and from the price of any other firm.

Let us actually assume $\theta'(s) \geq 0$,[7] which is equivalent to what Mrázová and Neary (2013) define as "subconvexity" of the demand function, and it is sometimes called "Marshall's Second Law of Demand". When demand elasticity is strictly increasing, the model has four main implications for pricing across firms. First, prices are lower but markups are higher for more productive firms, which differs from the Melitz (2003) model. Second, markups increase with the income of consumers and differently across firms: these effects are absent in any model based on homothetic preferences (Melitz, 2003; Feenstra, 2014) or quasilinear preferences (Melitz and Ottaviano, 2008). Third, and differently from models based on directly additive preferences (Behrens *et al.*, 2014; Bertoletti and Epifani, 2014 and Simonovska, 2015), markups are independent from market size L. Fourth, it is easy to verify that the elasticities of prices with respect to income and marginal cost sum to one:

$$\epsilon^E(c) \equiv \frac{\partial \ln p(c)}{\partial \ln E} = \frac{\theta\left(s\left(c\right)\right)+1-\zeta\left(s\left(c\right)\right)}{2\theta\left(s\left(c\right)\right)-\zeta\left(s\left(c\right)\right)} = 1 - \frac{\partial \ln p(c)}{\partial \ln c} \equiv 1-\epsilon^c(c) \in (0,1), \tag{6}$$

which shows the *inverse* relation between pricing to market and pass-through.[8]

The individual consumption of the variety produced by a c-firm is $x(c) = v'(p(c)/E)/\mu$, which is zero if its price (given by (5)) if above the choke level $\widehat{p}$. The equilibrium set of active firms is simply given by the interval $[0, \widehat{c}]$, where the marginal cost cutoff $\widehat{c} = aE$ is just the choke price.[9] The model is closed equating the expected gross profits:

$$\mathbb{E}\left\{\pi(c)\right\} = \int_0^{\widehat{c}} \frac{(p(c)-c)v'(p(c)/E)L}{\mu} dG(c)$$

[7] This implies $\theta' s/\theta = \theta + 1 - \zeta \geq 0$. See Bertoletti and Etro (2017) for an exploration of the case in which $\theta' < 0$ and there are fixed costs.

[8] Notice that $\theta' > 0$ implies also that the elasticity of demand with respect to income is higher for smaller firms with higher prices. However, by augmenting the model to endogenous quality as in Bertoletti and Etro (2017), we can obtain that more productive firms sell higher quality goods with higher prices and larger income elasticities. See Caron *et al.* (2014) on the role of income elasticities in trade.

[9] This assumes that the constraint $\widehat{c} \leq \overline{c}$ never binds.

to the entry cost F_e. Since $\mu = N\int_0^{aE} v'(s(c))s(c)dG(c)$, this gives:

$$N = \frac{EL}{\bar{\theta}F_e} \qquad \text{with } \overline{\theta} \equiv \left[\int_0^{\widehat{c}} \frac{1}{\theta(p(c)/E)} \frac{p(c)x(c)}{\int_0^{\widehat{c}} p(c)x(c)dG(c)} dG(c)\right]^{-1}, \tag{7}$$

where $\bar{\theta}$ is the harmonic average of demand elasticities weighted by the market shares. In particular, notice that the equilibrium distribution of normalized prices $F_s(s)$ has support $[b, a]$ and is given by:

$$F_s(s) = \Pr\{p(c) \leq sE; c \leq \widehat{c}\} = \Pr\{c \leq h(s)E; c \leq aE\} = \frac{G(h(s)E)}{G(aE)},$$

where $h(s) = s[1 - 1/\theta(s)]$ $(h' > 0)$. This allows us to express the average demand elasticity as:

$$\overline{\theta} = \left[\int_b^a \frac{1}{\theta(s)} \frac{sv'(s)}{\int_b^a sv'(s)dF_s(s)} dF_s(s)\right]^{-1}, \tag{8}$$

which is independent from market size (but can depend on income). Accordingly, the measure of consumed varieties $n = NG(\widehat{c})$ must be linear in population. Two simple examples are in order.

Isoelastic demand The familiar case of CES preferences should clarify the nature of the equilibrium. Consider $v(s) = s^{1-\theta}$ where $\theta \in (1, \infty)$ governs the constant demand elasticity. The Roy's identity delivers the *isoelastic demand* $x(\omega) = (\theta - 1)s(\omega)^{-\theta}/|\mu|$, or:[10]

$$x(\omega) = \frac{p(\omega)^{-\theta} E}{\int_\Omega p(\omega)^{1-\theta} d\omega}.$$

As well known, there is no *finite* choke price, the equilibrium prices are $p(c) = \frac{\theta c}{\theta - 1}$ since $\bar{\theta} = \theta$, and pass-through is complete. Therefore, the number of goods created is $N = \frac{EL}{\theta F_e}$ and all these goods are produced and consumed (since we have abstracted from fixed costs of production).

Linear demand Consider now a new example with $v(s) = (a - s)^2/2$. The Roy's identity (2) delivers the *linear demand* function $x(\omega) = (a - s(\omega))/|\mu|$, or:

$$x(\omega) = \frac{aE - p(\omega)}{\int_\Omega (aE - p(\omega))(p(\omega)/E)d\omega}.$$

[10]The indirect utility can be expressed (up to a mononotic transformation) as $V = E\left[\int_\Omega p(\omega)^{1-\theta} d\omega\right]^{1/(\theta-1)}$.

It is immediate to derive a familiar expression for the monopolistic price:

$$p(c) = \frac{c + aE}{2},$$

which is increasing less than proportionally in the marginal cost and in income, but is independent from the number of goods and population. Demand elasticity $\theta(s) = \frac{s}{a-s} \in (1, \infty)$ is increasing in the price-income ratio. The profits of an active firm with $c \leq \widehat{c} = aE$ are then given by $\pi(c) = \frac{(\widehat{c}-c)^2 L}{4E|\mu|}$. Further results can be obtained by assuming (as in Chaney, 2008, and the subsequent literature) that the cost distribution corresponds to a productivity distribution that is Pareto (unbounded above), namely $G(c) = (c/\overline{c})^{\kappa}$ with $\overline{c} > 0$ finite and $\kappa > 1$ as the shape parameter.

We can then compute $\mathbb{E}\{\pi(c)\} = \frac{L\widehat{c}^{\kappa+2}}{2(\kappa+1)(\kappa+2)E|\mu|\overline{c}^{\kappa}}$ and $|\mu| = \frac{N\widehat{c}^{\kappa+2}}{2(\kappa+2)E^2\overline{c}^{\kappa}}$. Accordingly, we obtain $\mathbb{E}\{\pi(c)\} = EL/\left[(\kappa+1)N\right]$ and thus, under free entry:

$$N = \frac{EL}{(\kappa+1)F_e},$$

which implies $\bar{\theta} = \kappa + 1$.[11] Since the choke price is finite only a fraction $G(aE)$ of the N firms that entered the market are active in this case.

Turning to the empirical literature, as discussed in detail in the introduction, the distinct prediction of IA preferences—the neutrality of population on prices—finds empirical support in markets of monopolistic competition with a large number of firms: see Simonovska (2015) and Dingel (2015). In addition, the model's prediction is in line with empirical findings by Handbury and Weinstein (2014) for U.S. cities: identifying varieties with barcode data, and controlling for all retail heterogeneity and purchasers' characteristics, the authors provide convincing evidence that larger cities do not feature different prices of individual varieties, but have more varieties available, which yields lower price indices there. Using the same data source, Broda *et al.* (2009) document also that richer consumers pay more for identical products even after controlling for average income in the zip code in which they live, where the latter aims to capture local costs to operate the store. In contrast, there is a literature that links productivity, prices, and mark-ups to city size. For example, Campbell and Hopenhayn (2005) find that retail establishments are larger in larger U.S. cities, which suggests that mark-ups should be decreasing with firm entry. Similarly, Hottman (2014) finds that retailers' mark-ups vary with the size of US cities. Both papers argue that these observations indicate that the retail sector is oligopolistic, rather than being monopolistically competitive.[12]

[11] $\bar{\theta}$ can be obtained directly also by using the equilibrium distribution $F_s = \left(\frac{2s-a}{a}\right)^{\kappa}$ to compute (8).

[12] It is well known that markets with a small number of firms would exhibit equilibrium markups decreasing in the size of the market due to strategic interactions (which depend on the number of competitors). For recent

With respect to the theoretical literature, a negative equilibrium relationship between market size and prices emerges in existing (non-CES) models of monopolistic competition (see Melitz and Ottaviano, 2008 and ACDR), and is often regarded as a pro-competitive effect. However, this relationship is not due to a strengthening of competition on the supply side, since strategic interactions are absent from these models. The mechanism is driven by equilibrium changes in the substitutability between products on the demand side, whose nature and direction cannot be easily verified empirically. While we consider the neutrality of population on prices an attractive feature of our monopolistic competition setting, competition effects could be easily introduced by adding demand externalities or strategic interactions.

2.2 Welfare and trade among identical countries

The effect of an expansion of the market size under IA replicates a key property of the Krugman (1980) model, for which a larger population (whose impact is equivalent to opening up to costless trade with identical countries) increases proportionally the number of firms/varieties created in equilibrium without affecting markups (which is not the case under direct additivity; see, for instance, Dhingra and Morrow, 2014). Since the range of created goods which are actually consumed and their prices are independent from population, this generates pure gains from variety. To see this, notice that welfare can be computed as follows:

$$V = n \int_0^{\widehat{c}} v\left(\frac{p(c)}{E}\right) \frac{dG(c)}{G(\widehat{c})}. \tag{9}$$

This is linear in the measure of consumed varieties $n = G(\widehat{c})N$, which in turn is linear in the population size. Therefore, costless trade leads to welfare gains that are due *only* to an increase of the mass of consumed varieties for any IA preferences and cost distribution.

With the notable exception of CES preferences, our setting implies an inefficient market allocation. To verify this, in Appendix A we solve the social planner problem for the maximization of utility under the resource constraint.[13] The optimal allocation delivers the following measure of firms:

$$N^* = \frac{EL}{(\bar{\eta}+1)F_e} \qquad \text{with} \quad \bar{\eta} \equiv \int_0^{\widehat{c}^*} \eta\left(s(c)\right) \frac{v(s(c))}{\int_0^{\widehat{c}^*} v\left(s(c)\right) dG(c)} dG(c), \tag{10}$$

trade models with strategic interactions see Atkeson and Burstein (2008) and Etro (2015). One could explicitly introduce an oligopolistic retail sector into our model and derive new pricing predictions, but such extension is beyond the scope of the present paper.

[13]Dhingra and Morrow (2015) and Nocco *et al.* (2014) have analyzed optimality with heterogeneous firms in the cases of direct additivity of preferences and of quasilinearity, respectively (without fixed costs, in the latter case). In the case of homogeneous firms, the characterization of the social planner problem for any symmetric preferences was first derived in Bertoletti and Etro (2016).

where $\bar{\eta} > 0$ is a weighted average of the subutility elasticity $\eta(s) = -v'(s)s/v(s) > 0$, with relative utilities as weights, and it is again independent from L. This allows for equilibrium entry either above or below optimum ($\overline{\theta}$ should be compared to $\bar{\eta} + 1$). More importantly, the social planner sets a constant mark up $m^* = 1/\bar{\eta}$, as needed to equalize the marginal rate of substitution between any two produced goods to their marginal cost ratio:

$$p^*(c) = \left(1 + \frac{1}{\bar{\eta}}\right) c. \tag{11}$$

Finally, the optimal marginal cost cutoff is smaller than the equilibrium one (when they are finite):

$$\widehat{c}^* = \frac{aE\bar{\eta}}{1 + \bar{\eta}} < \widehat{c}. \tag{12}$$

It follows that the equilibrium prices must be above optimal for the most efficient firms and below optimal for the most inefficient firms.[14] Therefore, a redistribution of production from high cost toward low cost firms would indeed improve the allocation of resources.

To gain further insights, our two examples are again useful. With CES preferences $\bar{\eta} = \theta - 1$ and the equilibrium is optimal (see Dhingra and Morrow, 2014). With any other IA preference relation with a finite choke price, if the cost distribution corresponds to a productivity distribution that is Pareto, we obtain that $\bar{\eta} = \kappa$ (see Appendix A). This result reveals an interesting property shared by our earlier example with linear demand and its generalization presented in Section 3.3 below: in equilibrium, the number of firms is optimal. Nevertheless, a pervasive inefficiency remains: too many goods are consumed relative to the mass of firms created, and low-cost firms produce too little while high-cost firms produce too much.

Our framework can be easily extended to trade frictions between identical countries for any IA preferences, which indeed include the CES case of Melitz (2003). First, notice that identical countries must have the same wage and the same value of μ. Second, given an iceberg transport cost $\tau > 1$, the pricing rules are the same as before, with $p(c)$ given by (5) for domestic sales and $p(\tau c)$ for foreign sales, both independent from the market size. However, as long as $\theta' > 0$, each exporting firm applies a lower markup on exports compared to the markup on domestic sales.[15] As long as there are positive fixed costs of production, such a model delivers an equilibrium partition of firms between exporters and non-exporters and selection effects of trade liberalization *à la* Melitz.[16] Most importantly, a reduction of the trade cost τ

[14]In fact, $p^*(0) = 0 \leq p(0)$ and $p^*(\widehat{c}^*) = \widehat{c} > p(\widehat{c}^*)$, with $p^{*\prime}(c) > 1 > p'(c)$ under the assumption that $\theta' > 0$.

[15]If varieties differ endogenously in qualities one can show that more efficient exporters can sell abroad better products with higher markups (see Bertoletti and Etro, 2017). The existing empirical literature points in this direction: see Bastos and Silva (2010), Manova and Zhang (2012), Martin (2012), and Dingel (2015).

[16]Assume that production in each market requires a fixed cost $F \geq 0$. The net profits from domestic sales are $\pi(c) - F$ and those from exports are $\pi(\tau c) - F$, where $\pi(c)$ is always given by (4). The cutoff cost for domestic

increases the markups of the previously imported goods due to incomplete pass-through, but does not affect the markups of the inframarginal domestic goods. As a consequence, trade liberalization induces a redistribution of production from high-cost non-exporting firms toward low-cost exporters, but it also increases the average markup of inframarginal firms, which tends to limit the welfare gains. The next section develops a fully-fledged multicountry model with the purpose of quantifying those gains.

3 Trade among heterogeneous countries

We now consider costly trade between countries that are heterogeneous in population, per-capita labour endowment and trade costs. The "iceberg" cost of exporting from country i (source) to country j (destination) is $\tau_{ij} \geq 1$, with $\tau_{ii} = 1$ for $i, j = 1, .., I$ where $I \geq 2$ is the number of countries. Country i has N_i firms paying the entry cost F_e, population L_i, wage w_i, marginal costs $\tau_{ij} w_i c$ for the destination country j and *per-capita* labor endowment e_i, so zero expected profits imply that individual income is $E_i = w_i e_i$. We assume that preferences exhibit a finite choke price a, that $\theta' > 0$ (to obtain pricing to market and incomplete pass-through) and that the cost distribution (that corresponds to a Pareto distribution of productivities) is given by $G(c) = (c/\overline{c})^{\kappa}$, where $\kappa > 1$ is the shape parameter and $\overline{c}$ is finite (but large enough to be never binding).

The profit that a c-firm from country i makes by selling to country j is:

$$\pi_{ij}(c) = \frac{(p_{ij}(c) - \tau_{ij} w_i c)\, v'\left(\frac{p_{ij}(c)}{E_j}\right) L_j}{\mu_j}, \tag{13}$$

where $|\mu_j|$ is as usual the marginal utility of income (times per capita income) of country j. Maximizing these profits delivers the price rule:

$$p_{ij}(c) = \tau_{ij} w_i c \left(\frac{\theta\,(p_{ij}(c)/E_j)}{\theta\,(p_{ij}(c)/E_j) - 1} \right). \tag{14}$$

As noticed earlier, this predicts that prices and markups are higher in countries that enjoy

sales is $\widehat{c} = \pi^{-1}(F)$, and the cutoff for the marginal exporting firm is $\widehat{c}_x = \widehat{c}/\tau$. The free-entry condition:

$$\int_0^{\widehat{c}} [\pi(c) - F]\, dG(c) + \int_0^{\widehat{c}/\tau} [\pi(\tau c) - F]\, dG(c) = F_e$$

closes the model. As long as $F > 0$, one can easily obtain by total differentiation that $\partial \widehat{c}/\partial \tau > 0$. The intuition is simple: lower trade costs increase the expected profits of exporting firms at the expense of non-exporters, which implies that the domestic cut-off firm must now be more efficient (for a similar result with directly additive preferences see Bertoletti and Epifani, 2014).

higher per-capita income levels, but independent of their population size. Moreover, since the markup in expression (14) is falling in the firm's intrinsic cost of production, c, the model predicts that more productive firms enjoy higher markups, which is in line with wide evidence, for instance with observations in Slovenian data documented by De Loecker and Warzynski (2012). Once again, the elasticities $\epsilon_{ij}^{E}(c) \equiv \partial \ln p_{ij}/\partial \ln E_j$ and $\epsilon_{ij}^{c}(c) \equiv \partial \ln p_{ij}/\partial \ln c$ add to 1.

The individual quantity sold by a c-firm of country i to destination j is given by $x_{ij}(c) = v'\left(p_{ij}(c)/E_j\right)/\mu_j$. The value of the corresponding sales $t_{ij}(c) = p_{ij}(c)x_{ij}(c)L_j$ is:

$$t_{ij}(c) = \frac{p_{ij}(c)v'\left(\frac{p_{ij}(c)}{E_j}\right)L_j}{\mu_j}.$$

The most inefficient firm in country i which is actually able to serve country j has the marginal cost cutoff:

$$\widehat{c}_{ij} = \frac{aE_j}{\tau_{ij}w_i}, \tag{15}$$

(remember that $v'(a) = 0$) which simplifies to $\widehat{c}_{ii} = ae_i$ for the domestic sales in country i. Therefore, in our model the range of the firms active domestically is wider in the country with higher individual labor endowment, and depends neither on the population size (since there are no fixed costs) nor on the trade costs. Instead, the set of exporters enlarges with the per capita income of the importing country, and shrinks with the bilateral trade cost and the exporter's wage. A key consequence is that trade liberalization does not affect the range of the firms active at home but just enlarges the set of exporters, and therefore the measure of imported varieties. However, like in other models (Melitz and Ottaviano, 2008, ACR and ACDR), production is reallocated across firms toward exporters and across countries. If trade costs are sufficiently high, exporters are more productive than non-exporters, represent a minority of the active firms as documented by Bernard *et al.* (2003), and they may sell tiny amounts per export market (the marginal exporter has zero sales), as documented in Eaton et al. (2011).

