NBER WORKING PAPER SERIES

FROM HETEROGENEOUS FIRMS TO HETEROGENEOUS TRADE ELASTICITIES: THE AGGREGATE IMPLICATIONS OF FIRM EXPORT DECISIONS

Rodrigo Adão
Costas Arkolakis
Sharat Ganapati

Working Paper 28081
http://www.nber.org/papers/w28081

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2020, Revised December 2025

This paper previously circulated with the title “Aggregate implications of firm heterogeneity: A nonparametric analysis of monopolistic competition trade models.” We thank Xiaohong Chen, Tim Christensen, Xiang Ding, Phil Haile, Keith Head, Federico Huneeus, Giovanni Compiani, Chang-Tai Hsieh, Samuel Kortum, Stephen Redding, and Yacan Wang for valuable comments. We thank Takashi Onada, Yongjie Xiong, Mengly Xu, Ryan Thurz, Sviatoslav Tiupin, and Linghui Wu for research assistance. We are grateful to Peter Morrow, Ana Fernandes and Feodora Teti for sharing data. The support of the National Science Foundation under grant 2018234 is gratefully acknowledged. Part of this research was conducted while Ganapati was a Fellow at Dartmouth College. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

From Heterogeneous Firms to Heterogeneous Trade Elasticities: The Aggregate Implications of Firm Export Decisions
Rodrigo Adão, Costas Arkolakis, and Sharat Ganapati
NBER Working Paper No. 28081
November 2020, Revised December 2025
JEL No. C14, F12, F14, F15

ABSTRACT

We study the consequences of globalization in monopolistic competition models where heterogeneous firms select into export markets. We summarize firm heterogeneity at the extensive and intensive margins of firm exports with two nonparametric elasticity functions that depend only on the share of active firms in a market. Given changes in trade costs, these elasticity functions are sufficient to compute the model's counterfactual predictions, and their shape generates heterogeneity in welfare responses across countries. We estimate these functions using the model's semiparametric gravity equations for firm export margins, which yield trade elasticity estimates that vary with the number of exporters in a market and the country's level of development. Compared to constant-elasticity gravity models, our estimates imply gains from trade that are larger in developed countries but smaller in developing countries, with differences arising mainly due to the entry and selection of heterogeneous firms.

Rodrigo Adão
University of Chicago
Booth School of Business
Department of Economics
and NBER
rodrigo.adao@chicagobooth.edu

Costas Arkolakis
Yale University
Department of Economics
and NBER
costas.arkolakis@yale.edu

Sharat Ganapati
Georgetown University
and CESifo
and also NBER
sg1390@georgetown.edu

A data appendix is available at http://www.nber.org/data-appendix/w28081

1 Introduction

International trade reflects the endogenous decisions of heterogeneous firms to select which countries to export to. Larger, more productive firms are more likely to export and their expansion induces smaller firms to exit (Melitz, 2003; Bernard and Jensen, 2004; Melitz and Redding, 2014). However, the relationship between firms and trade likely varies across countries, as large, high-productivity firms are more prevalent in developed economies (Hsieh and Olken, 2014). How does the decision of heterogeneous firms to export shape the impact of globalization across different countries? Following trade cost changes, do differences in firm characteristics translate into differences in aggregate responses?

Although the literature acknowledges the role of firm heterogeneity for selection into exporting and new firm entry, it often assigns this role a secondary importance when quantifying globalization gains. This is due to strong parametric restrictions on firm heterogeneity. Such restrictions can be useful: they easily link available data to tractable counterfactual predictions. But this is not a free lunch: they also come at the price of limiting how firm heterogeneity shapes responses to changes in trade costs. A canonical example is the assumption of a Pareto distribution for firm productivity, which implies that gains from trade are identical in neoclassical and heterogeneous-firm models and that new firm entry is invariant to globalization (Chaney, 2008; Arkolakis, Costinot and Rodríguez-Clare, 2012). Generally, parametric distributional assumptions determine both the aggregate implications of firm heterogeneity and the set of moments used for identification (e.g., Eaton, Kortum and Kramarz, 2011; Melitz and Redding, 2015; Head, Mayer and Thoenig, 2014; Bas, Mayer and Thoenig, 2017; Fernandes, Klenow, Meleshchuk, Pierola and Rodríguez-Clare, 2023).

In this paper, we propose a new methodology to analyze the aggregate consequences of globalization in monopolistic competition models without parametric restrictions on firm heterogeneity. We consider an extension of Melitz (2003) where heterogeneous firms in an origin decide whether to create a new variety (entry) and which destinations to sell it to (selection). We allow for an arbitrary distribution of firm fundamentals; namely, the joint distribution of destination-specific shifters of productivity, demand, and costs. Our setting generates rich patterns of heterogeneity in firm export decisions, both within and between destinations, such as those documented by Eaton et al. (2004, 2011) and Fernandes et al. (2023). To focus on the aggregate implications of these decisions, we abstract from markup heterogeneity across firms and maintain constant elasticity of substitution (CES) preferences, as in the extensive literature reviewed by Melitz and Redding (2014).

In this environment, we derive two functions that summarize all specified sources of firm heterogeneity, and govern firm exports to a destination through the extensive and intensive margins. These functions determine the elasticities of the two margins with respect to bilateral trade costs in semiparametric gravity equations, where origin and destination fixed-effects absorb endogenous country-level outcomes.

Our first main result is that the trade elasticity—the elasticity of trade flows with respect to trade costs, which is the sum of these two margins—can be expressed as a univariate function of the exporter firm share, the fraction of firms from the origin that sell to a given destination. The exporter firm share identifies the marginal firms whose responses to trade costs govern both exporter selection (extensive margin) and sales (intensive margin). The trade elasticity depends on the difference between marginal and inframarginal firms at the initial equilibrium, summarized by the decay rate of the distribution of firm-level shifters of entry and sales. Accordingly, the trade elasticity rises with the exporter firm share when the decay rate is steep, but remains constant or decreases when the decay rate is more gradual. A faster decay rate implies a lower relative mass of high-potential firms, which weakens export responsiveness in competitive markets where such firms are marginal exporters.[1]

Our second main result is that the same elasticity functions summarize how firm heterogeneity shapes the economy's aggregate response to changes in trade costs. Given these elasticity functions, the model's counterfactual predictions do not depend on its micro structure, including the joint distribution of firm fundamentals.[2] We further show that these functions govern how trade costs affect welfare, entry, and selection, both qualitatively and quantitatively. Their shape determines the extent to which selection into foreign markets leads to selection out of the domestic market, and whether globalization increases the number of firms in a country. Importantly, gains from trade depend on the shape of the entire elasticity function. In particular, gains tend to be higher when the trade elasticity function is decreasing with the exporter firm share. This implies that parametric assumptions that mis-specify how the trade elasticity varies with the exporter firm share lead to biased predicted welfare responses.

These theoretical results underscore the importance of reliable estimates of the trade elasticity as a function of the exporter firm share. It is important to let the data determine

[1] This result builds on the well-established observation that the trade elasticity is the sum of extensive and intensive margins of firm exports and is generally non-constant, except under a Pareto distribution of firm productivity (see Chaney, 2008; Head et al., 2014; Melitz and Redding, 2015; Bas et al., 2017; Fernandes et al., 2023). In fact, it decreases with the exporter firm share for several widely used distribution families, including truncated Pareto, multivariate log-normal, and mixtures of Pareto and log-normal (Eaton et al., 2011; Melitz and Redding, 2015; Bas et al., 2017; Fernandes et al., 2023). To our knowledge, however, prior work has not characterized the trade elasticity as a univariate function of the exporter firm share nor linked it to semiparametric gravity equations for firm export margins.

[2] These results directly generalize for models with multiple sectors, multiple factors, and multi-product firms. We also extend our results to allow for heterogeneous markups with single-aggregator demand functions, which requires additional elasticity functions in gravity equations for percentiles of firm sales.

the shape of the trade elasticity functions with minimal structural assumptions. This ensures that counterfactual predictions are driven by the estimated responses of firm export decisions, rather than parametric restrictions on firm heterogeneity. Accordingly, we extend conventional gravity tools (see e.g. Head and Mayer (2014)) to estimate the model's semiparametric gravity equations for the firm export margins. Our strategy identifies the trade elasticity functions from cross-market variation in the impact of trade cost shifters on firm export margins, given a market's exporter firm share. In practice, we implement a flexible semiparametric estimator that specifies the trade elasticity as a restricted cubic spline function of the exporter firm share.

We conduct Monte Carlo simulations to illustrate the properties of our estimator and highlight the limitations of restrictive parametric assumptions. We construct three economies in which the trade elasticity varies with the exporter firm share—increasing, constant, or decreasing. In all cases, our semiparametric gravity estimator delivers consistent estimates of the trade elasticity function and the gains from trade. In contrast, parametric approaches can be severely biased when they mis-specify the slope of the trade elasticity function, even if they perfectly fit the observed cross-sectional distribution of firm exports.

We apply our strategy to estimate trade elasticities across countries. We find that the trade elasticity varies systematically with the exporter firm share and the country's development level. In developing countries, trade flows are less responsive to trade costs in markets with fewer active firms than in those with more active firms. In contrast, developed countries display the opposite pattern: they have a larger responsiveness to trade costs in markets where only the few, most efficient firms are active.[3] Our estimates are statistically inconsistent with the trade elasticity functions implied by popular parametric restrictions on firm heterogeneity. However, they are consistent with the evidence in Hsieh and Klenow (2009) and Hsieh and Olken (2014) that, relative to developed countries, developing countries have a relatively lower mass of large firms.

We conclude by quantifying how the export decisions of heterogeneous firms affect the gains from trade across countries. Compared to a constant-elasticity benchmark, our estimates yield larger welfare gains in developed countries (by an average of 22%) but smaller gains in developing countries (by an average of 17%). These differences arise entirely from the entry and selection responses of domestic firms, as implied by the shape of our estimated elasticities of firm export margins. Similar patterns emerge for small, uniform reductions in trade costs across countries. We also show that the differences in welfare predictions can be even larger under alternative parametric specifications whose implied trade elasticity slopes

[3] We also show that these patterns are invariant to the sectoral composition of trade flows, and to other country characteristics (like their level of trade integration).

are rejected by our estimates.

Our paper is related to the extensive theoretical and empirical literature on firms in international trade (for reviews, see Bernard et al. (2007); Redding (2011); Melitz and Redding (2014)). From a theoretical perspective, we build on the insights in Chaney (2008), Arkolakis et al. (2012) and Melitz and Redding (2015) that the value of the trade elasticity summarizes the effect of firm heterogeneity on trade and welfare. Our main contribution is to generalize these insights without parametric restrictions on the distribution of firm fundamentals: we show that the elasticity functions of the intensive and extensive margins of firm exports summarize the general-equilibrium implications of firm entry and selection. These elasticity functions capture the components of the joint distribution of firm fundamentals that are sufficient for the model's counterfactual predictions to changes in trade costs.[4]

We use these elasticity functions to characterize the properties of monopolistic competition models with CES demand, as well as the welfare gains from globalization. Our work extends the decomposition proposed by Atkeson and Burstein (2010) to an environment with multiple asymmetric countries, domestic selection, and arbitrary heterogeneity. For any given country, welfare gains from firm entry and selection are non zero to a first-order when countries are asymmetric, but they are indeed second-order when countries are symmetric. We also derive a nonparametric extension of the sufficient statistics in Arkolakis et al. (2012) and Melitz and Redding (2015). It indicates that what matters for the gains from trade is not the (constant) "trade elasticity", but instead the (variable) "domestic elasticity" and the entry and selection decisions of domestic firms. As in Costinot and Rodríguez-Clare (2018), the gains from trade correspond to the area below the import demand curve, which incorporates variation in the domestic trade elasticity along the path from trade to autarky.

Empirically, our semiparametric approach bridges the literature that estimates constant-elasticity gravity trade models, reviewed by Head and Mayer (2014), and the one that uses granular data on firm outcomes to estimate the distribution of firm fundamentals behind exporting decisions (e.g. Head et al. (2014) and Egger et al. (2023) for productivity; Eaton et al. (2011) and Fernandes et al. (2023) for demand and trade costs). As Bas et al. (2017) point out, the "micro" heterogeneity in fundamentals across firms gives rise to "macro" heterogeneity in trade elasticities across markets. We extend existing tools to estimate variable trade elasticities in semiparametric gravity equations, while accounting for export decisions of

[4]Of course, the joint distribution of firm fundamentals identifies the two elasticity functions and the model's counterfactual predictions. However, the reverse is not true: the elasticity functions do not identify the joint distribution of firm fundamentals. For example, conditional on the elasticity functions, one does not need to know either the cross-destination correlation in firm fundamentals (like preference and trade costs) or the dispersion of firm revenue fundamentals conditional on entry determinants. In this sense, the elasticity functions reduce the dimensionality of the requirements for counterfactual analysis.

heterogeneous firms. We find that trade elasticities vary with the number of exporters and the country's development level. Our estimated elasticity functions are statistically different from those implied by "micro-to-macro" approaches that leverage parametric distributions of firm fundamentals to match the distribution of firm-level outcomes. We thus complement the literature estimating variable elasticities using parametric extensions of gravity models– e.g., Novy (2013), Fajgelbaum and Khandelwal (2016), and Lind and Ramondo (2018). Finally, we contribute to a literature that measures gains from varieties, as in Broda and Weinstein (2006) and Feenstra and Weinstein (2017), by leveraging our variable elasticity estimates to measure the welfare implications of firm entry and selection.

Our empirical approach builds upon recent advancements on the nonparametric identification of models with self-selection of heterogeneous agents (Berry and Haile, 2014; Adão, 2015). It provides a way of circumventing the difficulty in identifying the high-dimension distribution of firm fundamentals in the presence of firm selection into export markets, which has been typically tackled in practice with the use of strong parametric restrictions on firm heterogeneity. Our approach addresses the well-known challenge that observed outcomes among active firms in a market (i.e. cross-sectional moments) are insufficient to nonparametrically identify the distribution of fundamentals for firms that are not active in that market (Heckman and Honore, 1990).[5] Instead, we exploit *cross-market* variation in firm export margins induced by trade costs in order to nonparametrically identify the elasticity functions that summarize the role of firm heterogeneity in general equilibrium. Our simulations illustrate that this is relevant in practice. While our semiparametric approach consistently estimates the "macro" trade elasticity function for any distribution of firm fundamentals, severe bias may arise in existing parametric approaches that replicate well the "micro" cross-section distribution of firm exports but mis-specify how the trade elasticity varies with the exporter firm share.

Our work is closely related to recent papers conducting nonparametric counterfactual analysis in international trade models (Adao et al., 2017; Bartelme et al., 2019).[6] These flexible approaches require knowledge of multivariate functions whose nonparametric estimation is challenging in finite samples – for example, Adao et al. (2017) must estimate each country's demand function for all factors in the world economy. Compared to these papers, we consider

[5]In this sense, our estimation approach differs from those that measure firm heterogeneity by looking at the observed distribution of size and export decisions of active firms, such as Eaton et al. (2011), or other cross-sectional moments such as pass-through by firm size, as in Baqaee et al. (2024). It follows the insight in Heckman and Honore (1990) to deal with the problem that one does not observe outcomes in a destination for firms that decide not to operate there in the current equilibrium.