Defining $n_j^C \equiv \sum_{i=1}^{I} n_{ij}$ to be the measure of goods consumed in country j, we have:

$$\mu_j = \sum_{i=1}^{I} n_{ij} \int_0^{\widehat{c}_{ij}} v'(s_{ij}(c))s_{ij}(c)\frac{dG(c)}{G\left(\widehat{c}_{ij}\right)} = n_j^C \int_b^a v'(s)sdF_s(s),$$

where $F_s(s)$ is the equilibrium distribution of normalized prices. This distribution has support $[b, a]$, is identical across countries, and is independent from incomes and trade costs:

$$F_s(s) = \Pr\left\{c \leq h(s)\frac{\widehat{c}_{ij}}{a}; c \leq \widehat{c}_{ij}\right\} = \left(\frac{h(s)}{a}\right)^{\kappa},$$

where $h(s) = s[1 - 1/\theta(s)]$. The neutrality of the distribution of normalized prices as well as of markups from trade costs is due to the fact that liberalization reduces the prices of inframarginal exporting firms and increases their markups (due to incomplete pass-through), but it also attracts the entry of new exporters with higher prices (and smaller markups), and these effects exactly balance each other out. However, trade liberalization does not affect domestic pricing; therefore the average markup across all inframarginal firms (domestic and exporting) must increase. Finally, it is immediate to verify that the distribution of individual consumption is affected by any change in μ_i and, in particular, by changes in trade costs.[17]

Taking the expectations of sales and profits, we obtain the ratio:

$$\frac{\mathbb{E}\{\pi_{ij}\}}{\mathbb{E}\{t_{ij}\}} = \frac{1}{\overline{\theta}}, \tag{16}$$

where the constant $\overline{\theta}$ is the equilibrium harmonic average of demand elasticity defined in (8), which under the Pareto distribution is identical across countries.[18] Under endogenous entry, total expected profits $\mathbb{E}\{\Pi_i\} = \sum_{j=1}^{I} \mathbb{E}\{\pi_{ij}\}$ in country i must equate the fixed cost of entry $w_i F_e$. Let us define total (expected) sales from country i, as $Y_i = \sum_{j=1}^{I} T_{ij}$, where $T_{ij} = N_i \mathbb{E}\{t_{ij}\}$ are the (expected) sales in country j that originated from country i. The endogenous entry condition reads as:

$$\mathbb{E}\{\Pi_i\} = w_i F_e,$$

and the income/spending equality for country i implies $w_i e_i L_i = Y_i$, where $Y_i = \sum_j T_{ji}$ is GDP in country i. Therefore, we can derive the number of firms created in country i as:

$$N_i = \frac{w_i e_i L_i}{\sum_{j=1}^{I} \mathbb{E}\{t_{ij}\}} = \frac{e_i L_i}{F_e} \frac{\sum_{j=1}^{I} \mathbb{E}\{\pi_{ij}\}}{\sum_{j=1}^{I} \mathbb{E}\{t_{ij}\}} = \frac{e_i L_i}{\overline{\theta} F_e}, \tag{17}$$

which is the same as in autarky. Accordingly, trade does not affect the measure of firms created nor that of domestic firms active in each country.

[17] This is just the opposite of models based on directly additive preferences with a finite choke price, where changes in trade costs are neutral on the distribution of individual consumption and modify the distribution of prices.

[18] Our setting satisfies the Assumptions R1 and R2 of ACR (p. 102).

3.1 Trade margins and general equilibrium

We can now derive the measure of firms actually exporting to any country j from country i, $n_{ij} = N_i G(\widehat{c}_{ij})$, the so-called extensive margin of trade, as:

$$n_{ij} = N_i \left(\frac{\widehat{c}_{ij}}{\overline{c}} \right)^{\kappa}. \tag{18}$$

This depends negatively on the trade cost and positively on per-capita income of the destination country through $\widehat{c}_{ij}$, and on the "aggregate labor supply" $e_i L_i$ of the exporting country through N_i, but it is independent from the population size of the destination country. Hence, the model predicts that the extensive margin is falling in the distance to the destination (to the extent that trade costs are increasing in distance) and potentially rising in overall GDP of the destination country (which is the product of per-capita income and population), as reported in Bernard *et al.* (2007). A positive relationship between destination population size and the extensive margin can be restored by the introduction of fixed costs *à la* Melitz (2003).[19]

The evidence on the relationship between the extensive margin and population size is mixed. Authors who use internet data (e.g. Macedoni, 2015)[20] find that the extensive margin is neutral in population size as predicted by the baseline IA model. In contrast, authors who use traditional trade data such as firm-level or product-category-level manufacturing data (e.g. Macedoni, 2015 when using the Exporter Dynamics Database or Hummels and Klenow, 2005 when using disaggregate bilateral trade-flow trade), find that the extensive margin is increasing in population size and the coefficient estimates vary across countries and industries. This suggests that there is heterogeneity in fixed costs across countries and industries, ranging from nearly zero in online markets to positive and potentially large costs in traditional retail markets.

An implication of our model is that the extensive margin is increasing as the importing country gets richer, which is in line with the growth in the measure of imported varieties documented by Broda and Weinstein (2006) for the US over three decades. The authors document that, during the period 1972-2001, the number of imported varieties in the United States has increased by a factor of three. They also note that half of the rise can be attributed to new products sold by existing trade partners. We should remark that, contrary to the predictions of the IA framework, models based on directly additive or homothetic preferences (without fixed costs of production, as in ACDR) imply that the extensive margin is decreasing in the population of the destination country and it is neutral (increasing) with respect to income under homotheticity (direct additivity), while the Melitz-Chaney model (with fixed costs expensed

[19]See footnote 33.

[20]The author collects data for Samsung and verifies the results using data for Zara, Apple, H&M, and Ikea from the Billion Prices Project.

in source country wages) generates an extensive margin that is increasing in both destination income and population.

The total measure of varieties consumed in country j can be expressed as follows:

$$n_j^C = \frac{\sum_{i=1}^{I} e_i L_i \widehat{c}_{ij}^{\kappa}}{\overline{\theta}\overline{c}^{\kappa} F_e}, \tag{19}$$

which crucially depends on j's per-capita income and on its trade costs (through $\widehat{c}_{ij}$), and on the labor force of its trading partners. It follows that countries that are richer in per capita income terms (and larger in labor endowment) tend to consume more goods. This makes a remarkable difference with analogous models based on homothetic preferences and (untruncated) Pareto distribution (see Arkolakis *et al.*, 2010, and Feenstra, 2014), in which the measure of consumed goods is equal across countries and independent of their income, population and trade costs.

Expected sales from country i to country j can be derived, by computing $\mathbb{E}\{t_{ij}\}$, as follows:[21]

$$T_{ij} = N_i \mathbb{E}\{t_{ij}\} = Y_j \frac{n_{ij}}{n_j^C} = Y_j \frac{e_i L_i \widehat{c}_{ij}^{\kappa}}{\sum_{k=1}^{I} e_k L_k \widehat{c}_{kj}^{\kappa}} = n_{ij} \overline{t}_{ij},$$

where we decomposed trade into the product of the extensive margin and the intensive margin $\overline{t}_{ij}$. We can rewrite the latter as:

$$\overline{t}_{ij} = \frac{\mathbb{E}\{t_{ij}\}}{G(\widehat{c}_{ij})} = \frac{L_j E_j}{n_j^C}. \tag{20}$$

This is independent from the country of origin of the commodities, but it depends on both per-capita income and population of the destination country, contrary to the intensive margin of the Melitz-Chaney model, which is constant as long as the (fixed) costs of export are expensed in labor of the source country.[22] Two direct effects are immediately observable: in our model, the intensive margin is increasing in the destination's population size, and decreasing with respect to the destination's per capita income (through the impact of E_j on n_j^C). Therefore, the model can jointly generate a positive relationship between the intensive margin and the overall GDP of a destination, and a negative relationship with the destination's per-capita income, as documented for several source countries by Eaton *et al.* (2011). Notice that these implications are in contrast also with comparable models without fixed costs of production (ACDR): directly additive and homothetic preferences generate an intensive margin which is

[21] Accordingly, our setting also satisfies assumption R3 of ACR (pp. 103-4).

[22] The Melitz-Chaney model relates the intensive margin to the fixed cost to serve a destination and to the unit in which it is expensed. If this cost is parameterized to be source- and destination-specific as in Eaton *et al.* (2011), and if one assumes that the cost is systematically related to destination characteristics, the model can generate a systematic relationship between the intensive margin and destination characteristics.

always increasing in destination income.

To close the model in general equilibrium, notice that:

$$\frac{T_{ij}}{T_{jj}} = \frac{Y_i}{Y_j}\left(\frac{w_i}{w_j}\right)^{-(\kappa+1)}\tau_{ij}^{-\kappa}. \tag{21}$$

This simple result can be interpreted as follows: the assumption of a Pareto distribution gives rise to a generalized "gravity" equation that governs the trade shares (see e.g. Head and Mayer, 2014, and Allen *et al.,* 2014), where κ is the "trade elasticity" according to the terminology suggested by ACR. In particular, the *trade share* of i-goods in country j is given by:

$$\lambda_{ij} = \frac{T_{ij}}{Y_j} = \frac{n_{ij}}{n_j^C} = \frac{e_i L_i \widehat{c}_{ij}^{\kappa}}{\sum_{k=1}^{I} e_k L_k \widehat{c}_{kj}^{\kappa}}. \tag{22}$$

Finally, using the expressions for the trade shares we can express the income-spending equation of each country i as:

$$w_i e_i L_i = \sum_{j=1}^{I} \lambda_{ij} E_j L_j. \tag{23}$$

Using (22) and (23) provides the equilibrium wage system:

$$w_i = \sum_{j=1}^{I} \frac{w_j e_j L_j \left(\tau_{ij} w_i\right)^{-\kappa}}{\sum_{k=1}^{I} e_k L_k \left(\tau_{kj} w_k\right)^{-\kappa}} \quad i = 1, .., I, \tag{24}$$

which is similar to those of related models satisfying the restrictions of ACR. It implies wage equalization only under free trade or identical countries (as in the previous sections). Moreover, it can be proved (Alvarez and Lucas, 2007) that (24) has a unique solution (up to a normalization) and that the relative wage of country j is increasing in its aggregate labor supply $e_j L_j$ and decreasing in its trade costs $\boldsymbol{\tau_j'} = [\tau_{1,j}, .., \tau_{I,j}]$.

3.2 Welfare and comparison with other models

Utility for a consumer of country j can be expressed as:

$$V_j = n_j^C \int_b^a v(s)\, dF_s(s), \tag{25}$$

which is the product of the total mass of consumed (domestic and imported) varieties (19) and the expected utility from each good: the former depends on trade costs (as well as on size and income of all countries), but the latter depends only on preferences and the cost distribution.

Accordingly, trade liberalization affects welfare only through a change in the consumed varieties. More precisely, a reduction in trade costs reduces prices for each imported good in a less than proportional way due to incomplete pass-through and it does not affect the price of any domestic good (because the domestic cutoff cost does not change);[23] consumers exploit this by increasing the number of imported varieties without dropping any of the domestic varieties but consuming less of each. It is now clear that the higher is pass-through the more new imported varieties can be purchased, which increases the gains from trade liberalization.[24]

Our final aim is to derive a global quantitative measure of these gains from trade liberalization as in ACR (in spite of the non-homotheticity of our preferences). First of all, taking logs and differentiating (25) with respect to $\boldsymbol{\tau_j}$ and $\mathbf{w}' = [w_1, .., w_I]$ for a given E_j (we can always normalize wage changes in such a way that $d \ln w_j = 0$) we get:

$$d \ln V_j = d \ln n_j^C = \frac{-\kappa \sum_{i=1, i \neq j}^{I} n_{ij} \left(d \ln \tau_{ij} + d \ln w_i \right)}{n_j^C}, \tag{26}$$

where the last step exploits the differentiation of (19) with respect to $\boldsymbol{\tau_j}$ and $\mathbf{w}$ for a given E_j. Let us indicate the proportional change of a variable z from $\underline{z}$ to $\overline{z}$ as $\widehat{z} = \overline{z}/\underline{z}$. Integrating (26), we obtain that the proportional utility change $\widehat{V}_j$ due to a (possibly *large*) trade shock to $\boldsymbol{\tau} = [\boldsymbol{\tau_1}, .., \boldsymbol{\tau_I}]$ (and then to $\mathbf{w}$) is equal to the change $\widehat{n}_j^C$. In turn, through (22), the latter is simply related to the change in the domestic share λ_{jj} by $d \ln n_j^C = -d \ln \lambda_{jj}$ (i.e., $\widehat{n}_j^C = \widehat{\lambda}_{jj}^{-1}$).

Notice that, by (22) and (24), the changes in the domestic share λ_{jj} do not depend on the specific preferences. In fact, similarly to ACR, one can actually determine the impact of any reduction in trade costs by computing:[25]

$$\hat{w}_i Y_i = \sum_{j=1}^{I} \frac{\lambda_{ij} (\hat{w}_i \hat{\tau}_{ij})^{-\kappa}}{\sum_{k=1}^{I} \lambda_{kj} (\hat{w}_k \hat{\tau}_{kj})^{-\kappa}} \hat{w}_j Y_j \quad \text{and} \quad \hat{\lambda}_{jj} = \frac{(\hat{w}_j \hat{\tau}_{jj})^{-\kappa}}{\sum_{k=1}^{I} \lambda_{kj} (\hat{w}_k \hat{\tau}_{kj})^{-\kappa}}. \tag{27}$$

However, the specific preferences matter for translating these changes into a "quantitative" measure of the welfare gains from liberalization, which can be compared across models. In particular, here we want to derive the (proportional) variation of per-capita income in country

[23] Since domestic cutoffs do not change, this implies that no domestic firms exit during trade liberalization. Pavcnik (2002) finds that trade liberalization in Chile leads to domestic firm exit. Recently, Hsieh *et al.* (2016) document that domestic variety exit leads to significant welfare losses to Canada during the CUFTA. This mechanism is absent from our baseline model, but it can be restored by adding fixed market access costs to the model, albeit at the cost of loss of tractability in a multi-country environment. See footnote 33.

[24] Notice that the new imported varieties allow for the final equilibrium distribution of the (normalized) prices and markups to remain the same, therefore welfare changes in our setting can be measured only with changes in the number of imported varieties.

[25] Notice that the present model satisfies the requirements of what Allen *et al.* (2014) define as "universal gravity" and thus inherits their welfare results.

j, $\widehat{W}_j$, which is "equivalent" to the welfare change $\widehat{V}_j$ due to trade liberalization.

A suitable "money metric" is provided by the Equivalent Variation of income, EV_j, such that a consumer would be indifferent between the post-shock prices induced by the change of trade costs and the new income level $W_j = E_j + EV_j$ evaluated at pre-shock prices (see Varian, 1992, Par. 10.1), with proportional variation $\widehat{W}_j = W_j/E_j$. To understand how EV_j is computed, let us rewrite the equilibrium value of utility as:

$$V_j(W_j, E_j; F_{ij}) = \sum_{i=1}^{I} N_i \int_{bE_j}^{aW_j} v\left(\frac{p}{W_j}\right) dF_{ij}(p),$$

where F_{ij} is the unconditional distribution of prices p_{ij} posted by all firms of country i in country j (varieties with prices above the cut-off value aW_j are welfare irrelevant). By taking logs and differentiating the last expression with respect to W_j, one can get:

$$d\ln V_j = \frac{\kappa}{1 - \bar{\epsilon}_j^E\left(W_j, E_j\right)} d\ln W_j,$$

where $\bar{\epsilon}_j^E\left(W_j, E_j\right)$ is derived in Appendix C and captures the firm-sales-weighted average elasticity of price with respect to income, and it is thus related to the sales-weighted average pass-through and, ultimately, to the shape of the demand function. Therefore, the income variation that is equivalent to the impact of trade liberalization is implicitly determined by the solution of the equation:

$$\int_{E_j}^{W_j} \frac{\kappa}{1 - \bar{\epsilon}_j^E\left(t, E_j\right)} d\ln t = -\ln \widehat{\lambda}_{jj}. \tag{28}$$

To understand this formula, notice that a rather inelastic demand function (low elasticity with respect to price) implies that monopolistic firms set high prices and that prices react poorly to changes in costs and highly to changes in income ($\bar{\epsilon}_j^E\left(W_j, E_j\right)$ is high). Accordingly, high prices generate a high marginal utility of income (and low pass-through), which in turn reduces the income variation needed to match a given cost shock. As a consequence, low demand elasticity is associated with low welfare gains from trade liberalization. Instead, when the demand function becomes more elastic, cost reductions due to trade liberalization are shifted more into lower prices and the gains are higher.

We can approximate small changes in welfare defining $\bar{\epsilon}^E = \bar{\epsilon}_j^E\left(E_j, E_j\right) \in (0, 1)$ as the weighted average (with relative sales of firms as weights) of the elasticities of prices with respect to income, which is identical across countries. Under IA preferences, this is the complement to unity of the average pass-through $\bar{\epsilon}^c \in (0, 1)$. This allows to derive our main (local) result:

$$d\ln W_j = \bar{\epsilon}^c \frac{-d\ln \lambda_{jj}}{\kappa}, \quad \text{with } \bar{\epsilon}^c = 1 - \bar{\epsilon}^E \in (0, 1), \tag{29}$$

which shows that the welfare gains from trade liberalization are proportional to the average pass-through: intuitively, the lower is the pass-through on prices of reductions in trade costs, the lower must be the gains from trade liberalization.

Our result can be compared to those of a variety of traditional models. In particular, ACR have shown that a formula for the welfare gains as $d \ln W_j = -d \ln \lambda_{jj}/\sigma$, where σ is the "trade elasticity" of relative imports with respect to variable trade costs, applies to models different on the supply side but all based on *CES preferences* (as Anderson, 1979, Krugman, 1980, Eaton and Kortum, 2002, Melitz, 2003, Chaney, 2008 and others).[26] Thus, estimating such a trade elasticity allows one to measure the welfare gains from liberalization episodes. Notice that in models with homogenous firms, such as the Armington model (Anderson, 1979) and the Krugman (1980) model, the trade elasticity is related to the constant elasticity of substitution (namely $\sigma = \theta - 1$ in our notation): in these cases low substitutability between goods induces high imports of foreign varieties, which leads to high gains from trade liberalization. However, in a large class of heterogeneous firm models with an untruncated Pareto distribution of productivities, including the celebrated Melitz-Chaney model (Melitz, 2003; Chaney, 2008), ACR show that the trade elasticity σ is independent from preference parameters and just related to the shape of the Pareto distribution (namely $\sigma = \kappa$ in our notation), and therefore that the gains from trade liberalization are neutral with respect to the underlying model details.[27] This common result is based on the fact that all these models exhibit complete pass-through of cost reductions on prices due to CES preferences.

In a further generalization of the heterogenous firm models, ACDR confirm that the ACR welfare formula:

$$d \ln W_j = \frac{-d \ln \lambda_{jj}}{\kappa} \tag{30}$$

also applies to some prominent examples of homothetic preference. As already noticed by Arkolakis *et al.* (2010) and Feenstra (2014), in these cases trade liberalization induces consumers to replace the most expensive domestic goods with an identical number of cheaper imported varieties, which excludes gains from variety associated with trade. In the terminology of ACDR, reductions in marginal costs due to trade liberalization are here the only source of gains.[28] The reason is that a reduction in trade costs exerts two effects on markups that balance each other out as if pass-through was full: on the one hand, inframarginal exporting firms tend to increase

[26] Also see the results of Atkeson and Burstein (2010).

[27] A similar result applies even to Ricardian models as the one by Eaton and Kortum (2002), which adopts a Fréchet distribution of productivities. For a recent empirical investigation of the gains from variety in this class of models see Hsieh *et al.* (2016).