[6]Our paper also complements the literature offering sufficient statistics in neoclassical and gravity trade models (Allen et al., 2014; Baqaee and Farhi, 2019; Kleinman et al., 2024). This alternative approach relies on parametric assumptions for empirical and counterfactual analyses.

a different class of models that feature monopolistic competition. Our methodology has the advantage of only requiring the estimation of univariate elasticity functions.

Finally, we study the aggregate implications of selection of heterogeneous firms into exporting markets, as in Melitz (2003). While we extend our nonparametric approach to incorporate firm heterogeneity in markups, sectors, inputs, and products, we abstract from other potentially important dimensions of firm heterogeneity, such as skill intensity (Burstein and Vogel, 2017), sourcing locations (Blaum et al., 2015; Antras et al., 2017), information sets (Dickstein and Morales, 2018), product quality (Kugler and Verhoogen, 2008), innovation (Atkeson and Burstein, 2010; Bustos, 2011), and oligopolistic pricing behavior in granular settings (Atkeson and Burstein, 2008; Berman et al., 2012).

Our paper is organized as follows. Section 2 characterizes the trade elasticity functions that summarize the aggregate implications of the export decisions of heterogeneous firms. In Section 3, we show how to use these functions for computing counterfactual predictions to changes in trade costs. Section 4 outlines the methodology to estimate the trade elasticity functions using the model's semiparametric gravity equations for the margins of firm exports. We report estimates of the trade elasticity functions in Section 5, and counterfactual exercises in Section 6. Section 7 concludes.

2 From Heterogeneous Firms to Heterogeneous Trade Elasticities

We consider an economy with monopolistically competitive firms that differ in destination-specific shifters of productivity, demand, and trade costs. In this environment, we study how the endogenous export decisions of heterogeneous firms shape aggregate trade flows and welfare in general equilibrium. These aggregate effects are fully characterized by two functions that describe how the elasticities of firm export margins to trade costs vary with the number of exporters in a market. The shape of these elasticity functions depends on the relative mass of high-potential firms in terms of entry and sales. Standard parametric assumptions in the literature impose strong restrictions on the shape of these elasticity functions.

2.1 Environment

We first describe consumers and then the firm's problem and decisions.

Preferences. Each country j has a representative household that inelastically supplies $\bar{L}_j$ labor units, and has CES preferences over varieties ω. Demand is subject to a bilateral taste shifter $\bar{b}_{ij}$ that is common to all varieties from i sold in j, and an idiosyncratic shifter $b_{ij}(\omega)$ that is specific to a variety ω. The quantity that j demands of variety ω from origin i is

$$q_{ij}(\omega) = \left(\bar{b}_{ij} b_{ij}(\omega)\right) \left(\frac{p_{ij}(\omega)}{P_j}\right)^{-\sigma} \frac{E_j}{P_j}, \tag{1}$$

where σ is the elasticity of substitution, E_j is j's total spending, $p_{ij}(\omega)$ is the price of variety ω of i sold in j, and P_j is j's CES price index implicitly determined by j's budget constraint,

$$\sum_i \int_{\omega \in \Omega_{ij}} p_{ij}(\omega) q_{ij}(\omega) d\omega = E_j, \tag{2}$$

with Ω_{ij} the set of varieties of origin i available in j. This environment allows for variety-specific demand shifters. As shown by Eaton et al. (2011), heterogeneous taste shifters help to rationalize the heterogeneous decisions of firms to export to different destinations. Recently, Redding and Weinstein (2020, 2024) pointed out that heterogeneity in variety-specific demand shifters also plays an important role in determining variation in price indices and trade flows across countries and years.

Technology. A variety is produced by a single firm, so we refer to a variety as a firm. Production is subject to variable and fixed labor costs that are heterogeneous across firms. The cost of firm ω from i of selling q units in destination j is

$$C_{ij}(\omega, q) = w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} \frac{\tau_{ij}(\omega)}{a_i(\omega)} q + w_i \bar{f}_{ij} f_{ij}(\omega), \tag{3}$$

where w_i is the wage in origin i. The variable cost of selling q units in j includes both firm-specific iceberg shipping costs, $\bar{\tau}_{ij}\tau_{ij}(\omega)$, and productivity, $\bar{a}_i a_i(\omega)$. The second term, $w_i \bar{f}_{ij} f_{ij}(\omega)$, is the fixed labor cost necessary for firm ω from i to access consumers in j. Following Eaton et al. (2011), firms can differ both in their productivity and fixed costs. We further introduce heterogeneity in the variable cost of serving different destinations. This allows the model to flexibly replicate various patterns of firm-level exports across destinations.

Entry and revenue potentials. We now define the two variables that summarize the sources of firm heterogeneity determining export decisions.

Under monopolistic competition, the firm's profit maximization problem implies that its optimal price is $p_{ij}(\omega) = \frac{\sigma}{\sigma-1} \frac{\bar{\tau}_{ij}}{\bar{a}_i} \frac{\tau_{ij}(\omega)}{a_i(\omega)} w_i$ with an associated revenue of

$$R_{ij}(\omega) = (w_i^{1-\sigma} P_j^{\sigma-1} E_j) \bar{r}_{ij} r_{ij}(\omega) \tag{4}$$

where

$$r_{ij}(\omega) \equiv b_{ij}(\omega)\left(\frac{\tau_{ij}(\omega)}{a_i(\omega)}\right)^{1-\sigma} \quad \text{and} \quad \bar{r}_{ij} \equiv \bar{b}_{ij}\left(\frac{\sigma}{\sigma-1}\frac{\bar{\tau}_{ij}}{\bar{a}_i}\right)^{1-\sigma}. \tag{5}$$

We refer to $r_{ij}(\omega)$ as the *revenue potential* in j of firm ω from i and to $\bar{r}_{ij}$ as the revenue shifter in j that is common to all firms from i. Conditional on entering market j, $r_{ij}(\omega)$ is the ω-specific revenue shifter that combines different sources of firm heterogeneity.

Firm ω from i sells in j if its variable profit exceeds its fixed cost. Given the profits implied by CES demand, this is equivalent to $\pi_{ij}(\omega) = \frac{1}{\sigma}R_{ij}(\omega) - w_i\bar{f}_{ij}f_{ij}(\omega) \geq 0$, which yields the set of firms from i selling in j, Ω_{ij}:

$$\Omega_{ij} = \{\omega : e_{ij}(\omega) \geq e^*_{ij}\} \tag{6}$$

where

$$e_{ij}(\omega) \equiv \frac{r_{ij}(\omega)}{f_{ij}(\omega)} \quad \text{and} \quad e^*_{ij} \equiv \frac{\sigma\bar{f}_{ij}}{\bar{r}_{ij}}\left[\left(\frac{w_i}{P_j}\right)^{\sigma}\frac{P_j}{E_j}\right]. \tag{7}$$

We refer to $e_{ij}(\omega)$, the ratio between the firm's revenue potential and its fixed cost, as the *entry potential* of firm ω from i in j. Among firms with identical revenue potential, heterogeneity in fixed costs generates heterogeneity in entry potentials and, therefore, in entry decisions across destinations.

2.2 The Extensive and Intensive Margins of Bilateral Trade Flows

We now define the two functions that control the extensive and intensive margins of firm exports and, thus, bilateral trade flows. To do so, define the share of firms from i selling in j and their average sales as

$$n_{ij} \equiv \Pr[\omega \in \Omega_{ij}] \quad \text{and} \quad \bar{x}_{ij} \equiv \mathbb{E}[R_{ij}(\omega)|\omega \in \Omega_{ij}]. \tag{8}$$

Throughout the rest of the paper, we refer to an origin-destination pair as a market, and to n_{ij} and $\bar{x}_{ij}$ as the exporter firm share and the average firm exports in a market, respectively.

For each market, consider the distribution of $(r_{ij}(\omega), e_{ij}(\omega))$ generated by the underlying joint distribution of firm fundamentals, $\{a_i(\omega), b_{ij}(\omega), \tau_{ij}(\omega), f_{ij}(\omega)\}$. Without loss of generality, we assume that

$$r_{ij}(\omega) \sim G^r_{ij}(r|e) \qquad \text{and} \qquad e_{ij}(\omega) \sim G^e_{ij}(e). \tag{9}$$

This allows for any pattern of heterogeneity and correlation in revenue and entry potentials, $(r_{ij}(\omega), e_{ij}(\omega))$, both within and between markets. Accordingly, it departs from the literature that explicitly imposes functional form assumptions on the distribution of firm fundamentals. Our general formulation encompasses several distributional assumptions in the literature. For instance, in Melitz (2003), the only source of firm heterogeneity is productivity such that $r_{ij}(\omega) = e_{ij}(\omega) = (a_i(\omega))^{\sigma-1}$. In this special case, the distribution of $a_i(\omega)$ can be specified to be Pareto, as in Chaney (2008) and Arkolakis (2010), truncated Pareto, as in Helpman et al. (2008) and Melitz and Redding (2015), or log-normal, as in Head et al. (2014) and Bas et al. (2017). A single source of firm heterogeneity implies a strict hierarchy of entry across destinations and a perfect cross-firm correlation between the intensive and extensive margins of exports. To relax these implications, multiple papers incorporate additional sources of heterogeneity across firms. For example, Eaton et al. (2011) impose that the distribution of $a_i(\omega)$ is Pareto and of $(b_{ij}(\omega), f_{ij}(\omega))$ is log-normal, while Fernandes et al. (2023) assume a multivariate log-normal distribution of cost shifters.

We impose a regularity restriction on the distribution of entry potentials.

Assumption 1. $G^e_{ij}(e)$ *is continuous and strictly increasing on* $\mathbb{R}_+$*, with density* $g^e_{ij}(e)$*.*

This assumption implies that changes in trade costs induce a positive mass of firms to switch entry decisions, which is central for the invertibility argument used below.[7]

Extensive margin of firm exports. Given the definition in equation (8), the entry decision in equation (6) implies that the share of firms from i selling in j is equal to $n_{ij} = \Pr[e_{ij}(\omega) > e^*_{ij}] = 1 - G^e_{ij}(e^*_{ij})$. To connect the exporter firm share and determinants of entry, we define the entry potential function, $\epsilon_{ij}(n) \equiv (G^e_{ij})^{-1}(1-n)$. This function characterizes the level of $e_{ij}(\omega)$ for the firm ω that is below a fraction n of all firms ordered by their entry potentials.

Under Assumption 1, the entry potential function $\epsilon_{ij}(n)$ is strictly decreasing, with $\epsilon_{ij}(1) = 0$ and $\lim_{n\to 0} \epsilon_{ij}(n) = \infty$. These properties allow us to obtain the following equilibrium relationship between the exporter firm share, bilateral exogenous variables, and country-level endogenous variables:

$$\ln \epsilon_{ij}(n_{ij}) = \ln\left(\sigma \bar{f}_{ij}/\bar{r}_{ij}\right) + \ln\left(w_i^\sigma\right) - \ln\left(P_j^{\sigma-1} E_j\right). \tag{10}$$

Equation (10) determines the extensive margin elasticity, $\theta^e_{ij}(n_{ij}) \equiv \partial \ln n_{ij}/\partial \ln \bar{r}_{ij}$. This is the elasticity of the exporter firm share with respect to the bilateral shifter $\bar{r}_{ij}$, holding constant endogenous origin- and destination-level variables. It is a univariate function of the initial exporter firm share n_{ij}, $\theta^e_{ij}(n_{ij}) = -\left(\frac{\partial \ln \epsilon_{ij}(n)}{\partial \ln n}|_{n=n_{ij}}\right)^{-1}$. The exporter firm share

[7] The assumption also rules out bounds on the support of e. This simplifies our derivations, but is not essential, as discussed in Section 2.4.

defines the marginal firms whose export decisions respond to small changes in revenue shifters and, consequently, the strength of responses along the extensive margin of firm exports.

Given the definition of the entry potential function $\epsilon_{ij}(n)$, we can relate the extensive margin elasticity to the entry potential function and its hazard rate:

$$\theta^e_{ij}(n_{ij}) = \epsilon_{ij}(n_{ij}) h^e_{ij}(\epsilon_{ij}(n_{ij})). \tag{11}$$

In particular, $\theta^e_{ij}(n)$ is proportional to the hazard rate of entry potentials at the initial equilibrium, $h^e_{ij}(\epsilon_{ij}(n_{ij})) \equiv g^e_{ij}(\epsilon_{ij}(n_{ij}))/(1 - G^e_{ij}(\epsilon_{ij}(n_{ij})))$, which is the ratio of the mass of marginal firms, $g^e_{ij}(\epsilon_{ij}(n_{ij}))$, to the mass of inframarginal firms, $n_{ij} = 1 - G^e_{ij}(\epsilon_{ij}(n_{ij}))$. Accordingly, the sensitivity of $\theta^e_{ij}(n)$ with respect to the exporter firm share depends on the elasticity of the hazard rate: $\partial\theta^e_{ij}(n)/\partial \ln n = -\partial \ln h^e_{ij}(\epsilon_{ij}(n))/\partial \ln e - 1$. Whenever n is higher, the extensive margin becomes more sensitive if the hazard rate of entry potential decreases faster that it would under the Pareto distribution (for which $\partial \ln h^e_{ij}(\epsilon_{ij}(n))/\partial \ln e = -1$ and $\partial\theta^e_{ij}(n)/\partial \ln n = 0$).

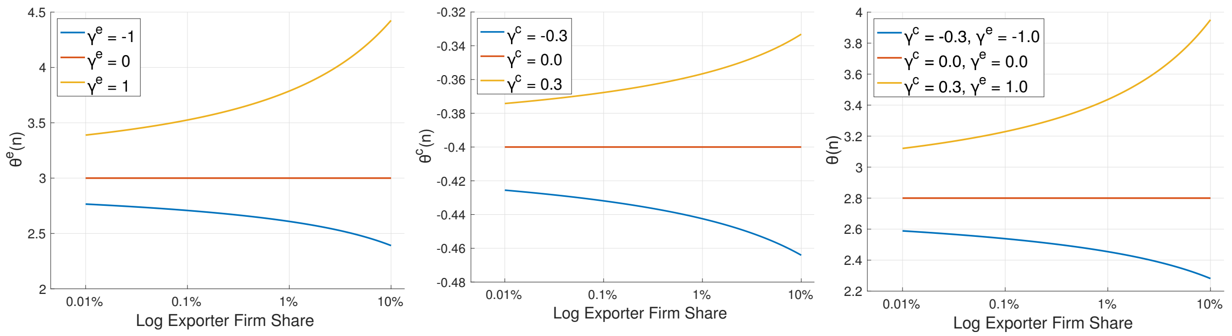
To illustrate further those points, we consider a modified Pareto distribution that allows for different decay rates over the support: $G^e(e) = 1 - (e/\underline{e})^{-\alpha^e}(\ln e/\ln \underline{e})^{-\gamma^e}$ for $e > \underline{e} > 1$, and $\alpha^e > \max\{0, -\gamma^e/\ln \underline{e}\}$, which implies that $\theta^e(n) = \alpha^e + \gamma^e/\ln \epsilon(n)$. The parameter γ^e controls the elasticity of the hazard rate and the slope of $\theta^e(n)$. Panel (a) of Figure 1 illustrates the extensive margin elasticity for different values of γ^e. For $\gamma^e = 0$, this example reduces to a Pareto distribution with a constant extensive margin elasticity, $\theta^e(n) = \alpha^e$. When $\gamma^e > 0$ in contrast, the hazard rate decreases at a faster rate compared to that of a Pareto distribution and, consequently, the extensive margin elasticity is lower in markets with few exporters (i.e., in which high-entry potential firms are marginal). The opposite happens when $\gamma^e < 0$: the hazard rate decreases more slowly, leading to a higher extensive margin elasticity in markets with few exporters. These patterns reflect the shape of the density of firm entry potentials: the smaller the γ^e, the larger the relative mass of high- to low-entry potential firms (see panel (a) of Appendix Figure OA.1).

Several popular distribution families yield an extensive margin elasticity that decreases with the exporter firm share, including log-normal, exponential, gamma, and Weibull.[8] In addition, $\theta^e_{ij}(n)$ is decreasing for specifications in which the log of firm fundamentals have a joint normal distribution, as in Bas et al. (2017) and Fernandes et al. (2023), since the distribution of entry potentials in each market is also log-normal.

[8]For examples, see Appendix Figure OA.2. When $\theta^e_{ij}(n)$ is decreasing, the Zero Profit Cutoff function in Melitz (2003) is also decreasing. His footnote 15 discusses that this property holds for many common distribution families.

Figure 1: Distributional Assumptions and the Elasticity of Firm Export Margins

(a) Extensive Margin Elasticity (b) Firm Composition Elasticity (c) Bilateral Trade Elasticity



Note. Panel (a) reports the extensive margin elasticity, $\theta^e_{ij}(n)$ defined in (11), Panel (b) reports the firm composition elasticity, $\theta^c_{ij}(n)$ defined in (13), and panel (c) reports trade elasticity, $\theta_{ij}(n)$ defined in (14). We report the elasticity functions obtained when the entry potential distribution is a modified Pareto function, $G^e(e) = 1 - (e/\underline{e})^{-\alpha^e}(\ln e/\ln \underline{e})^{-\gamma^e}$ with $e \geq \underline{e} > 1$, $\alpha^e = 3$ and $\underline{e} = \exp(1)$, and the conditional mean revenue potential is a modified power function, $\mathbb{E}[r|e = \epsilon_{ij}(n)] = n^{-\alpha^c}(1 - \ln(n))^{-\gamma^c}$ with $\alpha^c = 0.4$.

Intensive margin of firm exports. Given the definition in equation (8), the revenue expression in equation (4) and the entry decision in equation (6) imply that the average firm exports are $\bar{x}_{ij} = (w_i^{1-\sigma} P_j^{\sigma-1} E_j)\bar{r}_{ij}\mathbb{E}[r_{ij}(\omega)|e_{ij}(\omega) > e^*_{ij}]$. This depends on the mean revenue potential of the firms self-selecting into the market, $\mathbb{E}[r_{ij}(\omega)|e_{ij}(\omega) > e^*_{ij}]$. To characterize this term, we define the revenue potential function, $\rho_{ij}(n) \equiv \frac{1}{n}\int_0^n \mathbb{E}[r|e = \epsilon_{ij}(n')]\, dn'$. This function measures the mean revenue potential of the set of firms with the highest $n\%$ entry potentials in the market (i.e., those with $e_{ij}(\omega) > \epsilon_{ij}(n)$), where $\mathbb{E}[r|e = \epsilon_{ij}(n)]$ denotes the mean revenue potential of firms with the n-highest entry potential.

In equilibrium, the set of active firms in a market includes the fraction n_{ij} of firms with highest entry potentials (i.e., those with $e_{ij}(\omega) > e^*_{ij} = \epsilon_{ij}(n_{ij})$). The definition of $\rho_{ij}(n)$ yields the following relationship between average firm exports, firm composition, bilateral exogenous variables, and endogenous country variables:

$$\ln \bar{x}_{ij} - \ln \rho_{ij}(n_{ij}) = \ln(\bar{r}_{ij}) + \ln\left(w_i^{1-\sigma}\right) + \ln\left(P_j^{\sigma-1}E_j\right). \tag{12}$$

Equation (12) determines the intensive margin elasticity, $\theta^i_{ij}(n_{ij}) \equiv \partial \ln \bar{x}_{ij}/\partial \ln \bar{r}_{ij}$. This is the elasticity of average firm exports with respect to the bilateral shifter $\bar{r}_{ij}$, holding constant endogenous origin- and destination-level variables. It is a univariate function of the exporter firm share, since $\theta^i_{ij}(n_{ij}) = 1 + \theta^c_{ij}(n_{ij})\theta^e_{ij}(n_{ij})$ where we define the firm composition elasticity as $\theta^c_{ij}(n_{ij}) \equiv \frac{\partial \ln \rho_{ij}(n)}{\partial \ln n}|_{n=n_{ij}}$. The exporter firm share determines how changes in export decisions of marginal firms affect the composition of active firms in a market, $\theta^c_{ij}(n_{ij})$.

Given the definition of the revenue potential function $\rho_{ij}(n)$, we can relate the firm composition elasticity, $\theta^c_{ij}(n_{ij})$, to the difference in revenue potential of marginal and inframarginal firms:

$$\theta^c_{ij}(n_{ij}) = \frac{\mathbb{E}[r|e = \epsilon_{ij}(n_{ij})]}{\rho_{ij}(n)} - 1 = \frac{\mathbb{E}[r|e = \epsilon_{ij}(n_{ij})]}{\frac{1}{n_{ij}} \int_0^{n_{ij}} \mathbb{E}[r|e = \epsilon_{ij}(n)]\, dn} - 1. \tag{13}$$

The term $\theta^c_{ij}(n_{ij})$ measures exporter composition changes around the initial equilibrium, as captured by the ratio between the mean revenue potential of marginal exporters, $\mathbb{E}[r|e = \epsilon_{ij}(n_{ij})]$, and inframarginal exporters, $\rho_{ij}(n_{ij})$. When $\mathbb{E}[r|e = \epsilon_{ij}(n_{ij})]$ is smaller than $\rho_{ij}(n)$, the firm composition elasticity is negative. This is the case in the Melitz model with a single source of firm heterogeneity (i.e., $r_{ij}(\omega) = e_{ij}(\omega)$ and $\mathbb{E}[r|e = \epsilon_{ij}(n)] = \epsilon_{ij}(n)$, which always decreases with n). A negative firm composition elasticity is consistent with evidence that firms with higher revenue potential self-select into exporting (Melitz and Redding, 2014). Note however that the composition elasticity can vary with the exporter firm share: $\partial\theta^c_{ij}(n)/\partial \ln n = (1 + \theta^c_{ij}(n))(\partial \ln \mathbb{E}[r|e = \epsilon_{ij}(n)]/\partial \ln n - \theta^c_{ij}(n))$. Whenever n is higher, the composition elasticity is also higher if the decrease in the revenue potential of marginal firms is relatively weak, $\partial \ln \mathbb{E}[r|e = \epsilon_{ij}(n)]/\partial \ln n > \theta^c_{ij}(n)$.

To illustrate this discussion, we consider a modified power function: $\mathbb{E}[r|e = \epsilon_{ij}(n)] = n^{-\alpha^c}(1 - \ln(n))^{-\gamma^c}$ with $\alpha^c \in (0, 1)$ and $\gamma^c < \alpha^c$. Here, marginal firms have lower mean potential than inframarginal firms, since $\mathbb{E}[r|e = \epsilon_{ij}(n)]$ decreases with n and, thus, $\theta^c_{ij}(n) \in (-1, 0)$. The parameter γ^c controls the pace at which the mean revenue of marginal firms decreases: $\partial \ln \mathbb{E}[r|e = \epsilon_{ij}(n)]/\partial \ln n = -\alpha^c + \gamma^c/(1 - \ln(n))$. Panel (b) of Figure OA.1 illustrates the composition elasticity for three cases. For $\gamma = 0$, the composition term has a constant elasticity, $\theta^c_{ij}(n) = -\alpha^c$. A positive γ^c slows the decay of the mean revenue of marginal firms as n increases, so that $\theta^c_{ij}(n)$ increases with n. In contrast, a negative γ^c implies a faster decay and a decreasing $\theta^c_{ij}(n)$. The lower the value of γ^c, the faster the mean revenue potential of marginal firms decays as n increases (see Panel (b) of Appendix Figure OA.1).

We again note that, with a single source of firm heterogeneity, $\theta^c_{ij}(n)$ is decreasing for several popular parametric distribution families, like log-normal and truncated Pareto – see Panel (b) of Figure OA.2.

Bilateral trade flows. Next, we construct bilateral trade flows between countries as $X_{ij} \equiv N_i n_{ij} \bar{x}_{ij}$, with N_i the mass of firms in origin i. The trade elasticity, defined as the elasticity of bilateral trade flows with respect to bilateral revenue shifters holding constant other endogenous origin- and destination-level variables, $\theta_{ij}(n_{ij}) \equiv \partial \ln n_{ij}/\partial \ln \bar{r}_{ij} + \partial \ln \bar{x}_{ij}/\partial \ln \bar{r}_{ij}$, is:

$$\theta_{ij}(n_{ij}) = 1 + (\theta^c_{ij}(n_{ij}) + 1)\theta^e_{ij}(n_{ij}), \tag{14}$$

with $\theta_{ij}(n) > 1$ (as $\theta^c_{ij}(n) > -1$ and $\theta^e_{ij}(n) > 0$).

The trade elasticity is a function of the exporter firm share in each market, since it inherits the properties of the extensive and intensive elasticity functions through $\theta_{ij}^{e}(n)$ and $\theta_{ij}^{c}(n)$. The extensive and intensive margin elasticities regulate the entry of marginal exporters into a market and the differences between these marginal entrants and inframarginal firms, respectively. Consequently, they determine the responsiveness of bilateral trade flows to changes in trade costs.

To illustrate this point, Panel (c) of Figure OA.1 displays the trade elasticity function implied by the combination of the two examples above. The trade elasticity can be invariant to the exporter firm share, as in the benchmark of a constant-elasticity gravity model. Alternatively, depending on the decay rate of the entry and revenue potential functions, the trade elasticity may increase or decrease with the exporter firm share. Importantly, as discussed above, several popular distributional assumptions can only generate a decreasing trade elasticity (see Appendix Figure OA.2).

The following proposition summarizes these results.

Proposition 1. *Consider the monopolistic competition model with CES demand of Section 2.1 under Assumption 1. Then:*
a. For each market, the exporter firm share, n_{ij}, and the average firm exports, $\bar{x}_{ij}$, are given by equations (10) and (12), which are separable on country-level endogenous variables, exogenous bilateral shifters, and two functions, $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$.
b. Given country-level endogenous variables, $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ summarize the role of firm heterogeneity in the elasticity of bilateral trade flows to bilateral revenue shifters.

The proposition has a direct analogy to the inversion argument used to identify demand systems in Berry (1994) and Berry and Haile (2014), self-selection models in Adão (2015), and perfectly competitive trade models in Adao et al. (2017). Here, we leverage the structure of the monopolistic competition model to invert the equilibrium equations for the exporter firm share n_{ij} and the average firm exports $\bar{x}_{ij}$. These gravity-like expressions will be central for our strategy to estimate $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$. Since, as we noted earlier, these functions depend only on the exporter firm share, n_{ij}, they are relatively easy to estimate using cross-market variation in exporter firm shares and average firm exports. In contrast, in the settings cited above, the inversion produces elasticity functions that depend on vectors with dimensions that match the number of choices (such as markets or products).[9]

[9] For example, in Adao et al. (2017), the nonparametric gravity system depends on a destination-specific function whose dimension is equal to the number of factors in the world economy. We note that the univariate elasticity function emerges from the separability of export decisions across markets in our model. Such a separability does not hold if the firm's profitability in a destination depends on its decision to operate in other destinations, as in Tintelnot (2017) and Morales et al. (2019). In these cases, we have to invert the joint decision to export to all destinations, which increases the dimensionality of the elasticity functions.

2.3 Sufficient Statistics of Firm Heterogeneity in General Equilibrium

We next show that $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ are sufficient statistics for how export decisions of heterogeneous firms affect the aggregate variables in general equilibrium. To do so, we specify conditions for free entry, budget balance, and labor market clearing.

Firms incur a fixed labor cost, $\bar{F}_i$, to draw their fundamentals. With free entry, the equilibrium mass of firms in country i, N_i, expects to make zero profit:

$$\sum_j \mathbb{E}[\max\{\pi_{ij}(\omega), 0\}] = \sum_j n_{ij}(\bar{x}_{ij} - \bar{c}_{ij}) = w_i \bar{F}_i, \tag{15}$$

where $\bar{c}_{ij} \equiv \mathbb{E}[C_{ij}(\omega)|\omega \in \Omega_{ij}]$ is the sum of the mean variable and fixed costs of firms from i selling in j. With CES demand, $\bar{c}_{ij}$ is given by

$$\bar{c}_{ij} = (1 - 1/\sigma)\mathbb{E}[R_{ij}(\omega)|\omega \in \Omega_{ij}] + w_i \bar{f}_{ij}\mathbb{E}[f_{ij}(\omega)|\omega \in \Omega_{ij}], \tag{16}$$

and the free entry condition in (15) can be written in terms of $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$:[10]

$$\frac{1}{\sigma}\sum_j n_{ij}\bar{x}_{ij} = w_i \bar{F}_i + w_i \sum_j \bar{f}_{ij} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n)}(1 + \theta^c_{ij}(n))dn. \tag{17}$$

As argued above, $\bar{x}_{ij}$ and n_{ij} can also be written in terms of $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$.

To derive the budget constraint, we follow Dekle et al. (2008) and allow for exogenous international transfers $\{\bar{T}_i\}$ with $\sum_i \bar{T}_i = 0$. Total spending equals labor income and transfers in each country j, so that the budget constraint in (2) is equivalent to

$$\sum_i N_i n_{ij}\bar{x}_{ij} = w_j \bar{L}_j + \bar{T}_j = E_j. \tag{18}$$

Since labor is the only factor of production, labor income in i equals the total revenue of firms from i:

$$\sum_j N_i n_{ij}\bar{x}_{ij} = w_i \bar{L}_i. \tag{19}$$

Since (18)-(19) only depend on $\bar{x}_{ij}$ and n_{ij}, they can also be written as a function of $\rho_{ij}(n)$ and $\epsilon_{ij}(n)$. We can then state the following proposition.

Proposition 2. *Consider the monopolistic competition model with CES demand of Section 2.1 under Assumption 1. Assume knowledge of the exogenous fundamentals* $\{\bar{r}_{ij}, \bar{f}_{ij}, \bar{L}_i, \bar{T}_i, \bar{F}_i\}$,

[10] To see this, note that $\mathbb{E}[f_{ij}(\omega)|\omega \in \Omega_{ij}] = \mathbb{E}[r_{ij}(\omega)/e_{ij}(\omega)|e_{ij}(\omega) > e^*_{ij}] = (1/n_{ij})\int_{e^*_{ij}}^{\infty}(1/e)\mathbb{E}[r|e]dG^e_{ij}(e) = (1/n_{ij})\int_0^{n_{ij}} \mathbb{E}[r|e = \epsilon_{ij}(n)]/\epsilon_{ij}(n)dn$, and that $\mathbb{E}[r|e = \epsilon_{ij}(n)] = \rho_{ij}(n)(1 + \theta^c_{ij}(n))$.

the elasticity of substitution σ, and the functions $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$. Then:

a. The equilibrium vector $\{P_i, N_i, w_i\}$ solves the system of equations (17)-(19) with n_{ij} and $\bar{x}_{ij}$ given by (10) and (12).

b. The equilibrium is Pareto efficient.