[28] Bertoletti and Etro (2016) show that, in homogenous firms models of monopolistic competition with free entry, homothetic preferences generate complete pass-through of changes in the marginal cost and neutrality on the endogenous number of consumed goods.

their markups due to incomplete pass-through, but on the other hand, the reduction of the choke price (which creates a selection effect on the set of domestic firms) tends to reduce the markups of all firms. The welfare formula (30) applies also globally for the simple reason that the marginal utility of income can be normalized to be independent from income under homothetic preferences (as is the set of purchased varieties for given prices).

ACDR also consider the case of directly additive preferences. They derive the following local approximation (valid only for small welfare changes):

$$d \ln W_j = \left(1 - \frac{\rho}{\kappa + 1}\right) \frac{-d \ln \lambda_{jj}}{\kappa}, \tag{31}$$

where ρ, a weighted average (with relative sales as weights) of the elasticity of markups to productivity, is positive but smaller than unity in common models with incomplete pass-through. In this case, by reducing the choke price, trade liberalization not only creates a selection effect, but it also affects the equilibrium distribution of prices thus increasing the measure of consumed goods (while leaving unchanged the distribution of individual consumption levels). As discussed by ACDR, the gains from the reduction of the choke price compensate only in part the losses due to the increase in markup on imported varieties, leading to smaller gains compared to homothetic preferences. However, the difference is quite limited since $(1 - \frac{\rho}{\kappa+1}) \in (\frac{\kappa}{\kappa+1}, 1)$ when $\rho \in (0, 1)$.

Going back to our IA case, it is now clear why the gains from trade liberalization can be much smaller than in ACR or ACDR. A reduction in trade costs increases the markups of the inframarginal exporting firms due to incomplete pass-through without any counteracting forces because of the absence of selection effects. Accordingly, the average pass-through $\bar{\epsilon}^C \in (0, 1)$ is critical in determining the welfare gains in (29). When demand is very elastic, a high pass-through generates high gains from trade cost reductions (up to the ACR level in the limit case of perfectly elastic demand or full pass-through). Instead, when the demand is rather inelastic, pass-through is low and the gains are limited.

3.3 A specific functional form

In the remainder of the paper and for our quantitative analysis we adopt the following convenient specification of IA preferences:

$$V = \int_{\Omega} \frac{(a - s(\omega))^{1+\gamma}}{1 + \gamma} d\omega. \tag{32}$$

Here $\gamma \in (0, \infty)$ is the key preference parameter. By Roy's identity, the demand for each variety ω is:

$$x(\omega) = \frac{(a - s(\omega))^{\gamma}}{|\mu|}. \tag{33}$$

The elasticity of demand with respect to price is $\theta(s) = \frac{\gamma s}{a-s}$, which is increasing in the price. Demand is actually linear (as in the example of Section 2) for $\gamma = 1$, it tends to become perfectly elastic for $\gamma \to \infty$ and perfectly rigid for $\gamma \to 0$. The rest of the model is the same as above. We can summarize the relevant exogenous variables/parameters in our setting by the objects $\tilde{\mathbf{P}} = \{a, \kappa, \gamma, \boldsymbol{\tau}, \mathbf{e}, \mathbf{L}, F_e\}$ in matrix notation, where $\mathbf{e}'= [e_1, .., e_I]$ and $\mathbf{L}'= [L_1, .., L_I]$.

The optimal price of a c-firm from country i willing to sell to country j is easily derived as:

$$p_{ij}(c) = \frac{\gamma \tau_{ij} w_i c + a E_j}{1 + \gamma}, \tag{34}$$

which shows that the degree of pass-through is increasing in γ. Indeed, for $\gamma \to 0$ any reduction in costs would be exploited by the firms without price reduction (prices would approach the limit aE_j with full expropriation of consumer welfare), while for $\gamma \to \infty$ any reduction in costs would be fully translated into a price reduction (prices would approach the nominal marginal cost $\tau_{ij} w_i c$ as in perfect competition). The value of the sales of a c-firm from country i to country j is:

$$t_{ij}(c) = \frac{\gamma^{\gamma} (\gamma c + \widehat{c}_{ij}) (\widehat{c}_{ij} - c)^{\gamma} (\tau_{ij} w_i)^{1+\gamma} L_j}{(1+\gamma)^{1+\gamma} (E_j)^{\gamma} |\mu_j|}, \tag{35}$$

while the corresponding profits are given by:

$$\pi_{ij}(c) = \frac{\gamma^{\gamma} (\widehat{c}_{ij} - c)^{1+\gamma} (\tau_{ij} w_i)^{1+\gamma} L_j}{(1+\gamma)^{1+\gamma} E_j^{\gamma} |\mu_j|}, \tag{36}$$

and are a decreasing and convex function of c.

3.3.1 Markups, prices and sales

We now derive the model's key prediction regarding markup and price variation across destinations and across firms. Denote by $m_{ij}(c)$ the mark-up that a firm with cost draw c from country i enjoys in destination j (assuming that it actually serves that market, i.e. $c \leq \hat{c}_{ij}$):

$$m_{ij}(c) = \left(\frac{1}{1+\gamma} \right) \left(\frac{\widehat{c}_{ij} - c}{c} \right). \tag{37}$$

This markup is decreasing in γ, reflecting a more elastic demand, and rising in the cost cutoff $\hat{c}_{ij}$, reflecting pricing to market. Moreover, more productive firms set lower prices but enjoy higher markups.

Furthermore, from (34), the elasticity of prices with respect to the "intrinsic" marginal cost c (or the transport cost τ_{ij}, or the wage of the source country w_i) can be expressed as:

$$\epsilon_{ij}^{c}(c) = \frac{\gamma c}{\gamma c + \widehat{c}_{ij}} \in \left[0, \frac{\gamma}{1+\gamma}\right]. \tag{38}$$

Similarly, the elasticity of prices with respect to income of the destination country E_j is its complement to one:

$$\epsilon_{ij}^{E}(c) = \frac{\widehat{c}_{ij}}{\gamma c + \widehat{c}_{ij}} \in \left[\frac{1}{1+\gamma}, 1\right]. \tag{39}$$

It is easy to verify that the latter is also the elasticity of prices with respect to the real exchange rate between the source and the destination country, which is often the subject of empirical investigations.[29]

Both the degrees of pass-through $\epsilon_{ij}^{c}(c)$ and pricing to market $\epsilon_{ij}^{E}(c)$ vary with a firm's productivity in a monotonic way. For instance, pass-through is zero for the most efficient firms ($\epsilon_{ij}^{c}(0) = 0$) and the highest for the least efficient firms ($\epsilon_{ij}^{c}(\widehat{c}_{ij}) = \gamma/(1+\gamma)$). Pricing-to-market in turn is as high as 1 for the most productive firm and only $1/(1+\gamma)$ for the least productive one. These differences are due to the fact that efficient firms set their prices low and change them mainly on the basis of changes in income, while inefficient firms set high prices and change them mainly on the basis of changes in costs. These predictions are in line with empirical evidence provided by Berman *et al.* (2012) that pricing to market is more sensitive for more productive firms.[30]

Finally, let us consider the reaction of the sales of a firm of country i toward country j when the relevant trade cost decreases. We know from (38) that such a liberalization reduces prices $p_{ij}(c)$ more for the small (high-c) firms. As a consequence, these small firms increase more their production, as could be verified by evaluating the elasticity of quantity $x_{ij}(c) = (a - p_{ij}(c)/E_j)^{\gamma} / |\mu|$. Whether the sales $t_{ij}(c)$ of small firms are more or less reactive is not obvious. However, computing the elasticity $d \ln t_{ij}(c)/d \ln \tau_{ij}$ (after normalizing $d \ln w_i = 0$)[31]

[29] It is straightforward to extend the model to feature nominal and real exchange rates (see Bertoletti *et al.*, 2016).

[30] Using detailed French exporter data, these authors find that the exporter with average productivity raises prices by 0.8% when experiencing a 10% home currency depreciation. Furthermore, the response is 1.3% for exporters with a productivity level equal to the mean plus one standard deviation, namely, for more productive exporters.

[31] Taking logs of p_{ij} and x_{ij} and differentiating we get:

$$\frac{\partial \ln t_{ij}(c)}{\partial \ln \tau_{ij}} = \frac{\partial \ln x_{ij}(c)}{\partial \ln \tau_{ij}} + \frac{\partial \ln p_{ij}(c)}{\partial \ln \tau_{ij}} = 1 - \frac{\partial \ln |\mu_j|}{\partial \ln \tau_{ij}} + \left[\frac{\widehat{c}_{ij}}{\widehat{c}_{ij} + \gamma c} + \gamma \frac{c}{\widehat{c}_{ij} - c}\right] \frac{\partial \ln \widehat{c}_{ij}}{\partial \ln \tau_{ij}}.$$

and differentiating it with respect to c we obtain:

$$\frac{\partial}{\partial c}\left\{\frac{\partial \ln t_{ij}(c)}{\partial \ln \tau_{ij}}\right\}=\left[\frac{-\gamma\widehat{c}_{ij}}{\left(\widehat{c}_{ij}+\gamma c\right)^{2}}+\frac{\gamma\widehat{c}_{ij}}{\left(\widehat{c}_{ij}-c\right)^{2}}\right]\frac{\partial \ln \widehat{c}_{ij}}{\partial \ln \tau_{ij}}<0,$$

where $\partial \ln \widehat{c}_{ij}/\partial \ln \tau_{ij}=\partial \ln w_j/\partial \ln \tau_{ij}-1<0$. Since a reduction of τ_{ij} corresponds to trade liberalization, this shows that smaller firms respond more to trade liberalization, which is in line with the evidence presented by Eaton *et al.* (2008) and Arkolakis (2016). Together with the fact that trade liberalization induces entry of foreign varieties in our model, this implies that adjustments on the extensive margin (changes to new and least traded varieties) are critical in understanding the welfare gains from trade (as argued by Broda and Weinstein, 2006, and Kehoe and Ruhl, 2013).

3.3.2 Equilibrium distributions

Given our functional form, we can fully characterize the equilibrium in closed form (see Appendix B). The distribution of normalized prices on the support $[\frac{a}{1+\gamma}, a]$ can be derived as:

$$F_s(s)=\left(\frac{(1+\gamma)\,s}{\gamma a}-\frac{1}{\gamma}\right)^{\kappa}, \tag{40}$$

which depends only on the three parameters γ, κ and a. Analogously, prices in country j, given by expression (34), are distributed according to $F_j(p)=[(1+\gamma)\,p/\gamma aE_j-1/\gamma]^{\kappa}$, which is independent from trade costs and the identities of the exporting countries, but depends crucially on the income of the importing country j. The markup distribution can be derived as follows:

$$F_m(m)=1-\frac{1}{[1+(1+\gamma)\,m]^{\kappa}}, \tag{41}$$

which is also the same across countries.

The expected profit and expected sales of a firm from country i selling in country j can be expressed as (see Appendix B):

$$\mathbb{E}\left\{\pi_{ij}\right\}=\frac{a^{\gamma+1}\gamma^{\gamma}\kappa\widehat{c}_{ij}^{\kappa}B(\kappa,\gamma+2)E_jL_j}{(1+\gamma)^{1+\gamma}\overline{c}^{\kappa}\left|\mu_j\right|}\quad\text{and}\quad\mathbb{E}\left\{t_{ij}\right\}=\frac{a^{\gamma+1}\gamma^{\gamma+1}\widehat{c}_{ij}^{\kappa}B(\kappa+2,\gamma)E_jL_j}{(1+\gamma)^{\gamma}\overline{c}^{\kappa}\left|\mu_j\right|},$$

where $B(z,h)=\int_0^1 t^{z-1}(1-t)^{h-1}dt$ is the Euler Beta function.[32] Using its properties expression

[32] Its value is also given by:

$$B(z,h)=\frac{\Gamma(z)\Gamma(h)}{\Gamma(z+h)},$$

where $\Gamma(t)$ is the Euler Gamma function (see Appendix B). Its basic recursive properties are given by $B(z+$

(16) yields a value for $\bar{\theta}$ of:

$$\bar{\theta} = \kappa + 1, \tag{42}$$

which is independent from the preference parameters; this implies that the return on sales (16) is uniquely determined by the shape parameter of the Pareto distribution. Moreover, it allows us to solve for the number of firms created in country i in closed form as $N_i = e_i L_i / [(\kappa+1) F_e]$, which is the same as in autarky (both in the decentralized equilibrium and in the social optimum).

The extensive margin $n_{ij} = N_i G(\hat{c}_{ij})$ is independent from population and demand elasticity.[33] The number of goods consumed in country j, n_j^C (see (19)) is independent from the preference parameter γ, while it increases in the willingness to pay for each good, a.[34] Finally, we can also evaluate the market share in country j of an exporting c-firm from country i, $\alpha_{ij}(c) \equiv t_{ij}(c)/\bar{t}_{ij}$. This can be expressed as:

$$\alpha_{ij}(c) = \frac{\left(1 + \gamma \frac{c}{\hat{c}_{ij}}\right)\left(1 - \frac{c}{\hat{c}_{ij}}\right)^{\gamma}}{\gamma(1+\gamma)B(\kappa+2,\gamma)}, \tag{43}$$

and it can be verified that also the distribution of the market share is identical across countries and depends only on the two parameters γ and κ.[35] This result demonstrates an attractive feature of this framework relative to alternatives: the distribution of (normalized) firm sales is not uniquely tied to the distribution of firm productivities. Given a certain degree in productivity dispersion, governed by κ, the dispersion in firm sales is pinned down by γ.[36] Hence, the model can potentially reconcile both the measured productivity and sales advantages of exporters over

$1, h) = zB(z,h)/(z+h)$ and $B(z, h+1) = hB(z,h)/(z+h)$.

[33]Notice that our model can generate an extensive margin that is increasing in population simply by adding small fixed export costs. If these are in units of local labor, say F_j, it is easy to derive from (36) the modified cutoff:

$$\hat{c}_{ij} = \frac{aE_j}{\tau_{ij} w_i}\left[1 - \frac{1+\gamma}{a}\left(\frac{|\mu_j| F_j}{\gamma^{\gamma} e_j L_j}\right)^{\frac{1}{1+\gamma}}\right]$$

so that the extensive margin is directly increasing in the destination population L_j. Notice that the same extension induces selection effects on the measure of domestic firms through the negative impact of the price aggregator $|\mu_i|$ on the domestic cutoff $\hat{c}_{ii}$, without affecting the pricing formula (34).

[34]We can also compute $|\mu_j| = \frac{a^{\gamma+1}\gamma^{\gamma+1}B(\kappa+2,\gamma)}{(1+\gamma)^{\gamma}} n_j^C$ as a linear function of the number of consumed varieties.

[35]The crucial element is the distribution of $l = H(t) = (1+\gamma t)(1-t)^{\gamma}$, where $t = c/\widehat{c}_{ij}$ with $F_t(t) = \Pr\{c/\widehat{c}_{ij} \leq t\} = t^{\kappa}$ on the support $[0,1]$. Notice that $H'(t) < 0$ and $H''(t) < 0$ if and only if $t < 1/2$, therefore l is distributed on $[0,1]$ according to $F_l(l) = 1 - (H^{-1}(l))^{\kappa}$.

[36]Jung *et al.* (2015) demonstrate that the distributions of firm sales and productivity depend uniquely on the Pareto productivity shape parameter in existing models that feature consumers with directly additive preferences, including quadratic preferences (as in Melitz and Ottaviano, 2008, but without the outside good) and those of Behrens *et al.* (2014) and Simonovska (2015). Therefore, these models cannot jointly reconcile moments from the two distributions observed in US data. The authors outline a flexible, albeit not tractable, extension of Simonovska (2015) that falls within the directly-additive class and has more desirable quantitative features.

non-exporters reported by Bernard *et al.* (2003).

3.3.3 Welfare

Our specification of the indirect utility allows us to characterize the impact of trade in detail. The equilibrium value of utility in country j is now:

$$V_j = n_j^C \frac{(a\gamma)^{\gamma+1} B(\kappa, \gamma+2)\,\kappa}{(\gamma+1)^{\gamma+2}}, \tag{44}$$

which is linear in the number of consumed goods, depends on the willingness to pay a, on the preference parameter γ, which governs the level of market competitiveness/pass-through, and on κ, which governs the cost distribution.

As in the general model (see Appendix C), we calculate the equivalent variation on income EV_j keeping prices unchanged at their initial level before the trade shock.[37] Notice that the distribution F_{ij} of prices posted by all firms from country i in country j can be expressed as:

$$F_{ij}(p) = \left(\frac{(1+\gamma)\,p - aE_j}{\gamma\tau_{ij}w_i\overline{c}}\right)^{\kappa} \tag{45}$$

on the interval $[\underline{p}, \overline{p}]$, where $\underline{p} = p(0) = aE_j/(\gamma+1)$ and $\overline{p} = p(\overline{c}) = (\gamma\tau_{ij}w_i\overline{c} + aE_j)/(\gamma+1)$. Taking logs of $V_j(W_j, E_j; F_{ij})$ and differentiating with respect to W_j, we obtain:

$$d\ln V_j = \frac{(\gamma+1)(\kappa W_j + E_j)}{(\gamma+1)W_j - E_j} d\ln W_j. \tag{46}$$

For "small" income changes (i.e., evaluating the previous differential at $W_j = E_j$), we obtain the approximation $d\ln V_j = (\gamma+1)(\kappa+1)\,d\ln W_j/\gamma$. Recalling that gravity implies that $n_j^C = a^{\kappa}e_j^{\kappa}N_j/\overline{c}^{\kappa}\lambda_{jj}$, and that a shock to trade costs causes a proportional change of utility denoted by $d\ln V_j = d\ln n_j^C = -d\ln\lambda_{jj}$, this immediately delivers the local measure:

$$d\ln W_j = -\frac{\gamma d\ln\lambda_{jj}}{(\gamma+1)(\kappa+1)}, \tag{47}$$

whose coefficient is in the range $(0, \frac{1}{\kappa+1})$ for $\gamma \in (0,\infty)$. Notice that the upper bound of this range is the lower bound of the range obtained by ACDR for directly additive preferences in (31). In general, the model implies gains from any liberalization experiment that are approximately

[37] The optimal prices of the varieties unsold in country j are not uniquely defined above the cutoff price aE_j because demand and profits are zero. For the sake of simplicity (and to avoid any asymmetry between positive and negative equivalent variations), we assume that they follow the same pricing rule (34) as the varieties actually sold.

$\frac{\gamma\kappa}{(\gamma+1)(\kappa+1)}$ of the ACR gains in (30).

The global measure of the gains from trade liberalization, $\widehat{W}_j$, valid also for "large" trade shocks, can be obtained by integrating (46), which implicitly characterizes W_j as follows:

$$\int_{E_j}^{W_j} \frac{(\gamma+1)(\kappa t + E_j)}{(\gamma+1)t - E_j} d\ln t = -\ln \widehat{\lambda}_{jj}, \tag{48}$$

where we can further compute:

$$\int_{E_j}^{W_j} \frac{(\gamma+1)(\kappa t + E_j)}{(\gamma+1)t - E_j} \frac{dt}{t} = \left[(\gamma+\kappa+1)\ln(\gamma t + t - E_j) - (\gamma+1)\ln t\right]_{E_j}^{W_j}.$$

The approximation derived from (47) provides a lower bound for the exact welfare changes in (48),[38] therefore it can be used as a conservative measure of the benefits of trade liberalization. The welfare gains from trade liberalization depend on κ, as in ACR, but also on the preference parameter $\gamma \in (0, \infty)$, which governs pass-through and competitiveness in the markets. Accordingly, for given values of the Pareto shape parameter κ, the gains from trade liberalization are larger in more competitive markets with higher pass-through (higher γ).[39]

4 Quantifying the gains from trade

In this section, we quantify the gains from trade predicted by the parameterized IA model and we compare the results to those that arise form a benchmark model that relies on homothetic preferences.

4.1 Identification strategy

In order to quantify the welfare gains from trade predicted by the model, we need data on trade shares (and per-capita income in the case of large trade shocks) as well as estimates of two key parameters, κ and γ. There are a number of estimation strategies that would allow us to identify these two parameters.

In this section, we follow a parsimonious calibration strategy, which allows us to identify the two parameters of interest, compute the welfare gains from trade predicted by the model, and compare them to those predicted by standard homothetic models of trade. The starting point

[38]This follows from the concavity in W_j of the function on the left hand side of (48).

[39]Notice that a more competitive environment implies lower prices and higher pass-through, which in turn requires a larger equivalent income variation due to the decreasing marginal utility of income. Similarly, the exact income variation equivalent to a positive trade shock is larger when we take into account the decreasing marginal utility of income.

is the observation that the model falls within a large class of models that generate a log-linear gravity equation of trade. To derive the theoretical gravity equation of trade, take the log of the ratio of country j's import share from source i, λ_{ij} in expression (22), and j's domestic expenditure share, the corresponding expression for λ_{jj}, which yields

$$\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = \tilde{S}_i - \tilde{S}_j - \kappa \log \tau_{ij}, \tag{49}$$

where $\tilde{S}_i = \log(e_i L_i w_i^{-\kappa})$ for all $i = 1, .., I$.

From this expression κ can be interpreted as the partial elasticity of trade flows with respect to variable trade costs. Crucially, since homothetic (and non-homothetic models) studied by ACR and ACDR yield an identical gravity equation of trade, if the parameter could be estimated using the gravity moment alone, then these models would have to yield identical estimates of κ.

Estimating this parameter using the gravity equation alone is challenging and has been the focus of many papers—see Simonovska and Waugh (2014a,b) and Caliendo and Parro (2015) for recent contributions and a discussion of the related literature. Such a task is beyond the scope of this paper. Instead, for the purposes of the welfare exercise, we let $\kappa = 5$, which is the preferred estimate of ACR and ACDR, and approximately the average estimate of the trade elasticity across sectors obtained by Caliendo and Parro (2015). Since the gains from trade in homothetic models are entirely driven by the value of this parameter, this choice also sets a useful quantitative benchmark and allows us to relate our findings closely to those of the existing literature.

The parameter γ is a demand-side parameter in our model—it governs the elasticity of demand with respect to price as well as the elasticity of price with respect to income and the pass-through elasticity. Consequently, the parameter is at the heart of the model's pricing predictions, which are the key departure from the standard ACR framework.

Recall that a key testable prediction of our model relates to cross-country price variation. Prices should be increasing in destination per-capita income and independent of destination population size: $\partial p_{ij}/\partial E_j > 0$ and $\partial p_{ij}/\partial L_j = 0$. In the absence of data on firms' costs, the expected values of the elasticities of prices are of interest so as to be able to compare them with corresponding moments in the data. In particular, we can derive an explicit expression for the average elasticity of price with respect to per-capita income:

$$\mathbb{E}\left\{\epsilon^E\right\} = F_{2,1}(1, \kappa; 1 + \kappa; -\gamma), \tag{50}$$

where $F_{2,1}$ is the hypergeometric function defined in Appendix B. Hence, with a value of κ in hand, the average elasticity of price with respect to income identifies γ. Simonovska (2015) provides an estimate of this elasticity that amounts to 0.14. Given $\kappa = 5$, this estimate implies $\gamma = 7.62$ according to equation (50). Armed with these parameters, we can proceed to quantify the welfare gains form trade.

4.2 Welfare gains from trade

How do the welfare gains from trade compare to existing frameworks? First of all, a comparison of our approximate formula for the welfare gains (47) and the formula (30) that holds for CES and homothetic preferences provides an immediate "back of the envelope" calculation. Given $\kappa = 5$ and $\gamma = 7.62$, the IA model implies gains from any liberalization experiment that are approximately $\frac{\gamma\kappa}{(\gamma+1)(\kappa+1)} = 73.7\%$ of the ACR gains. Hence, the mismeasurement of welfare is quantitatively large (our model yields welfare gains that are almost 30% lower than ACR's gains).

We make this insight more precise looking at the global formula for the welfare gains (48) below. Specifically, we compute the welfare gains of moving from autarky to the observed trade share for each of 123 countries in year 2004 as predicted by the ACR framework for the CES model and by our IA model.[40] For the first case, we let $\kappa = 5$ and we use the formula (30) to arrive at the welfare measure. For our model, we let $\kappa = 5$ and use the welfare measure for large shocks (48) with $\gamma = 7.62$. For both models, we use the domestic expenditure shares (1-trade share) before and after the shock. In the case of the IA model, the welfare gains due to a global shock require values for per-capita income, and we let those correspond to the values before the shock. In this particular exercise, the domestic expenditure share goes from the observed share in the data to unity (autarky).

The left panel of Figure 1 plots the results for the ACR framework based on CES preferences, while the right shows the predictions of our IA model. The differences are significant. The average country enjoys a 14% welfare gain from trade according to the ACR framework. In contrast, our IA model yields a mean value of 10%, about 70% of that predicted by the homothetic model. While the ranking of countries according to welfare gains is identical in the two models, the dispersion in the ACR framework is much larger than the one predicted by the IA model.

To further understand the difference in magnitudes that we obtain relative to the literature, we focus attention on the US. ACR report that, for a US domestic expenditure share of 0.93 and a value for the Pareto shape parameter of 5-10, the welfare gains of moving away from autarky range from 0.7% to 1.4%. First, we point out that, in our database, for year 2004,

[40]We describe all data in Appendix E.

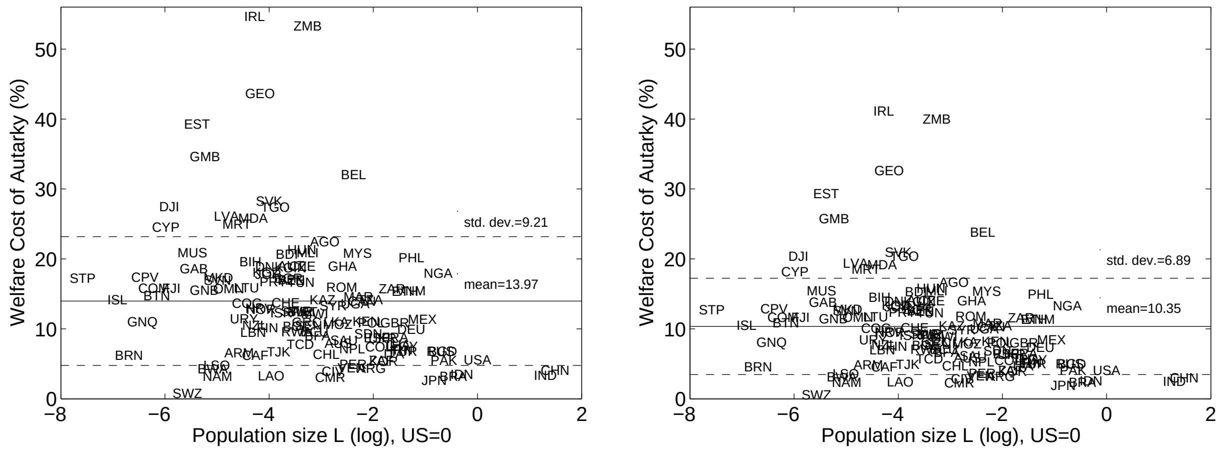

Figure 1: Welfare Cost of Autarky in CES and IA Models, 123 Countries

we obtain a US domestic expenditure share on *manufacturing*, adjusted for trade imbalances, of 0.75. Thus, we cannot directly compare the estimate for the US reported in the plot above to the estimate reported in ACR. Second, let us assume that indeed the domestic expenditure share is 0.93 as reported in ACR and that κ is 5. Then, the welfare gains from trade for the US predicted by the ACR framework would amount to 1.4%. In contrast, the welfare gains predicted by the IA model would amount to 1.07%. Thus, it is reasonable to conclude that the welfare gains from trade may be substantially overestimated, should one rely on preferences featuring full pass-through. It is important to remark that our findings also differ from those of the ACDR's estimate of a model based on directly additive preferences, since they obtain a measure of welfare gains that is 96% of the value predicted by the standard ACR framework.

4.3 Alternative identification strategies and the impact on welfare

Given the role that κ and γ play in governing the welfare gains from trade in our model, we offer several alternative identification strategies for these two parameters and we discuss the impact on welfare.

First, in Appendix D, we outline a strategy that uses commonly studied firm-level moments in the international trade literature.[41] In particular, we present the results from an overidentified estimation of γ and κ where we jointly target the following three moments: (i) the average elasticity of price with respect to income from Simonovska (2015) as described above, (ii) the domestic sales advantage of exporters over non-exporters for the US, and (iii) the measured productivity advantage of exporters over non-exporters for the US (see e.g. Bernard *et al.*, 2003). Since κ governs the underlying dispersion in firms' production costs, while γ controls the

[41] For a good review of quantitative trade models see Kehoe, Pau and Rossbach (2016).

demand side of the economy thus linking costs to equilibrium prices and quantities, the last two moments provide additional information regarding the value of the two parameters. Applying the Simulated Method of Moments (SMM) estimator delivers $\kappa = 2.77$ and $\gamma = 1.92$, which implies a slightly higher dispersion of costs relative to the baseline calibration and a demand function that is still convex but closer to the linear demand benchmark. In this exercise, we also evaluate the quantitative performance of the model along a number of firm-level dimensions such as the degree of cost past through, the level of markups, the fraction of exporters, and export intensity, as well as aggregate ones such as the intensive and extensive margins of trade.

The estimates of the two parameters generate lower gains from trade liberalization for the IA model. Indeed, using the approximation (47), the gains from a liberalization experiment are at most $\frac{\gamma\kappa}{(\gamma+1)(\kappa+1)} = 48.3\%$ of the ACR gains. We have obtained similar quantitative results in Bertoletti *et al.* (2016), where we have retained our calibrated value $\kappa = 5$ and we have estimated the preference parameter to match moments (i) and (ii) above, which delivers a demand function that is approximately linear ($\gamma \approx 1$).

Second, given a value for κ, one may refer to moments reported by the pass-through literature to identify γ (see e.g. Amiti *et al.*, 2014). The intuition behind such a strategy once again relies on the fact that the demand-side parameter γ governs equilibrium objects such as prices, which for given product costs, contain information about firms' markups and degree of pass-through. Third, one may arrive at a value for γ by directly estimating the demand system using micro-level data on prices and quantities at the firm- or product-level (see e.g. ACDR, Broda and Weinstein, 2006, and Feenstra and Weinstein, 2016). For a given value of κ, alternative estimates of γ would shrink or widen the gap in predicted welfare between the homothetic and the IA model. In particular, lower values of γ generate lower welfare gains in the IA model and therefore widen the difference in the predicted welfare levels between the two classes of models. We leave it for future research to exploit rich micro-level data and estimate the welfare gains from different liberalization episodes.

5 Conclusion

The contribution of this work is to introduce non-homothetic IA preferences to the literature on multicountry trade with heterogenous firms, quantify the welfare gains from trade under this class of preferences showing that it can be largely different from most alternative models, and propose a parametric specification that is highly tractable and useful for quantitative work. The model avoids the pervasive markup neutralities emerging in the CES model (Melitz, 2003) and the limits of quasilinear preferences in general equilibrium applications (Melitz and Ottaviano, 2008). Between variable markup models, this is the only one able to jointly deliver prices increasing in destination income, independent from population of the destination country and characterized by incomplete pass-through, with variable elasticities for firms of different productivity. Moreover, the model has novel implications for the extensive and intensive margins of trade that appear promising in front of the limited evidence. The implication of such a model for the gains from trade liberalization, however, is our main result: these gains can be much lower than those implied by the models based on homothetic or directly additive preferences analyzed in ACR and ACDR.

Our setting could be usefully extended to consider strategic interactions (Atkeson and Burstein, 2008 and Etro, 2015), heterogenous consumers and income distribution (for the case of identical firms see Bertoletti and Etro, 2017) and quality differentiation (Fajgelbaum *et al.*, 2011), more general preferences,[42] endogenous labor supply, and a 2x2x2 model with an outside good sold in a perfectly competitive setting to study the interplay with inter-industry trade. Our tractable non-homothetic preferences could also be exploited for dynamic analysis of structural change and business cycles.

References

Anderson, J. E., 1979, A Theoretical Foundation for the Gravity Equation, *American Economic Review*, 69, 106-16

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

Allen, T., C. Arkolakis and Y. Takahashi, 2014, Universal Gravity, mimeo, Northwestern University

Alvarez, F. and R. J. Lucas, 2007, General Equilibrium Analysis of the Eaton-Kortum Model of International Trade, *Journal of Monetary Economics*, 54, 1726–1768

Amiti, M., O. Itskhoki and J. Konings, 2014, Importers, Exporters, and Exchange Rate Disconnect, *American Economic Review*, 104, 7, 1942-78

Arkolakis, C., 2016, A Unified Theory of Firm Selection and Growth, *Quarterly Journal of Economics*, 131, 1, 89-155

Arkolakis, C., A. Costinot and A. Rodriguez-Clare, 2010, Gains From Trade under Monopolistic Competition: A Simple Example with Translog Expenditure Functions and Pareto Distributions of Firm-Level Productivity, mimeo, Yale University

Arkolakis, C., A. Costinot and A. Rodriguez-Clare, 2012, New Trade Models, Same Old Gains?, *American Economic Review*, 102, 1, 94-130

[42] For an analysis of monopolistic competition with general (non-additive and non-homothetic) preferences see Bertoletti and Etro (2016).

Arkolakis, C., A. Costinot, D. Donaldson and A. Rodríguez-Clare, 2015, The Elusive Pro-Competitive Effects of Trade, WP 21370, NBER

Arkolakis, C., S. Demidova, P. J. Klenow, and A. Rodriguez-Clare, 2008, Endogenous Variety and the Gains from Trade, *American Economic Review*, 98, 2, 444-50

Atkeson, A. and A. Burstein, 2008, Pricing-to-Market, Trade Costs, and International Relative Prices, *American Economic Review*, 98, 5, 1998-2031

Atkeson, A. and A. Burstein, 2010, Innovation, Firm Dynamics, and International Trade, *Journal of Political Economy*, 118, 3, 433-84

Bastos, P. and J. Silva, 2010, The Quality of a fFirm's Exports: where you export to matters, *Journal of International Economics*, 82, 2, 99-111

Behrens, K., G. Mion, Y. Murata and J. Südekum, 2014, Trade, Wages, and Productivity, *International Economic Review*, 55, 4, 1305-48

Berman, N., P. Martin and T. Mayer, 2012, How Do Difference Exporters React to Exchange Rate Changes?, *Quarterly Journal of Economics*, 127, 437-492

Bernard, A. B., J. Eaton, B. Jensen and S. S. Kortum, 2003, Plants and Productivity in International Trade, *American Economic Review*, 93, 4, 1268-90

Bernard, A. B., J. B. Jensen, S. J. Redding and P. K. Schott, 2007, Firms in International Trade, *Journal of Economic Perspectives*, 21, 3, 105-30

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, 1, 283-313

Bernard, A. B., S. J. Redding, and P. K. Schott, 2011, Multiproduct Firms and Trade Liberalization, *Quarterly Journal of Economics*, 126, 3, 1271-1318

Bertoletti, P. and P. Epifani, 2014, Monopolistic Competition: CES Redux?, *Journal of International Economics*, 93, 2, 227-38

Bertoletti, P. and F. Etro, 2016, Preferences, Entry and Market Structure, *RAND Journal of Economics*, 47, 4, 792-821

Bertoletti, P. and F. Etro, 2017, Monopolistic Competition when Income Matters, *Economic Journal*, in press

Bertoletti, P., F. Etro, and I. Simonovska, 2016, International Trade with Indirect Additivity, NBER Working Paper N. 21984

Broda, C., E. Leibtag, and D. E. Weinstein, 2009, The Role of Prices in Measuring the Poor's Living Standards, *Journal of Economic Perspectives*, 23, 2, 77-97

Broda, C. and D. E. Weinstein, 2006, Globalization and the Gains From Variety, *Quarterly journal of Economics*, 121, 2, 541-85

Caliendo, L. and F. Parro, 2015, Estimates of the Trade and Welfare Effects of NAFTA, *Review of Economic Studies*, 88, 1, 1-44

Campbell, J. and H. Hopenhayn, 2005, Market Size Matters, *Journal of Industrial Economics*, 53, 1, 1-25

Caron, J., T. Fally and J. Markusen, 2014, International Trade Puzzles: A Solution Linking Production and Preferences, *Quarterly Journal of Economics*, 129, 3, 1501-52

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, 2015, Prices, Markups and Trade Reforms, *Econometrica*, forthcoming

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

Dhingra, S. and J. Morrow, 2014, The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition, mimeo, L.S.E.

Dingel, J. I., 2015, The Determinants of Quality Specialization, mimeo, Chicago Booth.

Dixit, A. and J. Stiglitz, 1977, Monopolistic Competition and Optimum Product Diversity, *American Economic Review*, 67, 297-308

Eaton, J., M, Eslava, M. Kugler, and J. Tybout, 2008, The Margins of Entry into Export Markets: Evidence from Colombia, in Elhanan Helpman, Dalia Marin, and Thiery Verdier Eds., *The Organization of Firms in a Global Economy*, Cambridge, MA: Harvard University Press

Eaton, J. and S. Kortum, 2002, Technology, Geography, and Trade, *Econometrica*, 70, 5, 1741-9.