The main implication of Proposition 2 is that the distribution of firm fundamentals affects the economy's equilibrium only insofar it determines the shape of the functions $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$. In other words, conditional on these functions, equilibrium outcomes do not depend on other components of the joint distribution of firm fundamentals – for example, the cross-destination correlation in preference and trade costs, or the dispersion of firm revenue potentials given entry potentials. In the next section, we further show that $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ yield sufficient statistics for the welfare consequences of changes in trade costs.

We prove the second part of the proposition in Appendix A.2.1. It generalizes the equilibrium efficiency results of Dhingra and Morrow (2012) and Zhelobodko et al. (2011) to multiple countries and multiple sources of firm heterogeneity. While intuitive, our result is not trivial. Whereas relative quantities are efficient under CES preferences, endogenous entry and selection decisions of firms could be potentially distorted due to cross-market variation in the firm-level distribution of profit margins. Nevertheless, we find that these decisions are also efficient since CES demand implies that the profit share of all firms are identical and invariant of market conditions.

Distribution of Firm Exports and Aggregate Outcomes. We established that aggregate outcomes depend on two components of firm heterogeneity: (i) the distribution of entry potentials (i.e., G^e_{ij}, which determines $\epsilon_{ij}(n)$ and $\theta^e_{ij}(n)$) and (ii) the mean revenue potential conditional on the firm's entry potential (i.e., $\mathbb{E}[r|e]$, which determines $\rho_{ij}(n)$ and $\theta^c_{ij}(n)$). An extensive literature builds on Melitz (2003) to propose quantitative frameworks that match the distribution of firm export outcomes across destinations (see Melitz and Redding (2014), for a review). In light of this literature, it is natural to ask whether fitting this distribution is sufficient to recover the aggregate implications of firm export decisions.

Given Proposition 2, it is straightforward to see that the distribution of firm exports, conditional on $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$, plays no role in the model's aggregate predictions. The distribution of firm log-revenue in a market, $G^{\ln R}_{ij}(x) \equiv \Pr[\ln R_{ij}(\omega) < x | \omega \in \Omega_{ij}]$, is

$$G^{\ln R}_{ij}(x) = n_{ij}^{-1} \int_0^{n_{ij}} G^r_{ij}(e^x / \bar{R}_{ij} | e = \epsilon_{ij}(n)) dn, \tag{20}$$

with $\bar{R}_{ij} \equiv (w_i^{1-\sigma} P_j^{\sigma-1} E_j)\bar{r}_{ij}$. Note that $G^{\ln R}_{ij}(x)$ depends on the entire shape of the conditional distribution of revenue potentials $G^r_{ij}(r|e)$, not only on its mean $\mathbb{E}[r|e]$. Thus, one can construct two distinct economies with different $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ that nevertheless generate the same distribution of firm exports. Conversely, one can rationalize different distributions of firm exports, while maintaining the same $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$, and the same aggregate outcomes.[11] Intuitively, for different levels of bilateral trade costs, what matters for aggregate outcomes is the response of the extensive and intensive margins of firm exports, and not the response of higher moments of the distribution of firm sales. We will return to this discussion in Section 4.

2.4 Extensions

We next discuss extensions of Propositions 1 and 2.a. All derivations are in Appendix A.3.

Non-CES demand. Our first extension allows for variable markups by specifying a general single-aggregator demand, be it homothetic as in Matsuyama and Ushchev (2017) or non-homothetic as in Arkolakis et al. (2019). We extend our inversion argument to establish that the trade elasticity remains a function of the exporter firm share, as implied by the two gravity equations for the firm export margins. In the absence of fixed cost heterogeneity, the trade elasticity functions are sufficient to characterize the economy's equilibrium. However, when firms are heterogeneous in their fixed costs, the characterization of profit margins in the free entry condition requires additional elasticity functions that govern how revenue shifters affect percentiles of the distribution of firm exports; each elasticity is a univariate function of the exporter firm share.

Multiple sectors, multiple factors, input-output links, and import tariffs. Our second extension includes features common to quantitative trade models such as multiple factors of production, input-output links between multiple sectors, and import tariffs. Specifically, we extend the multi-sector, multi-factor gravity model of Costinot and Rodriguez-Clare (2013) to allow firms in each sector to be heterogeneous with respect to shifters of productivity, preferences, and variable and fixed trade costs. We restrict all firms in a sector to have the same nested CES production technology that uses multiple factors and multiple sectoral composite goods.[12] In this setting, we derive sector-specific analogs of (10) and (12) that

[11] As an extreme example, consider an economy in which revenue and entry potentials are independent—for instance, if fixed and variable costs vary proportionally across firms. In that case, $G^r_{ij}(r|e) = G^r_{ij}(r)$ and $G^{\ln R}_{ij}(x) = G^r_{ij}(e^x/\bar{R}_{ij})$, which implies that the distribution of firm exports does not depend on the trade elasticity functions.

[12] To simplify exposition, our derivations rely on nested CES preferences and technology. Note however that it is straightforward to extend our results to a more general structure of separable preferences and technology over sectoral composite goods while maintaining the assumption of CES preferences across varieties within each sector. We can use the alternative environment of Appendix Section A.3.1 to further relax the

determine aggregate variables in general equilibrium when combined with knowledge of the components of the production function that are common to all firms in each sector.

Allowing for zero trade flows. Next, we extend our model to allow for zero bilateral trade flows, as in Helpman et al. (2008). To do so, we consider a weaker version of Assumption 1 in which the support of the entry potential distribution is bounded: $G^e_{ij}(e)$ has full support over $[0, \bar{e}_{ij}]$. This does not affect the intensive margin equation in (12), but it introduces a censoring structure into the extensive margin equation in (10).

Multi-product firms. We finally extend our model to allow heterogeneous firms to produce multiple products, as in Bernard et al. (2011). We assume that firms face a convex labor cost of increasing the number of varieties supplied in each destination (see e.g. Arkolakis et al. (2021)). In this setting, the expressions for the extensive and intensive margins of firm exports are still given by (10) and (12), but Proposition 2.a also requires knowledge of the elasticity function controlling how the number of products per exporting firm responds to trade costs, which arises in an additional gravity equation for the extensive margin of products per exporter.

3 Nonparametric Counterfactual Analysis: The Aggregate Implications of Firm Export Decisions

This section establishes that $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ summarize how the export decisions of heterogeneous firms affect aggregate responses to changes in trade costs. We also provide expressions for welfare changes in terms of the trade elasticity functions to show how responses depend on the adjustment margins in our model.

3.1 Counterfactual Responses to Changes in Bilateral Trade Costs

We consider how the economy responds to counterfactual changes in revenue shifters $\{\bar{r}_{ij}\}$. For any variable y, we use y^0 to denote its value at the initial equilibrium, and $\hat{y} \equiv y'/y^0$ and $d \ln y$ to denote respectively its ratio and first-order log-change between the initial and counterfactual equilibria. Appendix A.2.2 establishes the requirements for computing the counterfactual responses of aggregate outcomes.

assumption of CES preferences within each sector.

Proposition 3. *Consider a counterfactual change in bilateral revenue shifters* $\{\bar{r}_{ij}\}$ *in the monopolistic competition model with CES demand of Section 2.1 under Assumption 1. Assume knowledge of the elasticity of substitution* σ*, and the bilateral trade matrix at the initial equilibrium* $\{X_{ij}^0\}$*. Then, we can compute counterfactual responses in aggregate outcomes* $\{X_{ij}, P_i, N_i, w_i\}$ *with knowledge of:*
a. for small shocks, the trade elasticity matrix at the initial equilibrium $\{\theta_{ij}(n_{ij}^0)\}$*;*
b. for large shocks, the functions $\epsilon_{ij}(n)$ *and* $\rho_{ij}(n)$*, and the exporter firm share matrix at the initial equilibrium* $\{n_{ij}^0\}$*.*

The first part of the proposition focuses on the (local) response of aggregate outcomes to small shocks in bilateral revenue shifters. It establishes that such responses are a function of the demand elasticity of substitution σ, as well as two matrices evaluated at the initial equilibrium, the matrix of bilateral trade flows $\{X_{ij}^0\}$ and its associated elasticity matrix $\{\theta_{ij}(n_{ij}^0)\}$. In this case, separate knowledge of the extensive and intensive margin elasticities –and the distribution of firm fundamentals– is not required conditional on knowing the elasticity matrix of bilateral trade flows. In other words, changes in export decisions of heterogeneous firms only affect aggregate responses to small shocks through the heterogeneous trade elasticities, $\theta_{ij}^0 = \theta_{ij}(n_{ij}^0)$.[13]

The result hinges on two key observations. First, by definition, the local response of trade flows combines local responses of n_{ij} and $\bar{x}_{ij}$, as measured by the trade elasticity θ_{ij}^0 at the initial equilibrium. Thus, what remains to show is that, in changes, equilibrium conditions can be written as a function of bilateral trade flows, or the aggregate outcomes that we solve for. Indeed, we argued in Section 2.3 that budget balance and labor market clearing, (18) and (19), can be expressed in terms of bilateral trade flows and aggregate variables. Additionally, as shown in Appendix A.2.2, the free entry condition in (17) links firm entry in country i, $d\ln N_i$, and firm selection into different markets, $d\ln \epsilon_{ij}(n_{ij})$,

$$d\ln N_i = \sum_j y_{ij}^0 \theta_{ij}^0 d\ln \epsilon_{ij}(n_{ij}) \tag{21}$$

with $y_{ij}^0 \equiv X_{ij}^0 / \sum_{j'} X_{ij'}^0$, which establishes the result immediately from equation (10).

Expression (21) reflects a key mechanism. If the shock makes all markets more attractive to firms (i.e., $d\ln \epsilon_{ij}(n_{ij}) < 0$ and $d\ln n_{ij} > 0$ for all j), then the economy must spend more resources in fixed entry costs and, because of free entry, experience a reduction in the mass of

[13] In Appendix A.2.2, we show that the same requirements are sufficient to compute responses of aggregate outcomes to small changes in population $\bar{L}_i$ and transfers $\bar{T}_i$. However, to compute responses to small changes in the fixed costs of exporting $\bar{f}_{ij}$ and entry $\bar{F}_i$, we need to know also the initial share of the country's labor force employed to cover fixed costs of exporting, which can be recovered using the functions $(\epsilon_{ij}(n), \rho_{ij}(n))$ and the initial matrix of exporter firm shares $\{n_{ij}^0\}$.

new firms (i.e., $d\ln N_i < 0$). The trade elasticity θ^0_{ij} simply controls how much firm selection changes translate into profitability changes when markups are constant.

Combined with the economy's resource constraint, expression (21) links selection into exporting and domestic firm entry: $d\ln\epsilon_{ii}(n_{ii}) = -\sum_{j\neq i}(y^0_{ij}/y^0_{ii})d\ln\epsilon_{ij}(n_{ij})$ and $d\ln N_i = \sum_{j\neq i}(\theta^0_{ij}-\theta^0_{ii})y^0_{ij}d\ln\epsilon_{ij}(n_{ij})$. For a reduction in the cost of exporting that induces firm selection into all foreign markets (i.e., $d\ln\epsilon_{ij}(n_{ij}) < 0$ and $d\ln n_{ij} > 0$ for $i \neq j$), the economy's resource constraint forces domestic firms to exit (i.e., $d\ln\epsilon_{ii}(n_{ii}) > 0$ and $d\ln n_{ii} < 0$). The effect of the exporter expansion on new firm entry depends on the shape of the trade elasticity function, since it summarizes cross-market variation in profit margins. When $\theta^0_{ij} > \theta^0_{ii}$ for all $j \neq i$, the mass of firms in origin i decreases with exporter expansion into all foreign markets, $d\ln N_i < 0$.[14] When $\theta^0_{ij} = \theta^0_{ii} = \bar{\theta}_i$ for all j, as in the class of constant-elasticity gravity trade models in Arkolakis et al. (2012), changes in resources used to export to different destinations mechanically compensate each other, shutting down firm entry, $d\ln N_i = 0$.

The second part of the proposition turns to the impact of large changes in bilateral revenue shifters. In this case, the trade elasticity matrix $\{\theta_{ij}(n_{ij})\}$ may endogenously change as the economy moves away from the initial equilibrium, and responses are shaped by a new set of marginal firms. One needs to track changes in $\theta_{ij}(n)$ induced by responses in n_{ij}, which requires the separate trade elasticity margins. Nonetheless, we do not need to know further details about the micro structure of the model, including the joint distribution of firm fundamentals and the initial matrix of exogenous fundamentals (e.g., $\bar{r}^0_{ij}$ or $\bar{f}^0_{ij}$). This part of the proposition is an application of the "hat-algebra" toolkit developed by Dekle et al. (2008), and a generalization of the sufficient statistics in Arkolakis et al. (2012) (Proposition 2) beyond the class of constant-elasticity gravity models.[15]

We can build more intuition for the connection between the two parts of the proposition using a constant-elasticity benchmark,

$$\theta^e_{ij}(n) = \bar{\theta}^e_{ij} \quad \text{and} \quad \theta^c_{ij}(n) = \bar{\theta}^c_{ij}. \tag{22}$$

This special case is a flexible extension of the Pareto variant of Melitz (2003) in Chaney (2008), with elasticities varying by origin and destination.[16] The first part of the proposition yields aggregate responses to large shocks based solely on knowledge of the trade elasticity matrix $\{\bar{\theta}_{ij}\}$, by integrating local responses without tracking changes in n_{ij}. When the trade

[14]We note that this condition holds if the trade elasticity decreases with the exporter firm share and only a small fraction of domestic firms export, $n^0_{ij} < n^0_{ii}$.

[15]As noted by Costinot and Rodriguez-Clare (2013), the "hat algebra" system for heterogeneous firm models also depends on the elasticity of substitution σ if the entry cost depends on the origin's wage.

[16]In Chaney (2008), $r_{ij}(\omega) = e_{ij}(\omega) = (a_i(\omega))^{\sigma-1}$ and $a_i(\omega) \sim 1 - a^{-\theta}$, leading to $\bar{\theta}^e_{ij} = -1/\bar{\theta}^c_{ij} = \theta/(\sigma-1)$.

elasticity functions are constant, the dispersion of firm entry and revenue potentials is also constant across the entire support. This leads to responses of firm export margins that are invariant to initial conditions.

3.2 The Margins of Welfare Responses to Changes in Bilateral Trade Costs

We now leverage the CES preferences in our model to characterize real wage responses to changes in trade costs. This is equivalent to welfare changes under trade balance (i.e., $\bar{T}_i = 0$). Equation (12) yields the change in the real wage of country j in terms of changes in exogenous and endogenous variables in any origin i:

$$\ln \frac{\hat{w}_j}{\hat{P}_j} = \frac{1}{\sigma - 1} \left(\ln \hat{\bar{r}}_{ij} + \ln \hat{N}_i - \ln \hat{x}_{ij} + \ln \hat{n}_{ij} \frac{\rho_{ij}(n^0_{ij} \hat{n}_{ij})}{\rho_{ij}(n^0_{ij})} \right) + \ln \frac{\hat{w}_j}{\hat{w}_i}, \tag{23}$$

with $x_{ij} \equiv X_{ij}/E_j$ the share of origin i in the expenditures of destination j.

To obtain a decomposition, we take the average of this expression weighted by initial trade shares, x^0_{ij}:

$$\begin{aligned} \ln \tfrac{\hat{w}_j}{\hat{P}_j} \;=\; & \underbrace{\sum_i \frac{x^0_{ij}}{\sigma - 1} \ln \hat{\bar{r}}_{ij}}_{\text{Technology}} + \underbrace{\sum_i x^0_{ij} \ln \frac{\hat{w}_j}{\hat{w}_i}}_{\text{Terms of trade}} - \underbrace{\frac{1}{\sigma - 1} \sum_i x^0_{ij} \ln \hat{x}_{ij}}_{\text{Demand substitution}} \\ + \; & \underbrace{\sum_i \frac{x^0_{ij}}{\sigma - 1} \ln \hat{N}_i}_{\text{Firm entry}} + \underbrace{\sum_i \frac{x^0_{ij}}{\sigma - 1} \ln \hat{n}_{ij} \frac{\rho_{ij}(n^0_{ij} \hat{n}_{ij})}{\rho_{ij}(n^0_{ij})}}_{\text{Firm selection}} . \end{aligned} \tag{24}$$

The first row measures the components of welfare responses that are present in neoclassical trade models. While the "technology" term captures the shock to the exogenous component of the cost of imported goods, the "terms of trade" term measures changes in the endogenous labor cost in origin i (relative to that of j).[17] These two channels are the "traditional" gains from trade in Hsieh et al. (2020), and capture the first-order impact of changes in trade costs on welfare in neoclassical models, such as the economy without wedges in Baqaee and Farhi (2019). In addition, the optimal adjustment of the consumption bundle creates an offsetting "demand substitution" effect. This component is approximately zero for small shocks (i.e., $\sum_i x^0_{ij} \ln \hat{x}_{ij} \approx \sum_i dx_{ij} = 0$), but it can be substantial for large shocks.[18]

[17]The technology term is scaled by $1/(\sigma - 1)$ because $\ln \hat{\bar{r}}_{ij}$ measures the demand shift associated with a cost shift (see the definition in (5)). To see this, consider shocks to iceberg trade costs for which $\ln \hat{\bar{r}}_{ij} = -(\sigma - 1) \ln \hat{\bar{\tau}}_{ij}$, and the technology term is $-\sum_i x^0_{ij} \ln \hat{\bar{\tau}}_{ij}$.

[18]The decomposition in Hsieh et al. (2020) does not have the demand substitution term, as Sato-Vartia weights cancel out substitution across origins.

The second row captures welfare changes stemming from endogenous firm entry in each origin i and the selection of heterogeneous firms from origin i into destination j. These terms measure welfare responses associated with the distinctive motives for trade in monopolistically competitive frameworks—love of variety, increasing returns to scale, and trade costs. They correspond to the indirect effect in Atkeson and Burstein (2010) and the "new" gains from trade in Hsieh et al. (2020), but written in terms of firm selection changes.

The contribution of "firm entry" in the first term is an average of the change in the mass of firms across origins, weighted by their initial expenditure shares. Because of CES preferences, the welfare value of these new varieties requires an adjustment by the parameter governing love for variety, $1/(\sigma - 1)$.

The second term measures the contribution of "firm selection" for welfare. It is also an average of the change in available varieties from different origins, weighted by expenditure shares and adjusted by love for variety. In addition, it takes into account the exporter composition change. That is, it incorporates the fact that the mean revenue potential of marginal entrants differs from that of inframarginal firms: $\ln \hat{n}_{ij} \frac{\rho_{ij}(n_{ij}^0 \hat{n}_{ij})}{\rho_{ij}(n_{ij}^0)} \approx (1+\theta_{ij}^c(n_{ij}^0)) d \ln n_{ij}$ with $\theta_{ij}^c(n_{ij}^0)$ capturing the composition elasticity defined in (13). A higher revenue potential of entrants compared to incumbents (i.e., a higher $\theta_{ij}^c(n_{ij}^0)$) generates a larger welfare gain from the selection of these marginal firms into destination j.

It is worth noting that the technology term only depends on exogenous shocks and initial spending shares and is invariant to the shape of the distribution of firm fundamentals. All other terms depend on the distribution of firm fundamentals. The firm terms, in particular, are generally nonzero when countries are asymmetric, since they directly affect welfare in each country j through variety availability.[19] However, Proposition 2.b shows that the equilibrium is efficient so that, up to a first-order, these terms are not important "on average" across countries. Indeed, only the technology term has a first-order impact on the global average real wage under trade balance:

$$\sum_j \frac{E_j^0}{E^0} d \ln \frac{w_j}{P_j} = \underbrace{\sum_j \sum_i \frac{X_{ij}^0}{E^0} \frac{d \ln \bar{r}_{ij}}{\sigma - 1}}_{\text{Global Technology Effect}}, \tag{25}$$

with $E^0 = \sum_j E_j^0 = \sum_{i,j} X_{ij}^0$. This formula for global welfare gains is closely related to the gains derived by Atkeson and Burstein (2010). Up to a first-order approximation, the

[19] One exception is the special case of symmetric countries studied in Atkeson and Burstein (2010) where the first-order impact of changes in trade costs on welfare is the technology term. Appendix A.2.2 shows that, when countries are symmetric, the link between firm entry and selection in (21) is identical across countries, which implies that the firm components of welfare exactly offset each other. In contrast, firm selection has a first-order impact on welfare when countries are asymmetric.

difference between welfare gains for country j in (24) and for the world in (25) can be interpreted as arising from between-country reallocation effects induced by responses in terms of trade, firm entry, and firm selection. At the global-level, these reallocation effects cancel each other when we use Negishi (1960) weights.[20]

Our welfare decomposition builds on the welfare-accounting tradition that dates back to Solow. Atkin and Donaldson (2022) highlight three complementary uses of such expressions. First, they clarify the theoretical mechanisms through which trade shocks affect welfare. Second, they enable ex ante decompositions of counterfactual predictions, identifying which mechanisms are quantitatively most important. Third, they provide a basis for ex post decompositions of observed welfare changes into their underlying margins. In our context, equation (24) serves these same purposes. It structures our theoretical analysis of how firm heterogeneity—summarized by the trade elasticity functions—affects welfare through changes in the terms of trade, entry, and selection. This expression guides the decomposition of welfare responses across countries in Section 6.

Welfare and Profits. Appendix A.2.4 derives the model's predictions for how trade cost changes affect the share of income accruing to profits from domestic and foreign sales. Even with constant markups, changes in firm composition create systematic movements in aggregate profits following trade shocks, which accrue to fixed costs paid in labor. Under free entry, changes in the mass of firms map directly to changes in the profit share of income, since $\pi_i \equiv N_i \sum_j \mathbb{E}[\max\{\pi_{ij}(\omega), 0\}]/w_i \bar{L}_i = N_i \bar{F}_i / \bar{L}_i$. As discussed in Section 3.1, the shape of the trade elasticity function determines whether the profit share rises or falls when exporter participation expands. Trade costs also affect the composition of profits between domestic and foreign markets. The response in the domestic profit share, $s^{\pi}_{ii} \equiv \mathbb{E}[\max\{\pi_{ii}(\omega), 0\}]/w_i \bar{F}_i$, is inversely related to welfare changes: $d \ln s^{\pi}_{ii} = -(\sigma - 1)\alpha^{\pi}_i d \ln w_i/P_i$ where $\alpha^{\pi}_i > 1$ denotes the ratio of variable profits to total profits (net of fixed costs) in the domestic market. Thus, trade shocks that reduce welfare lead firms to rely more heavily on domestic markets for their profits.[21]

[20] With any other set of weights, it is easy to show that these terms would affect global welfare through the impact of shock-induced transfers across countries. The presence of inefficiencies can also lead to additional reallocation terms in welfare. This has been discussed in a context of growth externalities by Perla et al. (2021), firm size wedges by Bai et al. (2024), variable markups by Arkolakis et al. (2019), or tariffs and exogenous markup wedges by Baqaee and Farhi (2019).

[21] For standard calibrations, $(\sigma - 1)\alpha^{\pi}_i > 1$. Hence, the profit share of exporter firms decreases disproportionately more than welfare in response to trade cost shocks. This prediction is qualitatively consistent with the large effects of increases in trade costs on firm profits and stock prices documented by Amiti et al. (2021).

3.3 The Gains From Trade

We now turn to a preeminent counterfactual exercise: the gains from trade defined as the impact on welfare of moving to autarky. In Appendix A.2.3, we characterize the gains from trade as a corollary of Proposition 3.

Corollary 1. *Consider a counterfactual change in trade costs that moves country j from the trade equilibrium to the autarky equilibrium: $\hat{\bar{r}}_{ij} \to 0$ for all $i \neq j$. Then,*

$$\ln \frac{\hat{w}_j}{P_j} = \frac{1}{\sigma - 1} \ln x_{jj}^0 + \frac{1}{\sigma - 1} \ln \hat{N}_j + \frac{1}{\sigma - 1} \ln \hat{n}_{jj} \frac{\rho_{jj}(n_{jj}^0 \hat{n}_{jj})}{\rho_{jj}(n_{jj}^0)} \tag{26}$$

where $\hat{n}_{jj}$ and $\hat{N}_j$ are given by

$$\frac{\epsilon_{jj}(n_{jj}^0 \hat{n}_{jj})}{\epsilon_{jj}(n_{jj}^0)} = x_{jj}^0 \hat{N}_j \hat{n}_{jj} \frac{\rho_{jj}(n_{jj}^0 \hat{n}_{jj})}{\rho_{jj}(n_{jj}^0)} \tag{27}$$

$$\hat{N}_j = \frac{1 - \gamma_{jj}(n_{jj}^0 \hat{n}_{jj})}{1 - \sum_d y_{jd}^0 \gamma_{jd}(n_{jd}^0)}, \tag{28}$$

with $\gamma_{ij}(n)$ the share of labor employed to cover the fixed costs of firms from i selling in j, as defined in (OA.21).

Equation (26) follows from expression (23) for $i = j$. Accordingly, the first term measures substitution towards domestic goods with $\hat{x}_{jj} = 1/x_{jj}^0$, and is no longer second-order. Conditional on domestic substitution, the two additional terms in equation (26) arise from the entry and selection decisions of domestic firms, which are given by equations (27)-(28). The discussion in Section 3.1 shows that these channels affect gains from trade only through the shape of the trade elasticity functions. When moving to autarky, selection out of exporting ($d \ln n_{ij} < 0$ for $i \neq j$) leads to higher domestic firm survival ($d \ln n_{jj} > 0$) and higher welfare, $\ln \hat{n}_{jj} \rho_{jj}(n_{jj}^0 \hat{n}_{jj})/\rho_{jj}(n_{jj}^0) = \int_{\ln n_{jj}^0}^{\ln n_{jj}^0 \hat{n}_{jj}} (1 + \theta_{jj}^c(u)) du > 0$. The magnitude of this effect depends on how different marginal and inframarginal domestic firms are, as measured by $\theta_{jj}^c(n)$ in (13), and how strong domestic selection changes are, as measured by $\hat{n}_{jj}$ in (27). Furthermore, the welfare contribution of domestic firm entry may be positive or negative, depending on whether the trade elasticity is increasing or decreasing on n.

Expression (26) is related to the sufficient statistic for the gains from trade in Arkolakis et al. (2012). Equation (27) implies that, locally, $\theta_{jj}^0 d \ln \epsilon_{jj}(n_{jj}) = -d \ln x_{jj} + d \ln N_j$, and

$$d \ln \frac{w_j}{P_j} = -\frac{1}{\tilde{\theta}_{jj}^0} (d \ln x_{jj} - d \ln N_j), \tag{29}$$

where $\tilde{\theta}_{jj}^0 \equiv (\sigma - 1)\theta_{jj}^0$ is the elasticity of domestic spending to domestic cost at the initial equilibrium. As in Arkolakis et al. (2012), expression (29) shows that the gains from trade depend on the domestic spending share. However, it indicates that what matters is the (variable) domestic trade elasticity, instead of the generic (constant) trade elasticity in Arkolakis et al. (2012). Our formula highlights the importance of accounting for trade elasticity heterogeneity and endogenous firm entry, in line with the insights implied by equations (32)-(33) of Melitz and Redding (2015). Intuitively, a higher domestic elasticity means that it is easier to substitute foreign varieties for domestic varieties (through both extensive and intensive margins), which attenuates the welfare consequences of having to spend more on domestic varieties.

We view the above discussion as a synthesis of the results in Melitz and Redding (2015), who stress the importance of variable trade elasticities and firm heterogeneity, and the results in Arkolakis et al. (2012), who stress the sufficient role of the trade elasticity parameter in constant-elasticity gravity models. Relative to them, our characterization indicates that what matters are the trade elasticity functions, which summarize the entry and selection decisions of heterogeneous firms. We now turn to the estimation of these functions.

4 Semiparametric Gravity

We develop a semiparametric approach to estimate $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ using the gravity-like equations for the margins of firm exports in (10) and (12). We then use Monte Carlo simulations to illustrate the properties of our estimator and the potential for bias of parametric approaches in the literature.

4.1 Estimation Strategy

Consider data on the share of firms from i selling in j, n_{ij}, and their average sales, $\bar{x}_{ij}$, across markets. To leverage cross-market variation for estimation, we assume that the distribution of firm fundamentals has the same shape in all markets belonging to the same group:

Assumption 2. *Markets are divided into groups, $\mathcal{G}_g$, such that*

$$G_{ij}(r, e) = G_g(r/\bar{\eta}_{ij}^r, e/\bar{\eta}_{ij}^e) \quad \textit{for all} \quad ij \in \mathcal{G}_g. \tag{30}$$

This assumption imposes that, for all markets in the same group, the distribution of entry and revenue potentials only differs with respect to the (unobserved) scalars $\bar{\eta}_{ij}^r$ and $\bar{\eta}_{ij}^e$.[22]

[22] Our notation allows groups to be defined as destination-origin pairs over different years. In this case, our

Importantly, we do not impose any parametric restriction on the shape of the distribution of firm fundamentals. The main implication of Assumption 2 is to restrict the entry and revenue potential functions to be identical across all markets in the same group:

$$\ln \epsilon_{ij}(n) = \ln \epsilon_g(n) + \ln \bar{\eta}^e_{ij} \quad \text{and} \quad \ln \rho_{ij}(n) = \ln \rho_g(n) + \ln \bar{\eta}^r_{ij} \quad \text{for all} \quad ij \in \mathcal{G}_g.$$

Assumption 2 follows a long tradition in the estimation of endogenous selection models. Without it, the results in Heckman and Honore (1990) imply that cross-sectional data from firms operating in a single market cannot nonparametrically identify the distribution of firm fundamentals. Observing only one market requires parametric restrictions to extrapolate from the outcomes of active firms to the unobserved fundamentals of inactive firms. In contrast, consistent with Heckman and Honore (1990), Assumption 2 exploits cross-market variation in both the extensive and intensive margins of firm exports to nonparametrically identify $\epsilon_g(n)$ and $\rho_g(n)$. In our empirical application, we estimate these functions for groups of countries defined by per-capita income, market integration, and other shared characteristics.

We also impose restrictions on the data generating process of the bilateral shifters of revenue and entry.