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

Etro, F., 2015, Endogenous Market Structures and International Trade, *Scandinavian Journal of Economics*, 117, 3, 918-56

Etro, F., 2016, Macroeconomic with Endogenous Markups and Optimal Taxation, Ca' Foscari University of Venice, Dept. of Economics WP 32

Fajgelbaum, P., G. Grossman and E. Helpman, 2011, Income Distribution, Product Quality, and International Trade, *Journal of Political Economy*, 119, 4, 721-65

Feenstra, R., 2003, A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity, *Economics Letters*, 78, 1, 79-86

Feenstra, R. C., 2014, Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity, WP 19833, NBER

Feenstra, R. C., R. Inklaar, and M. Timmer, 2013, The Next Generation of the Penn World Table, WP 19255, NBER

Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo., 2005, World Trade Flows: 1962-2000, WP 11040, NBER

Feenstra, R. C. and J. Romalis, 2014, International prices and endogenous quality, *Quarterly Journal of Economics*, 129, 2, 477-527

Feenstra, R. C. and David Weinstein, 2016, Globalization, Markups, and the U.S. Price Level, *Journal of Political Economy*, forthcoming

Fieler, A. C., 2011, Non-homotheticity and Bilateral Trade: Evidence and a Quantitative Explanation, *Econometrica*, 79, 4, 1069-01

Handbury, J. and D. E. Weinstein, 2015, Goods Prices and Availability in Cities, *Review of Economic Studies*, 82, 1, 258-96

Head, K. and T. Mayer, 2014, Gravity Equations: Workhorse, Toolkit, and Cookbook, Ch. 3 in *Handbook of International Economics*, Gopinath, G, E. Helpman and K. Rogoff (Eds), Vol. 4, 131-95, Elsevier

Head, K., T. Mayer and J.Ries, 2010, The Erosion of Colonial Trade linkages after Independence, *Journal of International Economics*, 81, 1, 1-14

Hottman, C., 2014, Retail Markups, Misallocation, and Store Variety in the US, mimeo

Hsieh, C., N. Li, R. Ossa and M. Yang, 2016, Accounting for the New Gains from Trade Liberalization, WP 22069, NBER

Houthakker, H. S., 1960, Additive Preferences, *Econometrica*, 28, 2, 244-57

Hummels, D., and P. J. Klenow, 2005, The variety and quality of a nation's trade, *American Economic Review*, 95, 2, 704-723

Jaimovich, N. and M. Floetotto, 2008, Firm dynamics, Markup variations, and the Business Cycle, *Journal of Monetary Economics*, 55, 1238-1252

Jung, J., I. Simonovska and A. Weinberger, 2015, Exporter Heterogeneity and Price Discrimination: A Quantitative View, WP 21408, NBER

Kehoe, T. J. and K. J. Ruhl, 2013, How Important Is the New Goods Margin in International Trade?, *Journal of Political Economy*, 121, 2, 358-392

Kehoe, T. J., S. Pau and J. Rossbach, 2016, Quantitative Trade Models: Developments and Challenges, WP 22706, NBER

Krugman, P., 1979, Increasing Returns, Monopolistic Competition and International Trade, *Journal of International Economics*, 9, 469-79

Krugman, P., 1980, Scale Economies, Product Differentiation, and the Pattern of Trade, *American Economic Review*, 70, 950-9

Macedoni, L., 2015, The Effect of Per Capita Income on the Product Scope of Exporters, mimeo, UC Davis

Manova, K. and Z. Zhang, 2012, Export Prices Across Firms and Destinations, *Quarterly Journal of Economics*, 127, 1, 379-436

Martin, J., 2012, Markups, quality, and transport costs, *European Economic Review*, 56, 4, 777-791

Mrázová , M. and J.P. Neary, 2013, Not so Demanding: Preference Structure, Firm Behavior, and Welfare, mimeo, Oxford

Melitz, M., 2003, The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica*, 71, 6, 1695-725

Melitz, M. and G. Ottaviano, 2008, Market Size, Trade, and Productivity, *Review of Economic Studies* 75, 1, 295-316

Muendler, M.-A., 2009, Converter from SITC to ISIC, mimeo, University of California, San Diego

Nocco, A., G. Ottaviano and M. Salto, 2014, Monopolistic Competition and Optimum Product Selection, *American Economic Review*, 104, 5, 304-09

Pavcnik N., 2002, Trade Liberalization, Exit, and Productivity Improvements: evidence from Chilean plants, *Review of Economic Studies*, 69, 1, 245-76

Simonovska, I., 2015, Income Differences and Prices of Tradables: Insights from an Online Retailer, *Review of Economic Studies*, 82, 4, 1612-56

Simonovska, I. and M. E. Waugh, 2014a, The Elasticity of Trade: Estimates and Evidence, *Journal of International Economics*, 92, 1, 34-50

Simonovska, I. and M. E. Waugh, 2014b, Trade Models, Trade Elasticities, and the Gains from Trade, WP 20495, NBER

Thisse, J.-F. and P. Ushchev, 2016, When Can a Demand System Be Described by a Multinomial Logit with Income Effect?, Higher School of Economics Research Paper No. WP BRP 139

Varian, H., 1992, *Microeconomic Analysis*, Third Edition, New York: Norton

Waugh, M. E., 2010, International Trade and Income Differences, *American Economic Review*, 100, 2093-2124

Appendix

A Social planner solution with IA preferences

Consider the Social Planner Problem for the model of Section 2:

$$\begin{aligned} & \max_{N,\widehat{c},x(c),s(c)} \left\{ N \int_0^{\widehat{c}} v\left(s(c)\right) dG(c) \right\} \\ s.v. \quad &: \quad N\left[\int_0^{\widehat{c}} cx(c)LdG(c) + F_e\right] = EL, \\ x(c) \quad &= \quad \frac{v'\left(s(c)\right)}{N\int_0^{\widehat{c}} v'\left(s(c)\right)s(c)dG(c)}, \end{aligned}$$

where the first is a resource constraint and the second is the demand associated with our preferences. Combining the two constraints we simplify them to the condition:

$$L\int_0^{\widehat{c}} v'\left(s(c)\right)cdG(c) = \left(EL - NF_e\right)\int_0^{\widehat{c}} v'\left(s(c)\right)s(c)dG(c).$$

Given positive values for N and $\widehat{c}$, consider the Lagrangian:

$$\ell = \int_0^{\widehat{c}} \left\{v\left(s(c)\right) - \lambda v'\left(s(c)\right)\left[\left(EL - NF_e\right)s(c) - Lc\right]\right\} g(c)dc.$$

Pointwise maximization for $s(c)$ provides:

$$v'\left(s(c)\right) - \lambda v''\left(s(c)\right)\left[\left(EL - NF_e\right)s(c) - Lc\right] - \lambda v'\left(s(c)\right)\left(EL - NF_e\right) = 0,$$

which can be rearranged as:

$$s(c) = \frac{\lambda\theta(s\left(c\right))Lc}{\lambda\left[\theta(s\left(c\right)) - 1\right]\left(EL - NF_e\right) + 1},$$

assuming $\theta > 1$. Replacing in the constraint we have:

$$\int_0^{\widehat{c}} v'\left(s(c)\right)c\left[L - \frac{\left(EL - NF_e\right)\lambda\theta(s\left(c\right))L}{\lambda\left[\theta(s\left(c\right)) - 1\right]\left(EL - NF_e\right) + 1}\right]dG(c) = 0,$$

where we can choose $\lambda = 1/\left(EL - NF_e\right)$ to satisfy the above condition. This implies a linear optimal price function $s(c) = Lc/(EL - NF_e)$. Using this, we are left with the residual problem:

$$\max_{\widehat{c},N} \left\{ N \int_0^{\widehat{c}} v\left(\frac{Lc}{EL - NF_e}\right) dG(c) \right\}.$$

Due to the absence of fixed costs of production, it is always optimal to consume any good that provides positive utility by setting $\widehat{c}(N) = aE\left(1 - \frac{NF_e}{EL}\right)$. Therefore, the previous problem simplifies to:

$$\max_{N} N \int_0^{\widehat{c}(N)} v\left(\frac{Lc}{EL - NF_e}\right) dG(c),$$

whose first-order condition is:

$$\int_0^{\widehat{c}(N)} v\left(s(c)\right) dG(c) + \frac{NF_e}{EL - NF_e} \int_0^{\widehat{c}(N)} v'\left(s(c)\right) s(c) dG(c) = 0.$$

This can be solved for:

$$N^* = \frac{EL}{F_e(1+\bar{\eta})},$$

where we defined $\bar{\eta}$ as a weighted average of the elasticity of the subutility $\eta(s) = -v'(s)s/v(s) > 0$, that is:

$$\bar{\eta} \equiv \int_0^{\widehat{c}(N^*)} \eta\left(s(c)\right) \frac{v(s(c))}{\int_0^{\widehat{c}(N^*)} v\left(s(c)\right) dG(c)} dG(c) > 0.$$

It follows that the optimal cost cutoff is:

$$\widehat{c}^* = \frac{aE\bar{\eta}}{1+\bar{\eta}} < aE,$$

which implies that an excessive fraction of goods is consumed in equilibrium. Finally, the optimal price is:

$$p^*(c) = c\left(1 + \frac{1}{\bar{\eta}}\right),$$

which is linear in the marginal cost.

Notice that integration per parts (using the linearity of $s(c)$ and assuming that $v(s(0))$ is finite) delivers:

$$\int_0^{\widehat{c}(N^*)} v'\left(s(c)\right) s(c) dG(c) = \left[v\left(s(c)\right) cg(c)\right]_0^{\widehat{c}(N^*)} - \int_0^{\widehat{c}(N^*)} v\left(s(c)\right)\left[g(c) + cg'(c)\right] dc$$

$$= -\int_0^{\widehat{c}(N^*)} v\left(s(c)\right)\left[g(c) + cg'(c)\right] dc,$$

which allows one to simplify $\bar{\eta}$ as:

$$\bar{\eta} = \frac{\int_0^{\widehat{c}(N^*)} v\left(s(c)\right)\left[g(c) + cg'(c)\right] dc}{\int_0^{\widehat{c}(N^*)} v\left(s(c)\right) dG(c)}.$$

If G is a Pareto distribution we then have $g(c) + cg'(c) = \kappa g(c)$, therefore $\bar{\eta} = \kappa$ independently from the specification of non-homothetic IA preferences.

B Derivations for the parametrized model

Under our specification of preferences $v(s) = \frac{(a-s)^{1+\gamma}}{1+\gamma}$ and the assumption of a Pareto distribution, the prices p_{ij} of firms from country i which are actually active at destination j (i.e., conditional on $c \leq \hat{c}_{ij}$) are distributed on the support $[aE_j/(1+\gamma), aE_j]$ according to:

$$F_j(p) = \Pr\left\{p_{ij} \leq p\right\} = \Pr\left\{\frac{\gamma c + \widehat{c}_{ij}}{1+\gamma}\tau_{ij}w_i \leq p\right\} = \left(\frac{(1+\gamma)\,p}{\gamma a E_j} - \frac{1}{\gamma}\right)^{\kappa}.$$

This distribution is independent from trade costs and the identity of the exporting country, but depends on the income of the importing country j. However, the distribution of the normalized prices $s_{ij} = p_{ij}/E_j$ is identical across countries. Namely, on the support $[\frac{a}{1+\gamma}, a]$ it is given by (40), which depends only on the three parameters γ, κ and a. The average price in country j can then be easily calculated as follows:

$$\mathbb{E}\left\{p_j\right\} = \int_{\frac{aE_j}{1+\gamma}}^{aE_j} p dF_j(p) = \left[pF_j(p)\right]_{\frac{aE_j}{1+\gamma}}^{aE_j} - \int_{\frac{aE_j}{1+\gamma}}^{aE_j} F_j(p)dp$$

$$= \frac{aE_j}{\kappa+1}\left[\kappa + \frac{1}{\gamma+1}\right], \tag{51}$$

which is increasing in income and decreasing in γ.

To obtain the distribution of the corresponding markups, notice that they are distributed on $[0, \infty]$ with:

$$F_m(m) = \Pr\left\{m_{ij} \leq m\right\} = \Pr\left\{\left(\frac{1}{1+\gamma}\right)\left(\frac{\widehat{c}_{ij} - c}{c}\right) \leq m\right\}$$

$$= \Pr\left\{\frac{\widehat{c}_{ij}}{1+(1+\gamma)\,m} \leq c\right\} = 1 - \frac{G\left(\frac{\widehat{c}_{ij}}{1+(1+\gamma)m}\right)}{G\left(\widehat{c}_{ij}\right)}$$

$$= 1 - \frac{1}{\left[1+(1+\gamma)\,m\right]^{\kappa}}. \tag{52}$$

The average markup can be calculated as follows:

$$\mathbb{E}\left\{m\right\} = \lim_{z\to\infty}\left\{\int_0^z m dF_m(m)\right\} = \lim_{z\to\infty}\left\{\left[mF_m(m)\right]_0^z - \int_0^z F_m(m)dm\right\}$$

$$= \frac{1}{(\gamma+1)(\kappa-1)}.$$

This value averages low markups by marginal firms (selling virtually nothing) and high markups by better producers, especially by the extremely productive exporters. Given the skewed distribution, the median mark-up is also of interest: this can be computed directly from (52) as $m^{Med} = \left(2^{1/\kappa}-1\right)/(1+\gamma)$.

Furthermore, it is straightforward to derive the distributions of the pass-through and pricing-to-market elasticities across all producing firms and compute moments from them. Using the Pareto distribution, the distributions of pass-through and pricing-to-market elasticities, which are the same across countries and independent from trade cost, satisfy:

$$\Pr\left\{\epsilon^c \leq \epsilon\right\} = 1 - \left(\frac{\epsilon}{\gamma(1-\epsilon)}\right)^{\kappa} \quad \text{and} \quad \Pr\left\{\epsilon^E \leq \epsilon\right\} = 1 - \left(\frac{1-\epsilon}{\gamma\epsilon}\right)^{\kappa}, \tag{53}$$

respectively. Given these closed-form distributions, the mean and median values can be easily computed, while the means plus standard deviations can be derived numerically. The average elasticity of price with respect to income is:

$$\mathbb{E}\left\{\epsilon^E(c)\right\} = \frac{\widehat{c}_{ij}}{G(\widehat{c}_{ij})}\int_0^{\widehat{c}_{ij}} \frac{dG(c)}{\gamma c + \widehat{c}_{ij}} = \frac{\kappa}{\widehat{c}_{ij}^{\kappa-1}}\int_0^{\widehat{c}_{ij}} \frac{c^{\kappa-1}}{\gamma c + \widehat{c}_{ij}}dc$$

$$= \kappa\int_0^1 t^{\kappa-1}\left(1+\gamma t\right)^{-1}dt \quad \text{with } t \equiv \frac{c}{\widehat{c}_{ij}}$$

$$= F_{2,1}(1,\kappa;1+\kappa;-\gamma), \tag{54}$$

where $F_{2,1}$ is the hypergeometric function:

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

with vector $(\alpha, \beta) = (1, \kappa)$, scalar $\delta = \kappa + 1$ and argument $z = -\gamma$,[43] and

$$\Gamma(t) = \int_0^\infty z^{t-1} e^{-z} dz$$

is the Euler Gamma function (if the real part of t is positive). The median elasticity of price with respect to income is $\epsilon^E_{Med} = 1/\left(1 + \gamma 2^{-1/\kappa}\right)$. One can also evaluate a weighted average elasticity with relative sales as weights, which corresponds to:

$$\bar{\epsilon}^E = \frac{1 + \gamma + \kappa}{(1+\gamma)(1+\kappa)}$$

and is higher because more productive firms have larger market shares.

Finally, to derive the distribution of market shares in the text and to demonstrate that profits are a constant share of sales which does depend neither on the source country nor on the destination, we compute the expected value of the exports to country j of a firm based in country i as follows:

$$\mathbb{E}\left\{t_{ij}\right\} = \int_0^{\widehat{c}_{ij}} t_{ij}(c) dG(c) =$$

$$= \frac{-\gamma^\gamma \left(\tau_{ij} w_i\right)^{\gamma+1} L_j}{(1+\gamma)^{\gamma+1} E_j^\gamma \left|\mu_j\right|} \int_0^{\widehat{c}_{ij}} \left[\gamma \left(\widehat{c}_{ij} - c\right)^\gamma - \gamma \left(\gamma c + \widehat{c}_{ij}\right) \left(\widehat{c}_{ij} - c\right)^{\gamma-1}\right] G_i(c) dc$$

$$= \frac{\gamma^{\gamma+1} a^{\gamma+1} E_j L_j}{(1+\gamma)^\gamma \overline{c}^\kappa \left|\mu_j\right| \widehat{c}_{ij}^{\gamma+1}} \int_0^{\widehat{c}_{ij}} \left(\widehat{c}_{ij} - c\right)^{\gamma-1} c^{\kappa+1} dc,$$

where we integrated by parts. Changing the variable of integration with $t = c/\widehat{c}_{ij}$ we obtain:

$$\mathbb{E}\left\{t_{ij}\right\} = \frac{\gamma^{\gamma+1} a^{\gamma+1} E_j L_j \widehat{c}_{ij}^\kappa}{(1+\gamma)^\gamma \overline{c}^\kappa \left|\mu_j\right|} \int_0^1 (1-t)^{\gamma-1} t^{\kappa+1} dt$$

$$= \frac{a^{\kappa+\gamma+1} \gamma^{\gamma+1} B(\kappa+2, \gamma)}{(1+\gamma)^\gamma \overline{c}^\kappa} \frac{L_j E_j^{\kappa+1}}{\left|\mu_j\right| \left(\tau_{ij} w_i\right)^\kappa}. \tag{55}$$

This allows to derive the average sales and the expression for the market share (43). Similarly, the expected profit $\mathbb{E}\left\{\pi_{ij}\right\}$ in country j for a firm based in country i is given by:

$$\mathbb{E}\left\{\pi_{ij}\right\} = \int_0^{\widehat{c}_{ij}} \pi_{ij}(c) dG(c) = \frac{\gamma^\gamma \kappa B(\kappa, \gamma+2) a^{\kappa+\gamma+1}}{(1+\gamma)^{1+\gamma} \overline{c}^\kappa} \frac{L_j E_j^{\kappa+1}}{\left(\tau_{ij} w_i\right)^\kappa \left|\mu_j\right|}. \tag{56}$$

The ratio of the two aggregate objects is then obtained by the recursive properties of the Euler Beta

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

function:

$$\frac{\mathbb{E}\{\pi_{ij}\}}{\mathbb{E}\{t_{ij}\}} = \frac{\kappa B(\kappa, \gamma+2)}{\gamma(1+\gamma)B(\kappa+2, \gamma)} = \frac{1}{\kappa+1}.$$

C Equivalent variation for IA preferences

Consider the general case of IA preferences. In this case it is convenient to work with the "unconditional" distribution, $G_{ij}(\chi)$, of the marginal cost $\chi = \tau_{ij} w_i c$ in country j by firms from country i, which has a support $[0, \tau_{ij} w_i \overline{c}]$ and it is given by:

$$G_{ij}(\chi) = \Pr\left\{c \leq \frac{\chi}{\tau_{ij} w_i}\right\} = G(\frac{\chi}{\tau_{ij} w_i}) = \left(\frac{\chi}{\tau_{ij} w_i \overline{c}}\right)^{\kappa}.$$

Let $p_j(\chi)$ be the equilibrium mapping between marginal costs and prices which only depends on E_j.[44] We can then write welfare (25) as:

$$V_j = \sum_{i=1}^{I} N_i \int_{bE_j}^{aW_j} v\left(\frac{p}{W_j}\right) dF_{ij}(p)$$

$$= \sum_{i=1}^{I} N_i \int_{0}^{\overline{\chi}_j} v\left(\frac{p_j(\chi)}{W_j}\right) dG_{ij}(\chi)$$

$$= \int_{0}^{\overline{\chi}_j} v\left(\frac{p_j(\chi)}{W_j}\right) d(\chi)^{\kappa} \sum_{i=1}^{I} N_i \left(\tau_{ij} w_i \overline{c}\right)^{-\kappa},$$

where $W_j = E_j + EV_j$ (see the discussion concerning the definition of the Equivalent Variation EV_j in the text) and $\overline{\chi}_j$ is defined by the condition $p_j(\overline{\chi}_j) \equiv aW_j$. Accordingly, taking logs, differentiating and integrating by parts we obtain :

$$d\ln V_j = d\ln\left\{\int_{0}^{\overline{\chi}_j} v\left(\frac{p_j(\chi)}{W_j}\right) d(\chi)^{\kappa}\right\}$$

$$= \frac{-\int_{0}^{\overline{\chi}_j} v'\left(\frac{p_j(\chi)}{W_j}\right)\frac{p_j(\chi)}{W_j} d(\chi)^{\kappa}}{\int_{0}^{\overline{\chi}_j} v\left(\frac{p_j(\chi)}{W_j}\right) d(\chi)^{\kappa}} d\ln W_j$$

[44] This is given by (14) for all the varieties actually sold in country j when income is E_j, but it is not uniquely defined above the cutoff aE_j. One can make the mild assumption that $p_j(\chi)$ is everywhere monotonic and differentiable: however, in computing the EV_j for the functional form of our example we assume that all prices follow the same pricing rule (34).

$$= \frac{\kappa \int_0^{\overline{\chi}_j} v'\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1} d\chi}{\int_0^{\overline{\chi}_j} v'\left(\frac{p_j(\chi)}{W_j}\right) p_j'(\chi)\chi^{\kappa} d\chi} d\ln W_j$$

$$= \kappa \left[\int_0^{\overline{\chi}_j} \frac{p_j'(\chi)\chi}{p_j(\chi)} \frac{v'\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1}}{\int_0^{\overline{\chi}_j} v'\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1} d\chi} d\chi \right]^{-1} d\ln W_j$$

$$= \kappa \left[\int_0^{\overline{\chi}_j} \epsilon_j^c(\chi) \frac{x\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1}}{\int_0^{\overline{\chi}_j} x\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1} d\chi} d\chi \right]^{-1} d\ln W_j$$

$$= \kappa \left[1 - \int_0^{\overline{\chi}_j} \epsilon_j^E(\chi) \frac{x\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1}}{\int_0^{\overline{\chi}_j} x\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1} d\chi} d\chi \right]^{-1} d\ln W_j$$

$$= \kappa \left[1 - \bar{\epsilon}_j^E \left(W_j, E_j\right)\right]^{-1} d\ln W_j, \tag{57}$$

where we define

$$\epsilon_j^c(\chi) \equiv \frac{\partial \ln p_j(\chi)}{\partial \ln \chi} \equiv 1 - \epsilon_j^E(\chi),$$

and

$$\bar{\epsilon}_j^E \left(W_j, E_j\right) \equiv \int_0^{\overline{\chi}_j} \epsilon_j^E(\chi) \frac{x\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1}}{\int_0^{\overline{\chi}_j} x\left(\frac{p_j(\chi)}{W_j}\right) p_j(\chi)\chi^{\kappa-1} d\chi} d\chi. \tag{58}$$

The local approximation in the text, valid for small EV_j, can be obtained by letting $W_j = E_j$ and computing $\bar{\epsilon}_j^E$ as a weighted average (with relative sales as weights) of the elasticity of prices with respect to income, $\epsilon_{ij}^E(c) = \partial \ln p_{ij}(c)/\partial \ln E_j$, which explains our notation.