Assumption 3. *We observe a vector of bilateral variables,* $z_{ij} = \{z_{ij,k}\} \in \mathbb{R}^K$*, such that*

$$\begin{aligned} \ln \bar{\eta}^r_{ij} \bar{r}_{ij} &= z_{ij}\kappa^r + \bar{\delta}^r_i + \bar{\zeta}^r_j + \eta^r_{ij}, \qquad \mathbb{E}[\eta^r_{ij}|z_{ij}, D] = 0 \\ \ln \bar{f}_{ij}/\bar{\eta}^e_{ij}\bar{r}_{ij} &= z_{ij}\kappa^e + \bar{\delta}^e_i + \bar{\zeta}^e_j + \eta^e_{ij}, \qquad \mathbb{E}[\eta^e_{ij}|z_{ij}, D] = 0 \end{aligned} \tag{31}$$

where D *is the matrix of origin and destination dummies, and* (κ^r, κ^e) *are real vectors of length* K *with known first entries* (κ^r_1, κ^e_1)*.*

Assumption 3 plays the central role of specifying observable variables z_{ij} whose variation across markets allows us to trace out the elasticities of the extensive and intensive margins of firm exports. This assumption has three parts, which we now discuss separately.

The first part of Assumption 3 is the separability of (31). Given origin and destination fixed-effects, bilateral shifters of revenue and entry are the sum of two components: the impact of the observed vectors, $z_{ij}\kappa^r$ and $z_{ij}\kappa^e$, and the unobserved shifters, η^r_{ij} and η^e_{ij}. Together with the equilibrium conditions for entry and sales in (10) and (12) under Assumption 2, equation (31) yields our semiparametric gravity equations:

$$\ln \bar{x}_{ij} - \ln \rho_g(n_{ij}) = z_{ij}\kappa^r + \delta^r_i + \zeta^r_j + \eta^r_{ij} \tag{32}$$

$$\ln \epsilon_g(n_{ij}) = z_{ij}\kappa^e + \delta^e_i + \zeta^e_j + \eta^e_{ij} \tag{33}$$

strategy could exploit variation over time for the same market while allowing the shape of the distribution to vary across all markets as long as it is constant over time for the same market.

where $\delta_i^e \equiv \ln(w_i^\sigma) + \bar{\delta}_i^e$, $\zeta_j^e \equiv \bar{\zeta}_j^e - \ln(E_j P_j^{\sigma-1})$, $\delta_i^r \equiv \ln(w_i^{1-\sigma}) + \bar{\delta}_i^r$, and $\zeta_j^r \equiv \ln\left(P_j^{\sigma-1}E_j\right) + \bar{\zeta}_j^r$. Holding all else constant, the comparison of n_{ij} and $\bar{x}_{ij}$ across markets with different observed shifters, $z_{ij}\kappa^r$ and $z_{ij}\kappa^e$, identifies $\epsilon_g(n)$ and $\rho_g(n)$. Note that the origin and destination fixed effects include endogenous outcomes (like wages and prices). For this reason, we need to maintain the assumption that the bilateral shifters are separable in the effect of the observable vector z_{ij}.[23]

The second part of Assumption 3 is the orthogonality between the observed and unobserved components, $\mathbb{E}[(\eta_{ij}^r, \eta_{ij}^e)|z_{ij}, D] = 0$. This is the formal notion of "all else constant" that allows us to trace out $\epsilon_g(n)$ and $\rho_g(n)$ from the responses of n_{ij} and $\bar{x}_{ij}$ to z_{ij}. It is the typical exogeneity assumption in the estimation of gravity equations for trade flows, as reviewed by Head and Mayer (2014). We use the implied moment conditions for the estimation of (32)-(33): for any function $\mathcal{Z}_g(.)$, $\mathbb{E}[\mathcal{Z}_g(z_{ij})(\eta_{ij}^r, \eta_{ij}^e)|D] = 0$. In our empirical application, z_{ij} includes trade cost shifters that are commonly used in the literature estimating gravity trade models. We follow Chen et al. (2024) by using splines to specify $\mathcal{Z}_g(.)$.

The last part of Assumption 3 is that we know the pass-through from one element of z_{ij} to the bilateral shifters, which we specify to be the first without loss. This assumption is analogous to that imposed by Heckman and Honore (1990) and Berry and Haile (2014). It is necessary to separate the impact of the bilateral shifters on n_{ij} and $\bar{x}_{ij}$ from the impact of z_{ij} on bilateral shifters. We pin down the scale of the bilateral shifters in terms of one component of z_{ij}. Such an assumption is implicit whenever observed shifters of trade costs are used for the estimation of the trade elasticity in gravity trade models. In our application, we follow Caliendo and Parro (2014) and Boehm et al. (2023) by imposing that variable trade costs are proportional to the cost of ad-valorem import tariffs.

Finally, we impose a basis for $\rho_g(n)$ and $\epsilon_g(n)$.

Assumption 4. *The functions $\rho_g(n)$ and $\epsilon_g(n)$ are spanned by restricted cubic splines, $f_m(\ln n)$, over knots $m = 1, ..., M$,*

$$\begin{bmatrix} \ln \rho_g(n) \\ \ln \epsilon_g(n) \end{bmatrix} = \sum_{m=1}^{M} \begin{bmatrix} \gamma_{g,m}^{\rho} f_m(\ln n) \\ \gamma_{g,m}^{\epsilon} f_m(\ln n) \end{bmatrix}. \tag{34}$$

We approximate the shape of $\rho_g(n)$ and $\epsilon_g(n)$ with a cubic spline function over each interval $[\bar{n}_m, \bar{n}_{m+1}]$ of the support $[0, 1]$, as in Ryan (2012). To improve precision, we restrict the bottom and upper intervals to have a log-linear slope. As shown in Stone (1985; 1990) and implemented by Harrell Jr (2001), restricted cubic splines can approximate any function

[23]However, as in Berry and Haile (2014), we could consider arbitrary functions of z_{ij}, $\kappa_g^r(z_{ij})$ and $\kappa_g^e(z_{ij})$, instead of the linear functions, $z_{ij}\kappa^r$ and $z_{ij}\kappa^e$.

in the context of a nonparametric regression model.[24] Our main estimates are based on three intervals ($M = 3$).

Under Assumption 4, we then recover the residuals as a function of parameters, $\Theta \equiv (\kappa^e, \kappa^r, \{\gamma^\rho_{g,m}, \gamma^\epsilon_{g,m}\}_{g,m=1}^{G,M}, \{\delta^r_i, \delta^e_i, \zeta^r_j, \zeta^e_j\}_{i,j=1}^{N,N})$:

$$\begin{bmatrix} \eta^r_{ij} \\ \eta^e_{ij} \end{bmatrix} = \begin{bmatrix} u^r_{ij}(\Theta) \\ u^e_{ij}(\Theta) \end{bmatrix} \equiv \begin{bmatrix} \ln \bar{x}_{ij} - z_{ij}\kappa^r \\ -z_{ij}\kappa^e \end{bmatrix} + \sum_{m=1}^{M} \begin{bmatrix} -\gamma^\rho_{g,m} f_m(\ln n) \\ \gamma^\epsilon_{g,m} f_m(\ln n) \end{bmatrix} - \begin{bmatrix} \delta^r_i + \zeta^r_j \\ \delta^e_i + \zeta^e_j \end{bmatrix}.$$

We then construct a Generalized Method of Moments (GMM) estimator for Θ:

$$\min_{\Theta} v(\Theta)' \hat{\Omega} v(\Theta), \quad \text{where} \quad v(\Theta) \equiv \begin{bmatrix} \sum_{ij} \left(u^r_{ij}(\Theta)\mathcal{Z}_g(z_{ij}), u^r_{ij}(\Theta)D_{ij}\right)' \\ \sum_{ij} \left(u^e_{ij}(\Theta)\mathcal{Z}_g(z_{ij}), u^e_{ij}(\Theta)D_{ij}\right)' \end{bmatrix}, \tag{35}$$

and $\hat{\Omega}$ is the two-step optimal matrix of moment weights.

There are two ways to interpret our strategy to estimate $\rho_g(n)$ and $\epsilon_g(n)$. First, imposing that $\rho_g(n)$ and $\epsilon_g(n)$ are given by the flexible functional form in Assumption 4 implies that identification, consistency, and inference follow from usual results for GMM. As such, identification requires the typical GMM rank condition (Newey and McFadden, 1994). Alternatively, Assumption 4 can be seen as a functional basis for the nonparametric estimation of $\rho_g(n)$ and $\epsilon_g(n)$. Under this interpretation, our estimator is the sieve nonparametric instrumental variable (NPIV) estimator in Chen and Qiu (2016), Chen and Christensen (2018), and Compiani (2019). In this case, identification requires the assumption of completeness in Newey and Powell (2003) or, in the case of our model with a linear component, the weaker version of this assumption in Florens et al. (2012).[25] Chen et al. (2024) derive confidence intervals for sieve NPIV estimators, which we report in our robustness analysis below.

4.2 Monte Carlo Simulations

We now turn to Monte Carlo simulations for three economies that differ in how the trade elasticity varies with the exporter firm share: constant, decreasing, and increasing. These simulations illustrate how the shape of the trade elasticity function affects the performance of our semiparametric strategy and of other parametric approaches in measuring the gains

[24] In particular Stone (1990) writes that, in order to "reduce the standard errors of log-spline estimates of extreme quantiles," he imposed "linear restrictions on the fitted splines" at the tails such that "the estimated distribution of the transformed variable has exponential tails." This regularity condition implies that the tail does not asymptote to infinity.

[25] The completeness assumption is not testable (Canay et al., 2013), but it is generically satisfied (Andrews, 2011; Chen and Christensen, 2018). If $\rho_g(n)$ and $\epsilon_g(n)$ are bounded, identification can be achieved by the weaker condition of bounded completeness (Blundell et al., 2007). We conduct inference on the shape of ρ_g and ϵ_g, through parameter estimates of the vectors γ^ρ_g and γ^ϵ_g.

from trade.

Simulation. We first briefly describe each simulated economy, with details in Appendix B.1. Figure 2 plots the true trade elasticity function (black solid line) for each economy. Panels (a) and (b) are motivated by common parametric assumptions in the literature, which imply a trade elasticity function that is either constant or decreasing with the exporter firm share, as generated by Pareto or log-normal distributions of entry potentials, respectively. The economy in Panel (c) features a trade elasticity that increases with the exporter firm share, implied by the modified Pareto distribution of entry potentials introduced in Section 2.2.

For each economy, we use Proposition 2 to simulate $b = 1, \dots, B$ realizations of the world equilibrium without international transfers. The simulated world consists of 100 ex-ante identical countries that differ only in their realizations of bilateral revenue shifters. In each simulation b, we independently draw the observed and unobserved components of revenue shifters from normal distributions.

In all three economies, we impose sufficient conditions to obtain the log-normal distribution of firm revenue specified by Head et al. (2014) and Bas et al. (2017).[26] As we formally established in Section 2, we can specify economies that have different trade elasticity functions, but the same distribution of firm-level exports. To highlight the distinction between the determinants of the trade elasticity function and the cross-sectional distribution of firm exports, our calibration imposes $G^r_{ij}(r|e) = \Phi(\ln r/\nu^r)$, so that the distribution of firm exports follows the log-normal distribution in equation (7) of Head et al. (2014): $G^{\ln R}_{ij}(x) = \Phi((x - \ln \bar{R}_{ij})/\nu^r)$ with $\Phi(.)$ the standard normal CDF.

Results. We consider two hypothetical researchers seeking to incorporate firm heterogeneity into the measurement of the gains from trade.

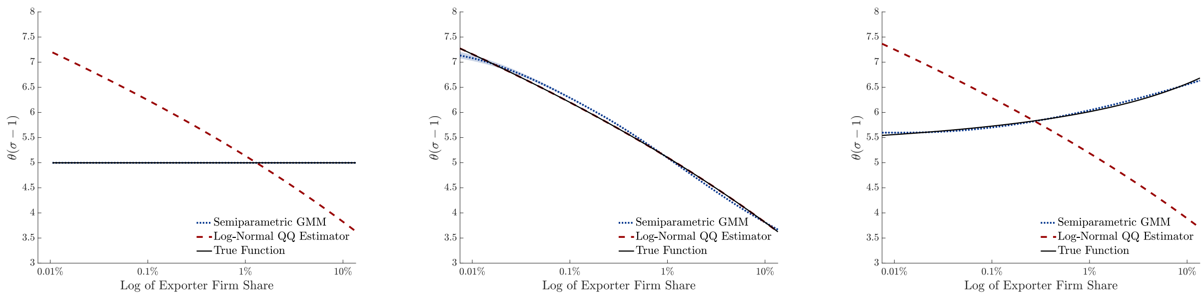
Researcher SP relies on the semiparametric approach introduced in Section 4.1: she implements the semiparametric GMM estimator in (35) to recover the trade elasticity functions and then applies Corollary 1 to measure the gains from trade.

Figure 2 shows that, across all three economies, our semiparametric approach performs well in recovering the shape of the trade elasticity function: the average estimate (blue dotted line) closely tracks the true function (black solid line). As a result, the semiparametric approach provides accurate estimates of the gains from trade. Across the three economies, the average mean squared error of the predicted gains from trade in the world is 3% of the

[26] The evidence in these papers motivate our specification of a log-normal distribution of firm sales, which generates quantiles of firm-level exports that fit almost perfectly their empirical analogs in France, China and, as we show below, Colombia (i.e., the R^2 of a regression of empirical on predicted quantiles is above 0.97).

Figure 2: Recovering Trade Elasticities: Monte Carlo Simulations

(a) Constant Trade Elasticity (b) Decreasing Trade Elasticity (c) Increasing Trade Elasticity



Note. Estimates obtained through simulation of 100 economies following the procedure summarized in Section 4.2 and detailed in Appendix B.1. All panels reports the trade elasticity function, $\theta_{ij}(n)$ defined in (14). The solid black line is the true underlying trade elasticity. The blue line is the average trade elasticity function recovered using our semiparametric estimator (i.e., Researcher SP's approach). The dashed red line is the average trade elasticity function recovered using the QQ log-normal estimator (Researcher P's approach). Panel (a) represents a model in which the trade elasticity is constant. Panel (b) represents a model in which the trade elasticity is decreasing, as implied by a log-normal distribution of entry potentials. Panel (c) represents a model in which the trade elasticity is increasing, as implied by the modified Pareto distribution of entry potentials in Section 2.2.

average gains from trade.[27]

Researcher P adopts a parametric micro approach: she implements an estimator that imposes a parametric distribution family for firm productivity to match the distribution of firm exports. In our setting, because the distribution of firm exports is log-normal, the best performing estimator must yield a log-normal distribution of firm sales. Hence, a minimum-distance estimator that matches the empirical distribution of firm sales is equivalent to the quantile-on-quantile (QQ) log-normal estimator proposed by Bas et al. (2017) – see Appendix B.1 for details.[28] In all three economies, the R^2 of the QQ log-normal estimator equals one, which indicates that it has a perfect fit for the distribution of firm exports, as illustrated for one realization of our simulations in Appendix Figure OA.5.

Figure 2 also reports the performance of researcher P's parametric approach that perfectly matches the distribution of firm exports in every market. In terms of recovering the true trade elasticity function, it performs poorly when its underlying parametric assumptions are violated. For instance, Panel (b) shows that the parametric approach accurately recovers the true trade elasticity function when the underlying distribution of firm fundamentals is indeed log-normal. However, in Panels (a) and (c), the parametric approach fails to recover the true elasticity functions because it imposes a decreasing relationship between trade elasticity and

[27]For each economy, the corresponding panel of Appendix Figure OA.3 reports the histogram across simulations of the mean square error of the predicted gains from trade in the world equilibrium normalized by the average gain from trade.