For our specific functional form we obtain $\overline{\chi}_j = a\left[(\gamma+1)W_j - E_j\right]/\gamma$ and

$$\bar{\epsilon}_j^E \left(W_j, E_j\right) = \int_0^{\overline{\chi}_j} \frac{aE_j}{\gamma\chi + aE_j} \frac{\left\{a\left[(\gamma+1)W_j - E_j\right] - \gamma\chi\right\}^{\gamma} \left(\gamma\chi + aE_j\right)\chi^{\kappa-1}}{\int_0^{\overline{\chi}_j} \left\{a\left[(\gamma+1)W_j - E_j\right] - \gamma\chi\right\}^{\gamma} \left(\gamma\chi + aE_j\right)\chi^{\kappa-1} d\chi} d\chi \tag{59}$$

$$= \frac{aE_j \int_0^{\overline{\chi}_j} \left\{1 - \frac{\gamma\chi}{a[(\gamma+1)W_j - E_j]}\right\}^{\gamma} \chi^{\kappa-1} d\chi}{\int_0^{\overline{\chi}_j} \left\{1 - \frac{\gamma\chi}{a[(\gamma+1)W_j - E_j]}\right\}^{\gamma} \left(\gamma\chi^{\kappa} + aE_j^{\kappa-1}\chi\right) d\chi}$$

$$= \frac{aE_j \int_0^1 \{1-t\}^{\gamma} t^{\kappa-1} dt}{\int_0^1 \{1-t\}^{\gamma} \left(\gamma\overline{\chi}_j t^{\kappa} + aE_j t^{\kappa-1}\right) dt}$$

$$= \frac{aE_jB(\kappa,\gamma+1)}{\gamma\overline{\chi}_jB(\kappa+1,\gamma+1)+aE_jB(\kappa,\gamma+1)}$$
$$= \frac{(\kappa+\gamma+1)\,E_j}{[\kappa W_j+E_j]\,(\gamma+1)},$$

which gives (46) in the text.

D Quantitative analysis

In this section, we quantify the model's key predictions. First, we revisit the model's implications and we derive testable predictions that can be compared to moments in the data. Second, we outline a strategy to identify the model's parameters and we evaluate the quantitative fit of the model to observations from cross-firm and cross-country data. Third, we design a counterfactual exercise to quantify the welfare gains from trade predicted by the model.

D.1 Background

As described in the main text, in order to quantify the welfare gains from trade predicted by the model, we need data on trade shares and per-capita income as well as estimates of two key parameters, κ and γ. In this section, we describe a structural approach toward identifying these two parameters. We discuss the merits of this approach amid data limitations in Section D.4, Step 3 below.

Our identification approach demands that we also take a stand on a number of additional parameters that characterize the model. Since the model falls within a large class of models that generate a log-linear gravity equation of trade, the two key parameters, κ and γ, together with a set of country- and country-pair-specific parameters that can be estimated using the model's structural gravity equation of trade, are sufficient to generate a set of moments that can be used to judge the model's fit to the data.[45]

Recall the theoretical gravity equation of trade derived in expression (49). Let $S_i = \exp(\tilde{S}_i)$, a transformation that will be used extensively as we proceed.

Once we have obtained estimates for the parameters κ and γ, as well as for the objects S_i for all $i = 1, .., I$ and τ_{ij} for all country-pairs i, j (see Section D.4 below for estimation), we can compute predicted per-capita income levels for each country from the model's predicted market clearing conditions using data for actual trade shares λ_{ij} for all i, j pairs and population size L_j for all countries j. In particular, refer to expression (23), where by definition per-capita income

[45] Simonovska and Waugh (2014,a,b) demonstrate this fact for models that rely on homothetic preferences, while Jung *et al.* (2015) analyze models that belong to a class of directly additive preferences.

in country i is $E_i = w_i e_i$. After normalizing population size, L_i, relative to a numeraire, we can obtain per-capita incomes, E_i, for any i, relative to a numeraire, using data on L_i and λ_{ij} for all i, j from the system of equations (23) for al i. Let $\mathbf{P}$ denote the vector of the parameters necessary for simulation in matrix notation, namely $\mathbf{P} = \{\kappa, \gamma, \boldsymbol{\tau}, \mathbf{E}, \mathbf{L}, \mathbf{S}\}$, where $\mathbf{E}' = [E_1, .., E_I]$ and $\mathbf{S}' = [S_1, .., S_I]$, and let $\boldsymbol{\Lambda}$ denote the bilateral trade-share matrix with typical element λ_{ij}.

With $\mathbf{P}$ and $\boldsymbol{\Lambda}$ in hand, we can compute all endogenous objects in the model that are necessary to derive a number of moments that we can use to identify κ and γ. We begin by computing cost cutoffs. Expression (15) suggests that a value for the parameter a would be needed in order to obtain cost cutoffs. As it turns out, it is sufficient to compute cost cutoffs relative to a numeraire cutoff, in order to derive the moments of interest; hence, we will not need to take a stand on the value of a because the parameter scales all cutoffs. Below we will describe how we select the numeraire cutoff. We begin by revisiting the model's predictions and translating them into moments from the model that are expressed as functions of normalized cutoffs, $\mathbf{P}$, and $\boldsymbol{\Lambda}$.[46]

D.2 Empirical predictions

The key empirical prediction that differentiates our model from existing frameworks relates to the behavior of prices across destinations; namely prices are higher in richer countries, but they do not vary with respect to market size. We discussed this prediction in the main text and we derived moment conditions for the key price elasticities there. Below, we revisit the remaining quantitative predictions of the model.

D.2.1 Pass-through and mark-ups across firms

In the model, more productive exporters price to market more or, alternatively, they enjoy lower cost pass-through. In expression (53) in Appendix B we have derived the distribution of the elasticity of price with respect to income, which, as already mentioned, corresponds to the elasticity with respect to the real exchange rate. We reproduce it below for convenience:

$$\Pr\left\{\epsilon^E \leq \epsilon\right\} = 1 - \left(\frac{1-\epsilon}{\gamma\epsilon}\right)^{\kappa}.$$

Given estimates of κ and γ, in addition to the mean of this distribution as illustrated above, we can also compute the mean plus one standard deviation response to exchange rate changes. This allows us to compare both measures to the corresponding moments in the data so as to test the prediction that small (or less productive) firms pass through cost changes more.

[46] It is worth to note that we will not need to estimate the parameter F_e for the purpose of our exercises.

Finally, the distribution of mark-ups is given in expression (41). With estimates of κ and γ at hand, we can derive moments from the mark-up distribution and compare them to data. In particular, the mean markup is $\mathbb{E}\{m\} = \frac{1}{(\gamma+1)(\kappa-1)}$, which is decreasing in $\kappa > 1$ and γ.

D.2.2 Extensive margin of trade

The extensive margin of trade was derived in expression (18) for general IA preferences. It varies across source and destination countries. Given a source country i, let j^* denote a numeraire destination. The ratio of the extensive margin for destination j, relative to the numeraire, is:

$$ext_{ij} = \left(\frac{E_j}{E_{j^*}}\right)^{\kappa} \left(\frac{\tau_{ij}}{\tau_{ij^*}}\right)^{-\kappa}. \tag{60}$$

Taking logs of the above expression allows us to obtain elasticities of the extensive margin with respect to destination specific characteristics. Hence the model predicts that, for a given source country, the extensive margin of trade is increasing in per-capita income with an elasticity of κ and falling in trade costs with an elasticity of $-\kappa$. With an estimate of trade costs at hand, we can also compute the elasticity of the extensive margin with respect to distance to the destination, and compare it to data.

D.2.3 Intensive margin of trade

The intensive margin of trade was derived in expression (20) for general IA preferences. It measures the average sales for firms in a particular destination and it is independent of the source country. Letting country j^* be a numeraire destination, and using the definition of the gravity object S_i, the ratio of the intensive margin for destination j, relative to the numeraire, can be rewritten as:

$$int_j = \frac{E_j L_j}{E_j^* L_j^*} \left(\frac{E_j}{E_{j^*}}\right)^{-\kappa} \left(\frac{\sum_k S_k \tau_{kj}^{-\kappa}}{\sum_k S_k \tau_{kj^*}^{-\kappa}}\right)^{-1} \tag{61}$$

Taking logs of both sides of the expression yields the following: controlling for aggregate effects (summation terms), the elasticity of the intensive margin with respect to destination GDP is 1 and the elasticity of the intensive margin with respect to destination per-capita income is $-\kappa$.

D.2.4 Sales and measured productivity advantage of exporters

More efficient firms realize higher sales. Moreover, trade barriers prevent less efficient firms from exporting, which implies that exporters enjoy an efficiency and sales advantage over non-exporters. Below, we derive two moments of interest from the distributions of measured productivity and sales of firms: the measured productivity and sales advantage of exporters over non-exporters. We derive these moments because we can readily compare them to their data counterparts.

Exporter sales advantage In order to derive moments for exporters and non-exporters from any source country i, it is useful to define a cost cutoff that separates firms into these two groups. In particular, using the characterization for cost cutoffs in expression (15), define the cost cutoff for exporters from country i as

$$\tilde{c}_{ij} \equiv \max_{k \neq i} \frac{aE_k}{\tau_{ik} w_i} \tag{62}$$

Notice that any firm from country i with cost $c < \tilde{c}_{ij}$ is an exporter to some country k and any firm with cost $c \in [\tilde{c}_{ij}, \hat{c}_{ii}]$ serves the domestic market only.[47] This follows from the fact that firms differ only along the cost dimension, so there is a strict ordering of markets by toughness, with the destination k'' being toughest for i's producers if $\hat{c}_{ik''} \leq \hat{c}_{ik'}$ $\forall k'$. Thus, we can refer to country j that satisfies the definition in expression (62) as the most accessible foreign destination for firms from i.

Having categorized firms into exporters and non-exporters, the first moment we are interested in is the ratio between the average domestic sales of exporters and the average sales of non-exporters from any country i.[48] Consider any firm from country i; its domestic sales are given by expression (35), where destination $j = i$. Integrating over all exporters, then integrating over all non-exporters, and finally taking the ratio of the two yields the exporter sales advantage at home, which we denote by $\tilde{H}_1$:

$$\tilde{H}_1 = \frac{\left[\left(\frac{\hat{c}_{ii}}{\tilde{c}_{ij}}\right)^{\kappa} - 1\right]\left[B\left(\frac{\tilde{c}_{ij}}{\hat{c}_{ii}}; \kappa, \gamma+2\right) + (1+\gamma)\, B\left(\frac{\tilde{c}_{ij}}{\hat{c}_{ii}}; \kappa+1, \gamma+1\right)\right]}{B(\kappa, \gamma+2) - B\left(\frac{\tilde{c}_{ij}}{\hat{c}_{ii}}; \kappa, \gamma+2\right) + (1+\gamma)\left[B(\kappa+1, \gamma+1) - B\left(\frac{\tilde{c}_{ij}}{\hat{c}_{ii}}; \kappa+1, \gamma+1\right)\right]},$$

where $B(u; z, h)$ is the incomplete (Euler) Beta function:

$$B(u; z, h) = \int_0^u t^{z-1}(1-t)^{h-1} dt.$$

To see that $\tilde{H}_1$ depends on $\mathbf{P}$ only, define $y_{ij} \equiv \frac{max_{k \neq i} E_k \tau_{ik}{}^{-1}}{E_i}$. Using the expressions for cost cutoffs in (15) and (62), it immediately follows that $\frac{\hat{c}_{ii}}{\tilde{c}_{ij}} = y_{ij}^{-1}$. Then, our desired moment, now

[47] This implicitly assumes that trade barriers are high enough so that $\hat{c}_{ii} > \hat{c}_{ik}$ $\forall k \neq i$.

[48] We follow Bernard *et al.* (2003) and derive this ratio because we will be comparing the model's predicted moment to the corresponding moment from the US distribution reported by these authors.

denoted by H_1, can be rewritten as:

$$H_1(\mathbf{P}) = \left[y_{ij}^{-\kappa} - 1\right] \cdot \frac{B(y_{ij};\kappa,\gamma+2) + (1+\gamma)\,B(y_{ij};\kappa+1,\gamma+1)}{B(\kappa,\gamma+2) - B(y_{ij};\kappa,\gamma+2) + (1+\gamma)\left[B(\kappa+1,\gamma+1) - B(y_{ij};\kappa+1,\gamma+1)\right]}, \tag{63}$$

where the dependence on $\mathbf{P}$ only is made explicit.

Exporter measured productivity advantage The second moment of interest is the measured productivity advantage of exporters over non-exporters. In the absence of intermediate goods, the value added of a firm is the ratio of its sales to the number of employees. Firm sales are given in expression (35). To derive the number of workers, notice that the production function of a $c-$firm from country i selling in country j is $x_{ij} = l_{ij}/(\tau_{ij}c)$, where $\tau_{ij}c$ is its "unit labor requirement" and $l_{ij}(c) = \tau_{ij}cx_{ij}(c)$ its conditional labor demand. The corresponding number of employed workers is given by $\tau_{ij}cx_{ij}(c)/e_i$.

With this in mind, the measured productivity, or the value added per worker, of a non-exporter with cost draw $c \in [\tilde{c}_{ij}, \hat{c}_{ii}]$ from country i is:

$$va_i^{nx}(c) = \frac{e_i t_{ii}(c)}{c\tau_{ii}x_{ii}(c)} = w_i e_i\left[1 + m_{ii}(c)\right].$$

Similarly, the measured productivity, or the value added per worker, of an exporter with cost draw $c < \tilde{c}_{ij}$ is:

$$va_i^{x}(c) = \frac{e_i \sum_{k\in K_i(c)} t_{ik}(c)}{c\sum_{k\in K_i(c)} \tau_{ik}x_{ik}(c)},$$

where $K_i(c)$ is the set of destinations k such that $c \le \widehat{c}_{ik}$.

Taking logs of both variables, integrating over all exporters and non-exporters, respectively, and taking the difference of the two yields the exporter measured productivity advantage (in percentage terms)[49], which we denote by $\tilde{H}_2$:

$$\tilde{H}_2 = \int_0^{\tilde{c}_{ij}} \log(va_i^x(c))\frac{\kappa c^{\kappa-1}}{\tilde{c}_{ij}^{\kappa}}dc - \int_{\tilde{c}_{ij}}^{\hat{c}_{ii}} \log(va_i^{nx}(c))\frac{\kappa c^{\kappa-1}}{\hat{c}_{ii}^{\kappa} - \tilde{c}_{ij}^{\kappa}}dc.$$

As was the case for $\tilde{H}_1$ above, it can be shown that $\tilde{H}_2$ can be re-expressed in terms only of $\mathbf{P}$ and $\mathbf{\Lambda}$, and denoted by $H_2(\mathbf{P},\mathbf{\Lambda})$. Focus first on the value added of non-exporters and

[49] See the preceding footnote.

substitute out the mark-up equation to obtain:

$$va_i^{nx}(c) = \frac{E_i}{1+\gamma}\left(\gamma + \frac{\hat{c}_{ii}}{c}\right).$$

Taking logs yields:

$$\log(va_i^{nx}(c)) = \log\left(\frac{E_i}{1+\gamma}\right) + \log\left(\gamma + \frac{\hat{c}_{ii}}{c}\right).$$

Integrating over all non-exporters yields

$$VA_i^{nx} = \log\left(\frac{E_i}{1+\gamma}\right) + \frac{1}{\hat{c}_{ii}^{\kappa} - \tilde{c}_{ij}^{\kappa}}\int_{\tilde{c}_{ij}}^{\hat{c}_{ii}} \log\left(\gamma + \frac{\hat{c}_{ii}}{c}\right)\kappa c^{\kappa-1} dc.$$

Apply the following change of variables: $t_{ij} = \frac{c}{\hat{c}_{ij}}$. Then VA_i^{nx} becomes:

$$VA_i^{nx} = \log\left(\frac{E_i}{1+\gamma}\right) + \frac{\kappa}{1 - y_{ij}^{\kappa}}\int_{y_{ij}}^{1} \log\left(\gamma + t_{ii}^{-1}\right) t_{ii}^{\kappa-1} dt_{ii},$$

where y_{ij} is defined as above.