[28]Several papers follow a similar parametric approach (e.g., Arkolakis (2010), Eaton et al. (2011), and Egger et al. (2023)), where moments of the distribution of firm export outcomes are used in the estimation of parameters governing the distribution of firm fundamentals.

exporter firm share. Importantly, this misspecification produces large biases in the predicted gains from trade: the mean squared error is near zero in Panel (b), but rises sharply—to 251% and 215%—in Panels (a) and (c).

Consistent with Heckman and Honore (1990), our simulations demonstrate that parametric approaches can be severely biased despite (perfectly) replicating the cross-sectional distribution of firm exports. This follows from our calibration, which ensures that firm exports have a log-normal distribution in all three economies, despite their distinct trade elasticity functions.[29]

5 Estimation Results

In this section, we estimate the semiparametric gravity equations for the extensive and intensive margins of firm exports. Our results show how the trade elasticity varies with exporter firm shares and market characteristics.

5.1 Data

Our estimation sample contains 87 origin countries, and their firms' exports to 157 destination countries in 2012.[30] We measure the average firm exports as $\bar{x}_{ij} \equiv X_{ij}/N_{ij}$ with N_{ij} and X_{ij} denoting the number and sales of firms from i in j, respectively. We obtain N_{ij} and X_{ij} from the OECD Trade by Enterprise Characteristics (TEC) for a set of developed origins, the World Bank Exporter Dynamics Database (EDD) for a set of developing origins, and from administrative customs data for Australia and China. We consider the sales of the origins in our sample to all destinations in the dataset. Appendix Table OA.1 lists all origin countries in our sample, and the associated source for each variable used in estimation. Appendix Section B.2 presents further detail about data construction.

Turning to the exporter firm share, we note that n_{ij} is defined as the ratio between the number of firms from i selling in j, N_{ij}, and the number of entrants in i, N_i. The challenge to measure n_{ij} is that N_i is not easily available in national statistics, since it includes also entrants that decide to never produce. We circumvent this issue by noting that, although we

[29]Appendix Figure OA.4 presents qualitatively similar results for two additional hypothetical researchers who rely on a parametric version of the GMM estimator in (35). Specifically, instead of using the flexible function basis in Assumption 4, they implement the GMM estimator under the assumption that the elasticity functions follow either a constant-elasticity form, as specified in equation (22), or the shape implied by a log-normal distribution of firm productivity. The problem is that, as discussed in Section 2.2, such parametric assumptions impose a specific shape for the trade elasticity function, which may be inconsistent with its true shape.

[30]The choice of the year was determined by data availability, with the goal of maximizing coverage. Our sample accounted for 58% of world trade in 2012. We show that results are similar when we implement estimation in the period of 2010-2014, which has comparable coverage.

consider a static model to simplify exposition, our equilibrium is isomorphic to the stationary equilibrium of the dynamic setting in Melitz (2003), where the mass of successful entrants in market ij at any period, $n_{ij}N_i$, exactly replaces the mass of incumbents in ij exiting exogenously, δN_{ij} with δ denoting the exogenous death rate. Thus, we can measure the exporter firm share as $n_{ij} = n_{ii}N_{ij}/N_{ii}$ where, for origin i at any given period, N_{ii} is the number of active domestic firms and n_{ii} is the survival probability of new domestic entrants.[31] We measure n_{ii} as the one-year survival rate of tradable firms from the OECD Demographic Business Statistics (SDBS), and N_{ii} as the number of active tradable firms from the OECD Demographic Business Statistics (SDBS), the OECD Structural Statistics for Industry and Services (SSIS), and the World Bank Enterprise Surveys.[32]

We build on the gravity literature reviewed by Head and Mayer (2014) to include in z_{ij} the following variables: import tariffs, geographic distance, as well as dummies for trade agreements, shared language, shared currency, and colonial ties. The Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) is the main source of bilateral variables in z_{ij}. The only exception is the import tariff cost, which we define as the log of one plus the simple average of the bilateral tariffs across all 6-digit HS goods reported in the Global Tariff Database from Teti (2024).[33]

As stated above, we impose the common assumption in the literature that iceberg trade costs are proportional to import tariff costs – for a discussion, see Section 4 of Head and Mayer (2014). Thus, we set the pass-through parameters for tariffs to $\kappa_1^r = -\kappa_1^e = 1 - \sigma$. Throughout our analysis, we use estimates in the literature for the elasticity of substitution. We set $\sigma = 3.2$ to match the mean estimate of the cross-firm elasticity in Redding and Weinstein (2024).[34]

[31] Our approach implies that a low survival rate represents a large pool of entrants that pay the sunk entry cost but fail to be productive enough to survive. A high survival rate reflects instead that most firms paying the entry cost are successful in production. Notice that our approach is more general than that in prior research imposing that $N_i = N_{ii}$ and $n_{ii} = 1$ (e.g., Fernandes et al., 2023), which shuts down changes in domestic firm composition that were empirically documented by Pavcnik (2002) and Trefler (2004), and theoretically characterized by Melitz (2003).

[32] We have survival rates for 27 origins in our sample. We impute the survival rate for the remaining countries using the simple average of the survival rate for countries with available data. We show below that our results are robust to excluding from the sample countries without data on survival rates. We also show that our results are similar when we use survival rates over longer periods.

[33] As in standard gravity estimation, Assumption 3 implies that consistency requires exogeneity of the observed cost shifters. We follow the literature in our choice of the variables in z_{ij}, but it is possible that trade policy responds to unobserved components of bilateral trade costs. To ease such concerns, we show below that our estimates are similar, but less precise, if we use an instrumental variable for trade policy based on the strategy in Boehm et al. (2023), which leverages changes in MFN tariffs between 2002 and 2012.

[34] We show below that our main conclusions are robust to alternative assumptions about the pass-through from import tariffs to revenue and entry shifters. Alternatively, one can design a strategy to estimate σ using firm-level microdata on sales and prices for at least one market.

The availability of data on $\bar{x}_{ij}$, n_{ij}, and z_{ij} defines our estimation sample. For each origin i, Table OA.2 reports the number of destinations with positive trade, along with the average and the standard deviation of the exporter firm shares and average firm exports across destinations. In addition, Appendix Figure OA.6 summarizes the empirical distribution of $\ln(n_{ij})$. Since n_{ij} is the only input of the elasticity functions, we are only able to precisely estimate these functions in the part of the support for which we observe values of n_{ij}.

5.2 Estimates of Semiparametric Gravity of Firm Exports

5.2.1 Elasticity Heterogeneity with Respect to Exporter Firm Share

We start by estimating equation (35) for a single group pooling all markets. Under Assumption 2, we effectively restrict the shape of the distribution of entry and revenue potentials to be identical across all markets. Such an assumption is implicit in models imposing that all countries have identical gravity trade elasticities or the same shape parameters for the distribution of firm fundamentals. In our model, it yields common elasticity functions for all markets, which restricts the adjustment margins of firm exports to only vary across markets due to variation in the initial exporter firm share.

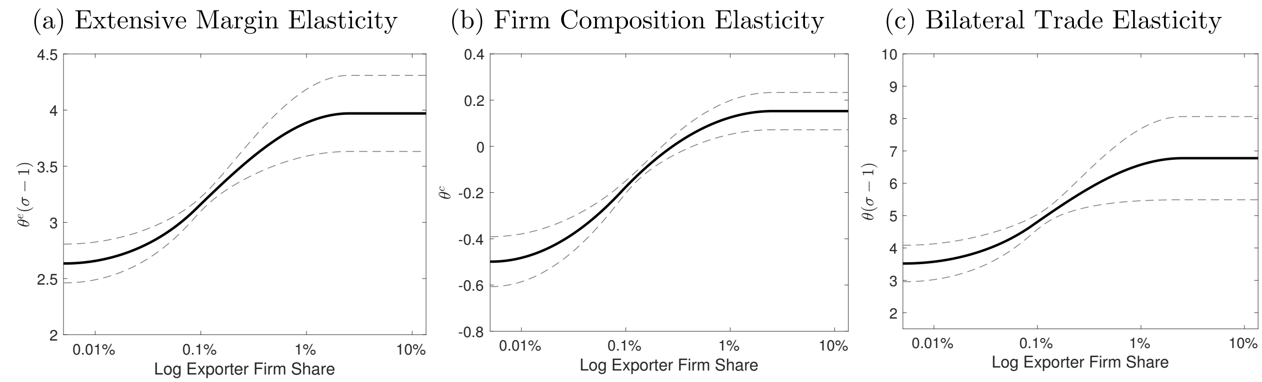
Figure 3 presents our semiparametric estimates. Panels (a) and (b) report our estimates of the extensive margin and firm composition elasticities, respectively. Panel (c) combines these elasticities to report the elasticity of bilateral trade flows with respect to bilateral trade costs. The solid lines are the baseline estimates obtained from (35), and the dashed lines are the associated 90% confidence intervals implied by robust standard errors.

Panel (a) shows that the extensive margin elasticity increases with the exporter firm share. For a 1 log-point increase in trade costs, we estimate a decrease in firm entry of 2.7 log-points in markets with low n_{ij}, but the estimated decrease is 4 log-points in markets with high n_{ij}. The entry potential distribution has a relatively lower mass among high-potential firms, which are the marginal firms in markets with few exporters.[35]

Turning to panel (b), we find an elasticity of firm composition that increases with the exporter firm share. For markets with low n_{ij}, our estimate of $\theta^c(n_{ij}) = -0.5$ implies that the average revenue potential of marginal entrants, $\mathbb{E}[r|e = \epsilon(n_{ij})]$, is 50% lower than that of incumbents, $\rho(n_{ij})$. In contrast, markets with high n_{ij} have $\theta^c(n_{ij})$ of roughly zero, implying that they have similar marginal and inframarginal firms. Accordingly, revenue potential differences are larger among firms with high entry potential that are marginal in markets with

[35] In line with the discussion in Section 2.2, the increasing extensive margin elasticity is consistent with an underlying distribution that decays faster than Pareto. In fact, Appendix Figure OA.7 shows that a log-Pareto distribution of entry potentials can approximate the positive slope of our estimated $\theta^e(n)$.

Figure 3: Semiparametric Gravity of Firm Exports – Single Group



Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination pairs for a single group pooling all pairs ($G = 1$). Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Thick lines are the point estimates and thin dashed lines are the 90% confidence intervals computed with robust standard errors.

a low n_{ij}. These markets exhibit stronger composition effects that attenuate the response of average firm exports to changes in trade costs. In fact, a 1 log-point increase in trade costs causes a decrease in average firm exports of 0.9 log-points in markets with low n_{ij} (i.e., $\theta^i(n)(\sigma - 1) \approx 0.9$ for $n < 0.01\%$), but the decrease is 2.3 log points in markets with high n_{ij} $\theta^i(n)(\sigma - 1) \approx 2.3$ for $n > 10\%$).

Panel (c) reports our estimates of the trade elasticity. Since the elasticities of both margins increase with the exporter firm share, so does the elasticity of bilateral trade flows. We estimate an elasticity of roughly 3.5 in markets with n_{ij} below 0.01%, which rises to about 6.5 in markets with n_{ij} above 10%. Our estimates are within the range reported in the literature using cross-country variation to estimate how bilateral trade responds to trade costs (Head et al., 2014), but are higher than the long-run estimates of Boehm et al. (2023) based on exogenous tariff changes over time. Appendix Figure OA.8 shows that trade elasticities are heterogeneous across markets. This heterogeneity reflects systematic variation with the exporter firm share, implying that firm heterogeneity shapes aggregate responses to trade costs changes.[36]

We note that our estimates are inconsistent with commonly used parametric assumptions in the literature, which imply a trade elasticity that is either constant or decreasing in the exporter firm share. Indeed, Appendix Figure OA.9 shows that, for at least part of the support, our estimated confidence intervals exclude the trade elasticity functions implied by

[36]In Table OA.4, we use a simple extension of standard gravity specifications to document that the trade elasticity increases with the exporter firm share. Columns (1) and (2) show that, relative to markets with n_{ij} below the median in our sample, the trade-to-distance elasticity is 0.22 higher in markets with n_{ij} above the median in our sample.

distributions of firm fundamentals that are Pareto, log-normal, or a combination of both.

Distribution of firm exports in Colombia. Are our trade elasticity estimates consistent with the empirical distribution of log exports across firms? As discussed in Section 2, this distribution depends on the shape of the conditional distribution of revenue potentials $G^r_{ij}(r|e)$, not only on its mean, $\mathbb{E}[r|\epsilon(n)] = (1+\theta^c(n))\rho(n)$. Thus, to recover the distribution of firm exports, we impose the additional assumption that $G^r_{ij}(r|e = \epsilon(n))$ follows a log-normal distribution with mean $\mathbb{E}[r|\epsilon(n)]$ and dispersion ν_{ij}, so that $G^{\ln R}_{ij}(x) = n^{-1}_{ij}\int_0^{n_{ij}} \Phi((x - \ln \mathbb{E}[r|\epsilon(n)])/\nu_{ij} + \delta_{ij})dn$ with δ_{ij} a market fixed-effect. Given our elasticity estimates, we choose (ν_{ij}, δ_{ij}) to match the quantiles of the empirical distribution of firm log exports, $Q_{ij}(p)$, by solving $\min_{(\nu_{ij},\delta_{ij})} \sum_{p=1}^{99}(G^{\ln R}_{ij}(Q_{ij}(p)) - p)^2$.

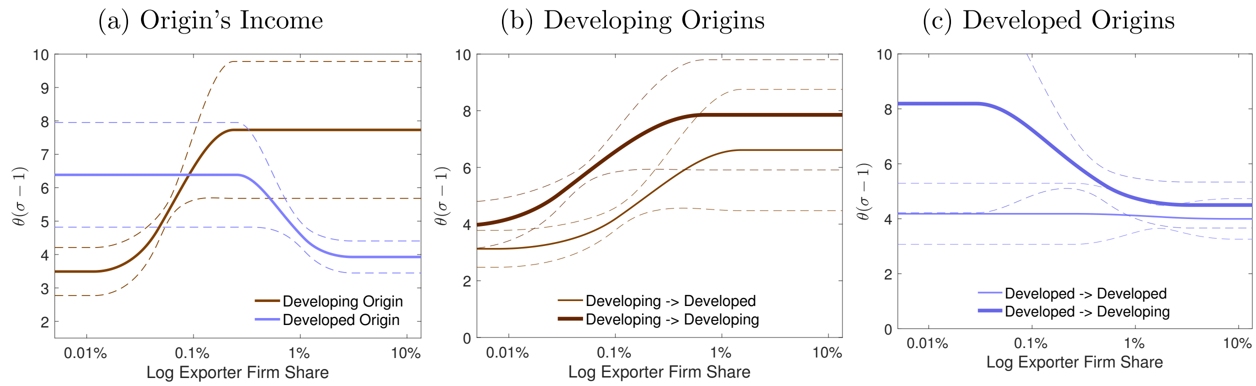
We implement this procedure for the distribution of log exports of Colombian firms to the ten largest destinations using the Exporter Firm Database from the World Bank (Fernandes et al., 2016). Appendix Table OA.5 shows that our model closely replicates the empirical quantiles of firm exports across all ten destinations. When we regress the empirical quantiles on their model-predicted counterparts, the estimated slope coefficient is close to one and the R^2 exceeds 0.98, as illustrated by Appendix Figure OA.10. Appendix Table OA.5 also reports that our model's fit for the distribution of firm exports is comparable to that of the QQ log-normal estimator in Bas et al. (2017). In line with our simulations, a joint log-normal distribution of firm fundamentals can reproduce the distribution of firm exports, but remains inconsistent with our estimated trade elasticity functions.