Next, focus on the value added for an exporter. For any exporter from country i with cost draw c define the following indicator function: $\delta_{ij}(c) = 1$ if $c < \hat{c}_{ij}$ and zero otherwise. Let $\Delta_{ij}(c)$ be a vector of size I with typical element $\delta_{ij}(c)$. Substituting in the equations for firm sales and output, the value added for an exporter can then be rewritten as

$$va_i^{x}(c) = \frac{e_i \sum_{k=1}^{I} \delta_{ik}(c) \frac{L_k(\tau_{ik} w_i)^{1+\gamma}(\gamma c + \hat{c}_{ik})(\hat{c}_{ik} - c)^{\gamma}}{(1+\gamma)(w_k e_k)^{\gamma}|\mu_k|}}{c\sum_{k=1}^{I} \delta_{ik}(c) \frac{\tau_{ik} L_k(\tau_{ik} w_i)^{\gamma}(\hat{c}_{ik} - c)^{\gamma}}{(w_k e_k)^{\gamma}|\mu_k|}}.$$

Furthermore, substituting in for $|\mu_k|$, and using the definition of λ_{kk} obtains:

$$va_i^{x}(c) = \frac{E_i}{1+\gamma}\left[\gamma + \frac{\sum_{k=1}^{I} \delta_{ik}(c) \frac{\tau_{ik}^{1+\gamma}(\hat{c}_{ik} - c)^{\gamma}\lambda_{kk} L_k}{(E_k)^{\gamma+\kappa} S_k} \frac{\hat{c}_{ik}}{c}}{\sum_{k=1}^{I} \delta_{ik}(c) \frac{\tau_{ik}^{1+\gamma}(\hat{c}_{ik} - c)^{\gamma}\lambda_{kk} L_k}{(E_k)^{\gamma+\kappa} S_k}}\right].$$

Taking logs and integrating over all exporters yields

$$VA_i^{x} = \log\left(\frac{E_i}{1+\gamma}\right) + \frac{1}{\tilde{c}_{ij}^{\kappa}}\int_{0}^{\tilde{c}_{ij}} \log\left[\gamma + \frac{\sum_{k=1}^{I} \delta_{ik}(c) \frac{\tau_{ik}^{1+\gamma}(\hat{c}_{ik} - c)^{\gamma}\lambda_{kk} L_k}{(E_k)^{\gamma+\kappa} S_k} \frac{\hat{c}_{ik}}{c}}{\sum_{k=1}^{I} \delta_{ik}(c) \frac{\tau_{ik}^{1+\gamma}(\hat{c}_{ik} - c)^{\gamma}\lambda_{kk} L_k}{(E_k)^{\gamma+\kappa} S_k}}\right]\kappa c^{\kappa-1} dc.$$

Applying the change of variables, $t_{ij} = \frac{c}{\hat{c}_{ij}}$, VA_i^x becomes:

$$VA_i^x = \log\left(\frac{E_i}{1+\gamma}\right) + \kappa y_{ij}^{-\kappa} \int_0^{y_{ij}} \log\left[\gamma + \frac{\sum_{k=1}^{I} \delta_{ik}(\hat{c}_{ii}t_{ii}) \frac{\lambda_{kk}L_k}{(E_k)^{\kappa-1}S_k}\left(1 - t_{ii}\frac{E_i\tau_{ik}}{E_k}\right)^{\gamma} \frac{1}{t_{ii}E_i}}{\sum_{k=1}^{I} \delta_{ik}(\hat{c}_{ii}t_{ii}) \frac{\tau_{ik}\lambda_{kk}L_k}{(E_k)^{\kappa}S_k}\left(1 - t_{ii}\frac{E_i\tau_{ik}}{E_k}\right)^{\gamma}}\right] t_{ii}^{\kappa-1} dt_{ii}.$$

Taking the difference between exporters and non-exporters yields the desired moment H_2:

$$H_2(\mathbf{P}, \mathbf{\Lambda}) = \kappa y_{ij}^{-\kappa} \int_0^{y_{ij}} \log\left[\gamma + \frac{\sum_{k=1}^{I} \tilde{\delta}_{ik}(t_{ii}) \frac{\lambda_{kk}L_k}{(E_k)^{\kappa-1}S_k}\left(1 - t_{ii}\frac{E_i\tau_{ik}}{E_k}\right)^{\gamma} \frac{1}{t_{ii}E_i}}{\sum_{k=1}^{I} \tilde{\delta}_{ik}(t_{ii}) \frac{\tau_{ik}\lambda_{kk}L_k}{(E_k)^{\kappa}S_k}\left(1 - t_{ii}\frac{E_i\tau_{ik}}{E_k}\right)^{\gamma}}\right] t_{ii}^{\kappa-1} dt_{ii}$$

$$- \frac{\kappa}{1 - y_{ij}^{\kappa}} \int_{y_{ij}}^{1} \log\left(\gamma + t_{ii}^{-1}\right) t_{ii}^{\kappa-1} dt_{ii}, \tag{64}$$

where the dependence on $\mathbf{P}$ and $\mathbf{\Lambda}$ is made explicit. In the above expression, $\tilde{\delta}_{ik}$ is a transformation of δ_{ik} that only depends on $\mathbf{P}$ and $\mathbf{\Lambda}$. Thus, it remains to show that t_{ii} and $\tilde{\Delta}_{ik}$ depend on $\mathbf{P}$ and $\mathbf{\Lambda}$. This argument can be found in the description of the simulation algorithm below.

Wages The two firm-level moments derived above rely on the endogenous wage, w_i, of the country whose exporters are simulated. In principle, should we simulate exporters for all countries, we would need to separately identify wages for all countries. However, the exporter moments that we will try to reconcile are only made available for US exporters by Bernard *et al.* (2003). Assuming that the two key parameters, κ and γ, are not country-specific, we can generate the moments from the model for US exporters by only simulating observations for US exporters. In this case, we let $w_{US} = 1$.

D.3 Simulation algorithm

In this model, there exists a continuum of firms; hence, the first step in the simulation is to recognize that the continuum needs to be discretized and the number of simulated draws has to be large enough so as to best approximate the entire continuum. In principle, one would need to simulate a very large number of draws for each country; which can be a daunting task. The task, however, is greatly simplified due to the fact that cost draws are transformations of random variables drawn from a parameter-free uniform distribution, where the transformation function depends on $\mathbf{P}$. This powerful insight draws on arguments first made transparent by Bernard *et al.* (2003) within the context of a model with a fixed measure of firms and subsequently by Eaton *et al.* (2011) within a model with an endogenous measure of firms.

Recall that our goal is to simulate a large number of cost draws, c, from the pdf given by

$g_i(c) = \kappa c^{\kappa-1}/\hat{c}_{ii}^{\kappa}$, which ensures that $c \in [0, \hat{c}_{ii}]$ for all i.[50] Given these draws, we can proceed to compute exporting costs and determine the subset of firms from each source country i that serve each destination j. With this in mind we proceed as follows. We draw 500,000 realizations[51] of the uniform distribution on the $[0, 1]$ domain, $U[0, 1]$, we order them in increasing order, and find the maximum realization, denoted by $u^{\max}$. Then, we let $c = (u/u^{\max})^{\frac{1}{\kappa}} \hat{c}_{ii}$. Notice that $c \in [0, \hat{c}_{ii}]$ by construction, and it has pdf of $g_i(c) = \kappa \frac{c^{\kappa-1}}{\hat{c}_{ii}^{\kappa}}$; yet the normalization allows us to utilize all draws. Multiplying each c by the appropriate wage rate and trade cost yields the cost to serve each market. Comparing this cost to the cost cutoff for each source-destination pair determines the set of exporters to every destination.

What remains is to decide the source-destination cost cutoff pair that serves as numeraire. This choice depends on the particular exercise that one intends to engage in. The objective of the normalization, however, is always the same: the numeraire is chosen so as to maximize the usage of the 500,000 draws from the uniform distribution. As we describe below, in at least one of the exercises, we choose to identify the key parameters of interest, κ and γ, from moments for US firms; thus, we need to simulate observations for all US producers—both domestic and exporting. To maximize the number of draws used, we choose the numeraire cost cutoff to be $\hat{c}_{US,US}$. Hence, all simulated firms serve at least the US and a subset of them serve different export markets.

D.4 Estimation

In order to numerically generate the moments from the model that we outline above, we first need to estimate the model's parameters and then simulate micro-level data. The estimation can be divided into the following three steps: (i) estimate a set of country(-pair) parameters using the model's predicted gravity equation of trade and data; (ii) use gravity-based estimates, together with data on population size, to estimate per-capita income levels from the model's market clearing condition; (iii) use parameters from (i) and (ii), together with moments of choice to identify the remaining parameters, γ and κ.

Step 1 The empirical gravity equation of trade that corresponds to the theoretical prediction derived in expression (49) is given by:

$$\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = \tilde{S}_i - \tilde{S}_j - \kappa \log \tau_{ij} + \varepsilon_{ij}, \tag{65}$$

[50]It would be futile to simulate firms with higher cost draws than this upper bound because they would immediately exit in equilibrium.

[51]The quantitative results are nearly identical when we use a grid of 2,000,000. One key difference is that US exporters serve a larger number of destinations in this case; namely, there are fewer zeros in the trade matrix.

where ε_{ij} is a country-pair residual error term. We assume that the bilateral trade cost takes on the following functional form:

$$\log \tau_{ij} = \beta_d \log d_{ij} + ex_i + \boldsymbol{\beta}_k \mathbf{d}_k + \beta_h d_h, \tag{66}$$

where β_d is the coefficient estimate on the log of the bilateral distance in kilometers, d_{ij}, ex_i is an exporter fixed effect as in Waugh (2010), d_h is an indicator that takes on the value of 1 if trade is internal with coefficient β_h, and $\boldsymbol{\beta}_k$ is a 5×1 vector of coefficients on a matrix of 5 indicators, $\mathbf{d}_k$, where each indicator takes on the value of 1 if countries i and j: (i) share a border, (ii) have a common official or primary language, (iii) have a common colonizer post 1945, (iv) have a regional trade agreement (RTA) in force, and (v) share a common currency.

After substituting expression (66) into (65), we estimate the coefficients for 123 countries via OLS using source and destination fixed effects.[52] We exclude trade share observations that take on the value of zero. A description of the (standard) datasets used in the estimation and the results from the gravity estimation can be found in Appendix E. We present the estimates of the gravity equation in Appendix F, and we plot the predicted and actual trade shares in Section D.5.

A couple of notes are in order. First, all parameter estimates pertaining to the trade costs are scaled by κ. Thus, the gravity equation allows us to estimate $\kappa \log \tau_{ij}$ only, rather than to separately identify κ from τ_{ij}. We present our identification strategy for κ in Step 3 below. Second, domestic trade costs are also estimated in this step and they are not necessarily equal to unity. Hence, before we proceed, we normalize all international trade costs, relative to their domestic counterparts.

Step 2 We compute per-capita incomes using the model's implied market clearing equation together with data on trade shares and population size for all 123 countries. In particular, we employ expression (23), where by definition per-capita income in country i is $E_i = w_i e_i$. After normalizing population size, L_i, relative to a numeraire country, which we take to be the US, we can obtain per-capita incomes, E_i, for all $i \neq US$, relative to the US, using data on L_i and λ_{ij} for all i, j from this system of equations. We describe the data sources in Appendix E and we plot the predicted and actual per-capita income in Section D.5.

Step 3 It remains to choose an identification strategy for the key remaining parameters that characterize the welfare gains from trade: κ and γ. In principle, these two parameters govern more than two moments in the model. Hence, different sets of moments will result in different

[52]For a detailed discussion on how to separately identify the coefficients S_i from ex_i see Simonovska and Waugh (2014a).

estimates for these parameters, different fits of model to data, and different estimates of the gains from trade. The main challenge is to select the moments that are (i) most informative about these two parameters and (ii) directly informative about the welfare gains from trade.

The parameter κ is a supply-side parameter; it is the shape parameter that governs the dispersion of the intrinsic productivity distribution, Pareto. The parameter γ is a demand-side parameter in our model—it governs the elasticity of demand with respect to price as well as the elasticity of price with respect to income and the pass-through elasticity. Consequently, the parameter is at the heart of the model's pricing predictions, which are the key departure from the standard ACR framework. In addition, for given κ, γ governs the distribution of firm sales. These two observations imply that γ is a critical input into the firm sales-weighted elasticity of price with respect to income, which quantifies the welfare gains predicted by our model.

With this in mind, we proceed with an overidentified estimation strategy for γ and κ by jointly targeting the following three moments: (i) the average elasticity of price with respect to income from micro-level data, (ii) the domestic sales advantage of exporters over non-exporters for the US, and (iii) the measured productivity advantage of exporters over non-exporters for the US.

We choose these moments for the following reasons. First, our model differs from existing alternatives precisely along the pricing dimension—it predicts that the price elasticity with respect to income is positive, while the price elasticity with respect to market size is zero. Therefore, targeting the price elasticity with respect to income seems to be a natural choice. Second, it is the price elasticity with respect to income (or, alternatively the pass-through elasticity: recall the definition of $\bar{\epsilon}_j^E$), weighted by each firm's relative sales, that constitutes the object that governs the welfare gains in our model. Therefore, the distribution of firm sales together with the price elasticity with respect to income are crucial elements in determining the magnitude of welfare gains from trade. Third, if we had access to firm-level sales data as well as firms' prices of identical goods sold across multiple destinations, then we could have directly estimated $\bar{\epsilon}_j^E$ (or the pass-through elasticity) instead of having to estimate γ. However, to our knowledge, such detailed data are not readily available. Given data limitations, we opted for the structural approach described above. Finally, the third moment seems natural as κ governs the shape of the productivity distribution in the model; thus, the measured productivity advantage of exporters is a very informative moment.

In sum, to identify κ and γ we combine objects from Steps 1 and 2 with three model-generated moments: (i) the domestic sales advantage of US exporters over non-exporters, or $H_1(\mathbf{P})$ given in expression (63); (ii) the value-added advantage of US exporters over non-exporters, or $H_2(\mathbf{P}, \mathbf{\Lambda})$ given in expression (64); and (iii) the average elasticity of price with respect to income from micro-level data, or $\mathbb{E}\left\{\epsilon_E\right\}$ given in expression (50). To compute the first two moments, we let the source country $i = US$ and we consider all 123 potential destinations referred to in Steps 1 and 2 above (the third moment is country invariant in the model). Finally, we employ the SMM estimator with an identity weighting matrix, which prevents us from computing meaningful standard errors.[53]

On the technical side, to compute the first two moments, we need to separately identify w_{US} from e_{US} because only w_{US} enters unit costs of production as well as cost cutoffs. Since we normalize all per-capita incomes (and sizes) relative to the US, this would imply that both $w_{US} = 1$ and $e_{US} = 1$. Notice that by construction this implies that we need to normalize all S_k's relative to S_{US} so that $S_{US} = 1$. Because $\log(S_k)$ is the object that we estimate from gravity, we first exponentiate this object for every k and then we divide it through by the exponent of the object for the US.

D.5 Quantitative results: moments, parameters and fit

Table 1: Moments and Parameters

Moment	Model	Data	Source
US Exporter Productivity Advantage	0.35	0.33	BEJK
US Exporter Sales Advantage	4.81	4.8	BEJK
Mean Price Elasticity of Income	0.43	0.14	Simonovska
Population, relative to US	(N-1)x1 vector		WDI
Bilateral trade shares	Gravity		Comtrade
Parameter	**Value**		
γ	1.919		
κ	2.772		
L	(N-1)x1 vector		
τ	Gravity		

Table 1 summarizes the three moments that we target, the data sources, as well as the resulting parameters that match those moments.[54] The price elasticity moment is the preferred estimate obtained by Simonovska (2015).[55] In this overidentified approach, we employ the identity matrix to weigh the three moments that the two parameters attempt to jointly match.

[53]Results obtained using the optimal weighting matrix are available upon request.

[54]The firm-level moments reported in Bernard *et al.* (2003) are for the universe of US firms in 1992. Bernard *et al.* (2007) document similar statistics using 2002 US firm-level data.

[55]In Table 1 of their paper, Alessandria and Kaboski (2011) estimate this parameter using HS-10 digit unit value data for US exports of final/consumption goods recorded at the port of shipping to 28 destinations during the 1989-2000 period. While the prices do not include shipping costs and destination-specific non-tradable components, they may reflect quality variations, which is why we opt to target the moment in Simonovska (2015) instead. Two observations are in order. First, our model predicts that the mean price elasticity is country invariant; thus, combining the sales moment for the US with the price elasticity moment for Spain is not problematic. Second, re-estimating the model with Alessandria and Kaboski's (2015) moment as a target yields nearly identical results.

When we let κ and γ jointly match moments from the sales, measured productivity, and price elasticity distribution, the estimated γ amounts to 1.92, while κ centers around 2.77.

Table 2: Predicted VS. Actual Moments

Moments	Model	Data	Source
US Exporters, % All firms	16.27	18 - 21	Bernard et al., BEJK
Export Intensity (%)			BEJK
0-10	80.2	66	
10-20	10.5	16	
20-30	4.4	7.7	
30-40	2.0	4.4	
40-50	1.2	2.4	
50-60	0.7	1.5	
60-100	1.0	2.8	
Mark-up, mean, %	19.34	5 - 40	Jaimovich & Floetotto
Cost pass-through, mean, %	0.57	0.36 - 0.57	De Loecker et al.
Log E, rel. US, mean	-3.03	-2.70	WDI
(standard deviation)	(2.67)	(1.66)	
Moments	Corr	(model,data)	
Log E, 123 countries	0.87		
λ , 123 countries	0.91		

In Table 2 we explore the basic fit of the estimated model to a variety of moments. Unlike the moments displayed in Table 1, which we target in the identification of the model's parameters, the moments in Table 2 serve for external validation as they are not targeted. The model predicts that the share of US firms that export is roughly 16%, in line with US data for year 1992 reported in Bernard *et al.* (2003). The resulting export intensity in the model is even more skewed than observed in the data and it reflects the prediction that a large number of US firms that export sell very little abroad — that is to say most exporters sell tiny amounts abroad even though they enjoy a large domestic sales advantage over non-exporters.

Furthermore, the mean price elasticity with respect to the real exchange rate is equivalent to the mean price elasticity with respect to per-capita income of 0.43 (the median being 0.4), while the mean plus one standard deviation estimate is 0.51. We interpret these elasticities as equilibrium elasticities for the broad manufacturing sector. These estimates are qualitatively in line with, but exceed in magnitude, the findings in Berman *et al.* (2012) for a set of French exporters to non-Eurozone destinations.

The average cost pass-through predicted by the model is exactly one minus the pricing-to-market elasticity reported above and amounts to 0.57, which is within the range of 0.36-0.57 reported by De Loecker *et al.* (2015) for Indian manufacturing firms. Finally, given our estimated parameters, the average markup amounts to 19%, which is in line with common findings in the macroeconomic literature (see Jaimovich and Floetotto, 2008).

D.5.1 Prices across destinations

Our model predicts that prices are increasing in destination per-capita income and are independent of destination population size. Given our estimates of κ and γ, the mean elasticity of price with respect to per-capita income is 0.43. In turn, the price elasticity with respect to population is zero. These estimates compare qualitatively to estimates reported in the empirical literature, although the predicted elasticity of price with respect to income does exceed the corresponding moment in the data. In particular, Simonovska (2015) finds that a Spanish apparel retailer systematically price discriminates according to the per-capita income of destinations, but does not vary prices across countries of different population sizes. The typical estimate of the elasticity of price with respect to income that the author obtains circles around 0.14, which corresponds to one of the targets that we use in our estimation. While the predicted moment exceeds the target, we interpret the model as representing the broad manufacturing sector rather than apparel alone.

Estimates of the elasticity of price with respect to destination income and population size for a broad set of manufacturing products are not available as detailed price data as the dataset employed by Simonovska (2015) are only available for a handful of producers/sectors. In the working paper version we find supporting evidence in favor of the author's findings, which are in line with the predictions of the IA model, using retail price data for products with identical characteristics sold in different cities around the world. A key limitation of the data is that they are not sufficiently detailed so as to be able to argue that price variation across destinations is entirely due to pricing to market; in particular, prices may differ across destinations due to differences in quality or non-tradable components as well. We leave for future research to arrive at estimates of the elasticity of price with respect to per-capita income and market size across different industries.

D.5.2 Aggregate moments

In this section, we present the model's fit to aggregate moments. The model generates per-capita income levels that are at par with the data. In particular, the model falls somewhat short of the mean per-capita income level among 123 countries, but it yields a higher variance. Despite the dispersion, the model's predictions line up with the data, as the correlation of the predicted and actual per-capita income among 123 countries is 0.87 (in logs). Figure 2 gives a visual representation of the model's fit along the income dimension. While countries line up along the 45-degree line, which represents a perfect fit, the model underpredicts the income levels of the poorest set of countries. Since per-capita incomes are chosen so as to match observed trade flows in the market clearing equation, this result may be due to the fact

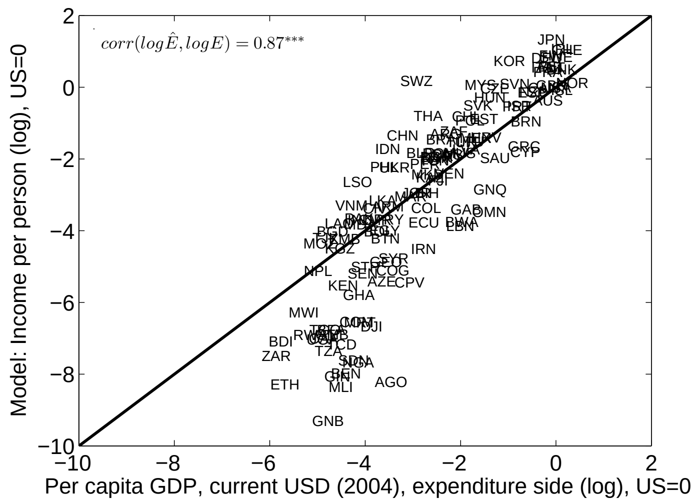

Figure 2: Predicted VS. Actual Per-Capita Income, 123 Countries

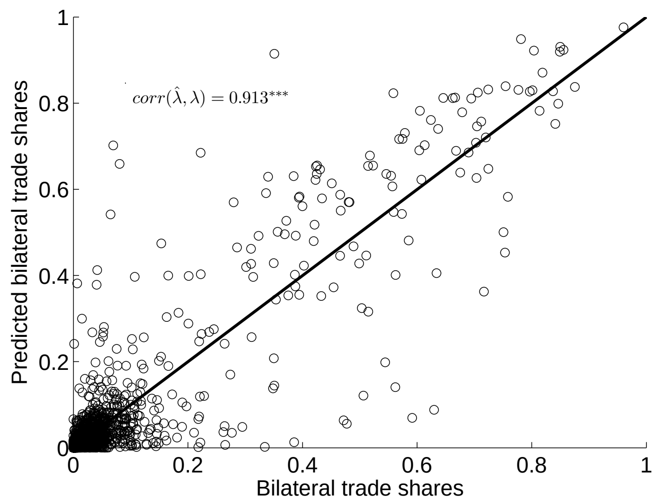

Figure 3: Predicted VS. Actual Trade Shares, 123 Country Pairs

that these countries have relatively low import and export shares, even conditional on trade barrier levels. This would suggest that these countries may simply be plagued by very low productivities.

Figure 3 plots (non-zero) predicted against actual bilateral trade shares for all country pairs. A large cluster of bilateral trade shares can be seen around the origin representing the fact that, for the majority of countries, each individual destination accounts for a tiny fraction of its total sales. On the other hand, large numbers that are dispersed around the top right corner mostly capture domestic expenditure shares. Despite the large variation in trade shares, the model can match the cross-section of trade shares quite well due to the flexible specification for trade costs in the structural gravity equation.

D.5.3 The margins of trade

In this section, we quantify the model's predictions about the extensive and intensive margin of trade. Recall from (60) that the model predicts that the elasticity of the extensive margin with respect to destination per-capita income equals κ, while the same elasticity with respect to trade costs equals $-\kappa$. Since trade barriers are increasing in distance, our model's predicted elasticity with respect to distance is necessarily negative.

Table 3: Predicted US Extensive Margin of Trade

	(1)
Log(pcincome)	2.779***
	(0.117)
Log(L)	0.046
	(0.073)
Log(distance)	-1.890***
	(0.294)
R^2	0.93
# Observations	61

Notes: All variables relative to Mexico—the most popular US export destination in terms of number of exported products. *** indicate significance at 1%-level. Standard errors in parentheses.

In Table 3, we quantify the elasticity with respect to distance. Since the extensive margin in the model is source-destination specific, we focus on the US as a source country. We regress the predicted extensive margin on destination per-capita income, size, and distance from the US, all in logs. The estimated elasticities with respect to the three variables are 2.8, 0.05, and -1.9, respectively, and only the first and the last are statistically significant. The coefficients on per-capita income and distance are consistent with the findings in Bernard *et al.* (2007) for US data. In particular, the authors document that the elasticity of the number of exported products by US exporters with respect to destination GDP is 0.52 and with respect to distance is -1.06. While the authors do not decompose the elasticity with respect to GDP into the two components: per-capita income and population size, our model suggests that the positive slope in the data may be due to the per-capita income component.

Along the intensive margin dimension, the model predicts that, controlling for aggregate effects, (i) the elasticity of the intensive margin with respect to destination GDP is 1; (ii) the elasticity of the intensive margin with respect to destination per-capita income is $-\kappa$, or -2.8 given our estimate. Accordingly, in our model the intensive margin of trade is increasing in a destination's overall GDP and decreasing in the destination's per-capita income, which can reconcile findings in Eaton *et al.* (2011). Eaton *et al.* (2011) find that the intensive margin (defined as average per-firm sales) is increasing in destination GDP and either increasing or decreasing in destination per-capita GDP depending on the source country analyzed.

Both findings are potentially in line with the existing empirical literature; however, the literature typically does not distinguish between the effects that per-capita income and market size have on the margins of trade. We leave for future research to conduct empirical investigations using detailed firm-level data across different industries (characterized by differing magnitudes of fixed costs) to help discern the role of these two variables in driving the margins of trade so as to better evaluate the performance of different models.

D.6 Counterfactual: trade cost reduction between USA and EU

Given the IA model's favorable performance with respect to data, in this section, we use the estimated version to evaluate the welfare gains from a bilateral reduction in trade costs between the US and the European Union.

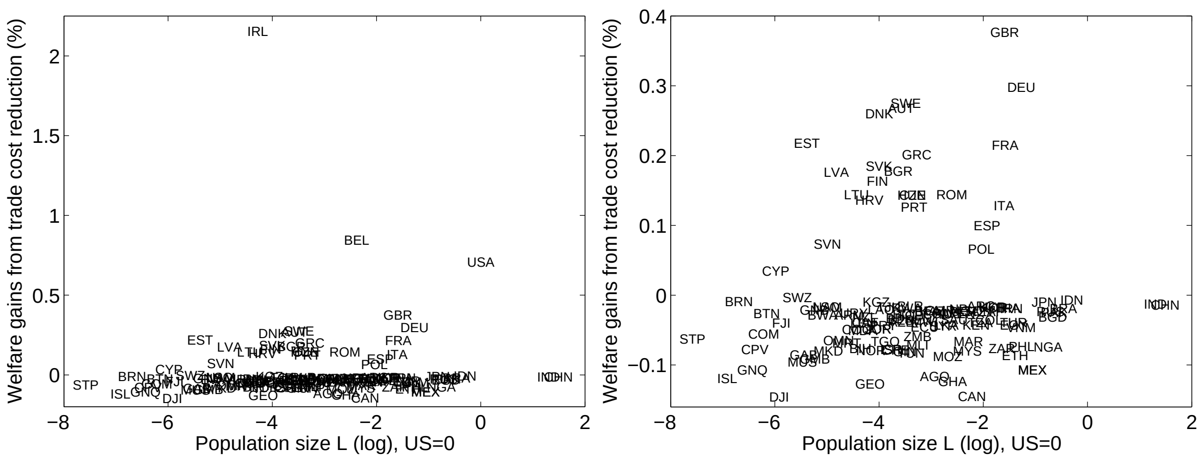

Figure 4: Welfare Gains From Trade Cost Reduction Between USA and EU, 123 Countries

To quantify the gains from bilateral trade cost reductions, we proceed as follows: First, we set the RTA indicator in the trade-cost function in expression (66) to unity for the country pairs that involve the USA and each of the EU members. Second, we use the gravity coefficient estimates, as well as the estimate of κ, to compute new bilateral trade barriers. Third, we compute the percent reduction on trade barriers for the US and the EU. The mean percent reduction in trade barriers among these countries is 16%, while the trade barriers for non-EU countries remain unchanged by construction. Finally, to compute the welfare gains, we plug the computed change in trade barriers into the system (27), using actual trade shares and predicted income, which jointly satisfy the market clearing conditions given by the system of equations in (23).

We report the results for all the countries in the left panel of Figure 4. Clearly, the EU and US gain from the trade cost reduction, but the gains are asymmetric. The US enjoys welfare gains of roughly 0.7% and Belgium (alongside Luxembourg and the Netherlands) gains by roughly 0.8%. Ireland is the biggest winner with gains amounting to more than 2%. To obtain a better sense of the results, in the right panel of Figure 4, we zoom in on the countries with gains below 0.4%. The mean gains among the free-trade agreement members are 0.3% with a standard deviation of 0.4%. Non-members suffer losses which amount to an average of -0.04%. Among non-members, USA's major trade partners Mexico and in particular Canada experience some of the largest losses. Overall, however, the gains far exceed the losses in world welfare.

E Data appendix

E.1 Gravity equation

The description below follows closely the work of Simonovska and Waugh (2014a). To construct trade shares, we used bilateral trade flows and production data as follows:

$$\lambda_{ij} = \frac{\text{Imports}_{ij}}{\text{Gross Mfg. Production}_j - \text{ Exports}_j + \text{Imports}_j},$$

$$\lambda_{jj} = 1 - \sum_{k \neq j}^{I} \lambda_{kj}.$$

To construct λ_{ij}, the numerator is the aggregate value of manufactured goods that country j imports from country i. Bilateral trade-flow data are for year 2004 from the update to Feenstra *et al.* (2005), who use UN Comtrade data. We obtain all bilateral trade flows for our sample of 123 countries at the four-digit SITC level. We then used concordance tables between four-digit SITC and three-digit ISIC codes provided by the UN and further modified by Muendler (2009).[56] We restrict our analysis to manufacturing bilateral trade flows only—namely, those that correspond with manufacturing as defined in ISIC Rev.#2.

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

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

$$\log\left(\frac{MVA}{GO}\right) = \beta_0 + \boldsymbol{\beta}_{GDP}\mathbf{C}_{GDP} + \boldsymbol{\beta}_L\mathbf{C}_L + \boldsymbol{\beta}_{MVA}\mathbf{C}_{MVA} + \boldsymbol{\beta}_{AVA}\mathbf{C}_{AVA} + \epsilon,$$

where β_x is a 1×3 vector of coefficients corresponding to C_x, an $N \times 3$ matrix which contains $[\log(x), (\log(x))^2, (\log(x))^3]$ for the sub-sample of N countries for which gross output data are available. Data on geographic barriers (distance, shared border, official common language, colonial relationship, common currency and RTA) are from Head *et al.* (2010). Data on population size for year 2004 is from the World Development Indicators. Data on per-capita income is from Feenstra *et al.* (2013) (Penn World Tables 8.0).

F Additional tables

Table 4: Gravity equation: Estimates

Barrier	Parameter Estimates	S.E
Log distance	-1.30	0.03
Border shared	0.75	0.11
Official Common Language	1.06	0.06
Colonial Relationship	1.35	0.08
Common Currency	-0.08	0.15
RTA	0.48	0.06
Internal trade	1.46	0.22
# Observations	15,129	
TSS	160,320	
SSR	27,694	
σ_ν^2	2.67	

Table 7: Gravity equation: Estimates

Country	$\hat{S}_i$	S.E	ex_i	S.E.	Country	$\hat{S}_i$	S.E	ex_i	S.E.	Country	$\hat{S}_i$	S.E	ex_i	S.E.
Angola	-1.03	0.2	-2.62	0.33	Fiji	-0.47	0.19	-2.32	0.3	Nepal	0.42	0.22	-2.83	0.31
Argentina	1.01	0.17	2.63	0.23	Finland	0.96	0.16	2.41	0.22	New Zealand	-0.38	0.16	3.45	0.23
Armenia	0.67	0.19	-3.58	0.28	France	0.33	0.15	5.19	0.21	Nigeria	-0.66	0.19	-1.44	0.28
Australia	0.15	0.16	3.78	0.22	Gabon	-0.94	0.18	-1.81	0.26	Norway	0.14	0.16	2.24	0.22
Austria	0.23	0.15	3.01	0.22	Gambia, The	-2.07	0.21	-3.01	0.32	Oman	-0.3	0.18	-0.49	0.25
Azerbaijan	-0.17	0.19	-2.52	0.27	Georgia	-3.1	0.18	1.37	0.26	Pakistan	0.77	0.15	1.61	0.22
Bangladesh	0.79	0.17	0.42	0.23	Germany	0.21	0.15	5.95	0.21	Paraguay	0.01	0.19	-0.68	0.27
Belarus	1.12	0.17	-0.66	0.24	Ghana	-1.07	0.2	-0.10	0.28	Peru	0.38	0.17	1.32	0.24
Belgium	-2.08	0.15	7.55	0.21	Greece	0.45	0.16	1.23	0.22	Philippines	-0.33	0.17	2.60	0.23
Benin	-0.56	0.21	-3.93	0.34	Guinea	-1.53	0.21	-2.62	0.31	Poland	0.61	0.15	2.20	0.22
Bhutan	0.19	0.28	-5.04	0.41	Guinea-Bissau	-0.47	0.27	-5.60	0.45	Portugal	-0.34	0.16	2.99	0.22
Bolivia	0.26	0.18	-1.60	0.27	Hungary	0.66	0.16	1.39	0.22	Romania	0.33	0.16	1.26	0.22
Bosnia and Herzegovina	0.72	0.22	-2.84	0.31	Iceland	-0.27	0.17	-0.53	0.25	Russian Federation	1	0.16	2.74	0.22
Botswana	1.27	0.24	-4.36	0.35	India	1.19	0.15	3.02	0.24	Rwanda	0.42	0.23	-5.68	0.35
Brazil	1.13	0.15	3.99	0.22	Indonesia	1.16	0.16	3.44	0.22	Sierra Leone	-0.8	0.27	-4.04	0.39
Brunei Darussalam	1.76	0.24	-5.36	0.35	Iran, Islamic Rep.	0.68	0.2	-0.18	0.27	Saudi Arabia	0.52	0.19	1.05	0.26
Bulgaria	0.03	0.16	0.92	0.23	Ireland	-3.14	0.15	6.26	0.22	Senegal	-0.57	0.16	-1.20	0.24
Burkina Faso	0.46	0.19	-4.36	0.29	Israel	0.89	0.17	1.10	0.23	Slovak Republic	-0.5	0.16	1.70	0.22
Burundi	-1.5	0.19	-3.19	0.32	Italy	0.26	0.15	5.18	0.22	Slovenia	0.77	0.17	0.30	0.23
Cameroon	1.79	0.2	-3.91	0.29	Japan	1.22	0.15	5.47	0.22	South Africa	0.51	0.15	3.42	0.22
Canada	-0.01	0.15	4.06	0.22	Jordan	-0.17	0.17	-0.84	0.24	Spain	0.18	0.15	4.30	0.21
Cape Verde	-0.44	0.2	-4.83	0.36	Kazakhstan	0.19	0.17	0.13	0.25	Sri Lanka	-0.1	0.17	0.57	0.24
Central African Republic	0.6	0.24	-4.89	0.35	Kenya	-0.2	0.16	-0.75	0.22	Sudan	-0.09	0.2	-3.59	0.3
Chad	0.66	0.23	-6.61	0.39	Korea, Rep.	0.77	0.15	4.91	0.21	Swaziland	2.48	0.23	-4.00	0.31
Chile	0.29	0.18	1.98	0.25	Kyrgyz Republic	0.05	0.19	-2.96	0.28	Sweden	0.63	0.15	3.58	0.22
China	0.89	0.15	6.23	0.22	Lao PDR	1.23	0.26	-3.56	0.34	Switzerland	0.09	0.18	3.70	0.26
Colombia	0.23	0.16	0.77	0.23	Latvia	-0.42	0.18	-0.12	0.25	Syrian Arab Republic	-0.35	0.18	-0.90	0.25
Comoros	-0.8	0.26	-4.74	0.4	Lebanon	0.58	0.19	-2.24	0.26	Tajikistan	1.09	0.24	-3.14	0.33
Congo, Dem. Rep.	-0.66	0.23	-2.31	0.33	Lesotho	1.64	0.29	-6.35	0.42	Tanzania	-0.69	0.21	-2.07	0.3
Congo, Rep.	-0.82	0.2	-1.36	0.29	Lithuania	0.7	0.2	-0.95	0.28	Thailand	0.55	0.19	4.19	0.26
Côte d'Ivoire	0.96	0.2	-1.60	0.28	Macedonia, FYR	0.15	0.18	-2.22	0.26	Togo	-1.22	0.17	-1.77	0.26
Croatia	0.76	0.16	-0.68	0.23	Malawi	-0.17	0.18	-3.50	0.27	Tunisia	0.52	0.16	-0.65	0.23
Cyprus	-0.83	0.17	0.34	0.23	Malaysia	-1.04	0.15	6.19	0.22	Turkey	0.63	0.16	2.95	0.22
Czech Republic	0.24	0.15	2.38	0.22	Mali	-0.95	0.22	-2.73	0.3	Uganda	-0.4	0.17	-2.81	0.25
Denmark	-0.4	0.16	3.95	0.22	Mauritania	-1.97	0.22	-1.78	0.31	Ukraine	1.11	0.19	1.47	0.27
Djibouti	-1.85	0.23	-2.72	0.37	Mauritius	-1.07	0.17	0.32	0.23	United Kingdom	-0.21	0.15	5.44	0.21
Ecuador	-0.31	0.17	0.23	0.25	Mexico	0.21	0.15	2.58	0.23	United States	0.13	0.15	6.73	0.21
Egypt, Arab Rep.	0.28	0.16	0.94	0.22	Moldova	-0.65	0.18	-1.79	0.28	Uruguay	-0.56	0.19	1.52	0.25
Equatorial Guinea	0.6	0.23	-4.50	0.38	Morocco	-0.06	0.16	0.67	0.22	Venezuela, RB	0.61	0.18	-0.39	0.25
Estonia	-1.72	0.16	1.58	0.23	Mozambique	-0.36	0.21	-1.69	0.31	Vietnam	-0.62	0.2	3.02	0.27
Ethiopia	-0.58	0.2	-2.34	0.29	Namibia	1.15	0.22	-3.83	0.31	Zambia	-3.61	0.17	1.79	0.26