5.2.2 Elasticity Heterogeneity with Respect to Country Development

How do the elasticity functions of firm exports vary across markets? Formally, we relax the assumption of a common elasticity function for all markets. We instead assume that the distribution of revenue and entry potentials is only common within groups of markets with similar observable characteristics.

We first estimate elasticity functions that are specific to the level of development of the origin country. This allows developing and developed countries to differ in terms of the dispersion of firm-level fundamentals, in line with the evidence in Hsieh and Olken (2014). Panel (a) of Figure 4 reports estimates of the trade elasticity function separately for markets whose origin country is developing (dark brown) and developed (light purple), as defined by the World Bank (see Appendix Table OA.2). Estimated elasticity functions for developing origins are increasing in the exporter firm share and qualitatively similar to the pooled estimates reported in Figure 3 (due to developing origins being over-represented in our sample). In contrast, estimated elasticity functions for developed origins are decreasing

Figure 4: Bilateral Trade Elasticity – Developed and Developing Origins



Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination. Panel (a) assumes that there are two groups of markets ($G = 2$) defined by whether the origin country is developed (light purple) or developing (dark brown), as defined in Table OA.2. Panels (b) and (c) assume that there are four groups of markets ($G = 4$) defined by whether either the origin or destination is developed or developing, as defined in Table OA.2. We report estimates of the elasticity functions for developing origins in panel (b) and for developed origins in panel (c). The solid lines in all panels correspond to the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta_g(n)$. The estimator imposes that κ^r and κ^e are common across markets, as in our baseline specification. Dashed lines are the associated 90% confidence intervals computed with robust standard errors.

on the exporter firm share. We note that the estimates for the two groups are statistically different not only from each other, but also across levels of the exporter firm share.[37]

Our estimates suggest that developing countries have a higher density of firms with low entry and revenue potentials. As a result, they exhibit more pronounced effects of trade shocks on trade flows in markets with high n_{ij} (which have a relatively larger number of firms with low entry potential). In comparison, developed origins have a higher relative mass of firms with high entry and revenue potentials, which leads to more pronounced responses to trade shocks in markets with a low n_{ij} (which have a relatively larger number of firms with high entry potential). We also note that the qualitatively distinct shape of the trade elasticity function for the two country groups is consistent with Hsieh and Klenow (2009) and Hsieh and Olken (2014), who show that developing countries have a fatter tail of low-productivity firms (which have low entry and revenue potentials).[38]

We then examine whether the elasticity functions vary systematically with the development level of the destination country. This may occur if the destination's income affects the distribution of trade costs or the quality of available varieties (Waugh, 2010; Khandelwal,

[37] Appendix Table OA.4 shows again that this pattern emerges in a simple extension of a gravity specification. We now estimate the differential elasticity of trade to distance in markets with above median n_{ij} separately for developed and developing origins. While the estimated coefficient is positive for developing origins, it is negative for developed countries. In addition, Appendix Figure OA.11 shows that, even in a constant-elasticity benchmark, the trade elasticity is higher in developing countries compared to developed countries.

[38] Appendix Figure OA.12 reports estimates for four country groups based on income level, as defined by the World Bank. Despite wider confidence intervals due to fewer markets in each group, we estimate elasticity functions that are steeper for less developed countries, in line with Figure 4.

2010). Accordingly, panels (b) and (c) of Figure 4 present estimates based on four groups defined by whether either the origin or destination is developed or developing.

For developing origins, panel (b) shows that the trade elasticity functions have a similar shape regardless of the development level of the destination. In contrast, panel (c) shows that the destination's income level matters for the trade elasticity of developed origins. Interestingly, our estimated trade elasticity between developed countries is roughly constant at 4, remarkably close to existing estimates (Simonovska and Waugh, 2014). The trade elasticity is decreasing for exports from developed to developing countries, suggesting that firms in developed economies differ in their ability to serve developed versus developing markets.

5.3 Robustness

Within-sector variation. Our estimates so far have pooled together firms in all sectors. It is possible however that the elasticity heterogeneity documented above is driven by cross-market variation in sectoral firm composition. To address this concern, we now use within-sector variation to estimate the trade elasticity functions by defining markets as origin-destination-sector triplets and fixed effects as origin-sector and destination-sector.[39] Appendix Figure OA.13 shows that the shape of the within-sector estimates are similar to the baseline estimates above, with wider confidence intervals due to the smaller number of countries in our sample. In addition, Appendix Figures OA.14-OA.16 report sector-specific estimates of the elasticity functions that are consistent with the multi-sector model discussed in Section 2.4. We do so under the assumption that the sector-specific elasticity functions are the same in all origin-destination pairs due to the lower number of countries in our sector-level sample. While most sectors have similar shapes for the extensive elasticity function, they differ in their intensive margin elasticities. Combining the two margins, the bilateral trade elasticity has a similar shape in all sectors, despite its level varying across sectors.

Other dimensions of elasticity heterogeneity. Appendix Figure OA.17 investigates whether trade elasticity functions vary with determinants of market integration. This could be the case for example if deeper levels of integration have a disproportional impact on the trade costs of smaller firms. Panel (a) shows that deeply integrated markets, defined as those with a trade agreement and a common currency, also have a trade elasticity that is increasing

[39] We build the sample of origin-destination-sector triplets using the same data sources described in Section 5.1, which provide sector-level data for 46 origins (see Table OA.1). Our sector definition follows Boehm et al. (2023), as reported in Appendix Figure OA.16. The vector z_{ij} remains the same, but we use instead the simple average of import tariffs across 6-digit HS in each sector.

in the exporter firm share, with a threshold shifted to the right due to the higher levels of exporter firm share in this subsample of markets. Finally, panel (b) reports that estimates are also similar when we consider two market groups defined by whether they either have a common language or colonial ties. Overall, our estimates indicate that these determinants of market integration do not affect the strength of the adjustment margins of firm exports conditional on the exporter firm share.

Alternative cost shifter. Our baseline estimates impose that import tariffs do not affect fixed costs. In panels (a) and (b) of Appendix Figure OA.18, we show that estimates are similar if we assume instead that import tariffs affect both variable and fixed trade costs. In addition, panels (c) and (d) of Appendix Figure OA.18 report estimates with σ given by the 25th and 75th percentiles of the estimates in Redding and Weinstein (2024). Furthermore, our estimator is not affected much by the exact data source used for tariffs. We use those from Teti (2024), but Appendix Figure OA.19 shows that estimates are similar if we use the raw data from UN TRAINS.

Our estimation strategy also assumes that bilateral average tariffs are orthogonal to unobserved determinants of bilateral trade conditional on the bilateral gravity controls in z_{ij} and on origin and destination fixed effects. We assess the robustness of this assumption by using an instrumental variable for import tariffs inspired by Boehm et al. (2023). In particular, we define $z_{ij}^{\text{tariff,IV}} = \Delta_{2002-2012} \ln(1 + \text{MFNtariff}_j) \times 1[i, j \in \text{WTO}_{2002}] \times 1[i, j \notin \text{FTA}_{2002}]$, so that we exploit only bilateral variation in tariffs stemming from MFN tariff reductions over the decade preceding our sample year that affected WTO members without a free trade agreement. As such, our estimation does not rely on time-invariant bilateral determinants of tariffs or on variation arising from partner-specific tariff reductions. Appendix Figure OA.20 shows that the estimates obtained using this alternative set of moments are similar to our baseline results, though the confidence intervals are wider because $z_{ij}^{\text{tariff,IV}}$ explains only 18% of the variation in bilateral tariffs.

Alternative specifications. Our baseline confidence intervals are valid under Assumption 4. We now instead follow Chen et al. (2024) to provide confidence intervals under the assumption that (34) is a basis for the nonparametric estimation of $\epsilon(n)$ and $\rho(n)$. Appendix Figure OA.21 shows that this weaker assumption only slightly widens confidence intervals.

Our baseline estimates allow $\epsilon(n)$ and $\rho(n)$ to differ across three intervals of the support, as specified in Assumption 4 with $M = 3$. In Appendix Figure OA.22, we investigate the robustness with respect to this specification choice by allowing functions to vary across five intervals of the support; that is, we specify $M = 5$ in Assumption 4. This alternative specification yields similar estimates, albeit less precise.

Alternative sample. We show that estimates are similar when (i) we use data for the years of 2010 or 2014 with similar sample coverage (Appendix Figures OA.23–OA.24), (ii) we measure n_{ii} using three-year survival rates (Appendix Figure OA.25), and (iii) we exclude from the sample origin countries with imputed survival rates (Appendix Figure OA.26).[40]

6 The Aggregate Implications of Firm Export Decisions

We conclude by quantifying the contribution of firm export decisions for the aggregate impact of changes in trade costs. This exercise combines the elasticity estimates of Section 5 with the theoretical results of Section 3.

6.1 Welfare Gains from Trade

We begin by using Corollary 1 to compute the welfare gains from trade—that is, the change in welfare when we move from the 2012 equilibrium to autarky. We compare the gains implied by our semiparametric estimates in panels (b)–(c) of Figure 4 to those from two parametric benchmarks commonly used in the literature, Pareto and log-normal. Each panel of Figure 5 plots the difference in predicted gains from trade—relative to each benchmark—against the average exporter firm share across countries.

Panel (a) shows that, relative to a constant-elasticity gravity model, our estimates imply larger gains from trade in developed countries (by an average of 21%) but smaller gains in developing countries (by an average of 17%). The differences are often substantial—for instance, the predicted gains are about 33% lower for Lebanon but 25% higher for Luxembourg. These differences reflect the heterogeneity in our estimated trade elasticity functions, which rise with the exporter firm share in developing countries and indicate a higher relative mass of low-potential firms.

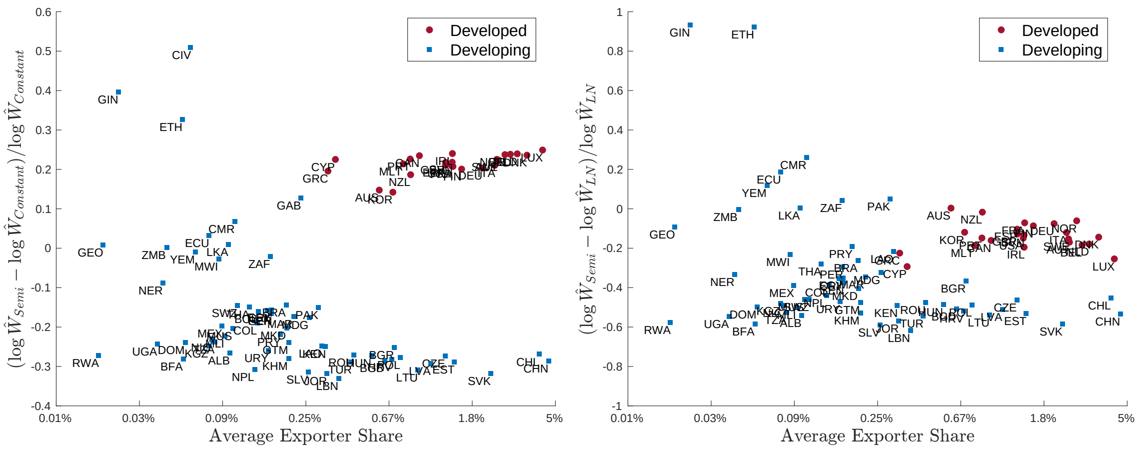
Panel (b) compares our baseline estimates to those implied by the log-normal benchmark. Differences are again larger for developing countries (by an average of 44%), since the log-normal specification imposes a decreasing trade elasticity. For developed countries, our approach yields smaller predicted gains, consistent with the milder slope of our estimated trade elasticity function relative to the slope of the log-normal benchmark.

These findings are consistent with the theoretical intuition developed in Section 3.1. Differences in the predicted gains from trade arise entirely from domestic firm entry and

[40]We note that we drop 75% of the origin-destination pairs in this case, which widens confidence internals.

Figure 5: The Gains From Trade: Comparison to Other Parametric Forms

(a) Baseline vs Pareto (b) Baseline vs Log-normal



Note. Gains from trade is minus the real wage change implied by moving from the observed equilibrium in 2012 to autarky, computed with the formula in Corollary 1. The vertical axis is the difference in welfare responses predicted by our semiparametric estimates and alternative parametric specifications for each country, divided by the welfare response implied by the alternative specification, and the horizontal axis is the log of the average exporter share of that country in 2012. In panel (a), semiparametric and constant-elasticity predictions are based on the elasticity estimates in Figures 4 and OA.9, respectively. For the log-normal specification in panel (b), we consider the elasticity functions implied by a log-normal productivity distribution with dispersion parameter based on the QQ log-normal estimates reported in Head et al. (2014) (Table 1, column 1).

selection. When the average exporter firm share is higher, a larger share of resources is allocated to exporting, which increases competitive pressure on domestic entry and selection — entirely through the slope of the trade elasticity function. Appendix Figure OA.27 shows that entry and selection of domestic firms are systematically related to the country's initial average exporter share, with the larger adjustments for developing countries reflecting the positive slope of the trade elasticity function. In addition, Appendix Figure OA.28 shows that our estimates indicate potentially large responses in the margins of firm profits, as defined in Section 3.2.[41]

6.2 Uniform Changes in Bilateral Trade Costs

We next consider a uniform reduction of 1% in bilateral trade costs between all origins and destinations starting from the observed equilibrium for $\{X^0_{ij}, n^0_{ij}\}$ in 2012. We focus on a uniform shock because a heterogeneous impact across countries comes entirely from heterogeneity in trade elasticities and initial conditions. We use Proposition 3.b to compute the model's counterfactual predictions for changes in all outcomes, which we feed into expression (24) to obtain the associated welfare responses and its components.

[41]We note that similar qualitative results hold for other parametric specifications in the literature, such as those illustrated in Appendix Figure OA.9. This is because these alternative parametric specifications yield trade elasticity functions whose slopes are inconsistent with our estimates.

Table 1: Impact of Uniform Reductions in Trade Costs on Welfare and its Components

Group of Countries	Welfare Elasticity ($\times$100)	Contribution to Welfare Elasticity				
		Neoclassical Components			Firm Components	
		Technology	Terms of trade	Substitution	Entry	Selection
All	3.17	101.7 %	-2 %	0.9 %	4 %	-4.5 %
Developed	3.97	95.1 %	-2.2 %	0.7 %	-4.7 %	11.2 %
Developing	1.80	126.4 %	-1.4 %	1.6 %	36.5 %	-63.1 %

Note. Starting from the observed equilibrium in 2012, we compute the counterfactual equilibrium implied by a reduction of 1% in bilateral trade costs, i.e. $\tau_{ij} = 0.99$ for all $i \neq j$, between all countries. For each group of countries, the second column of each panel reports 100 times the average log-change in real wage, weighted by each country's aggregate expenditure in 2012 and normalized by the shock size of 0.01. The remaining columns report the average of each component in (24) divided by the value reported in the second column. Counterfactual predictions computed with Proposition 3.b, given the elasticity estimates reported in panels (b) and (c) of Figure 4. We display welfare changes and its components for each country in Appendix Figure OA.29.

Table 1 reports counterfactual predictions by country group. We use the semiparametric estimates in Figure 4, which allow for elasticity heterogeneity with respect to both exporter firm shares and country development levels. The second column reports the average welfare response across all countries (first row), the subset of developed countries (second row) and the subset of developing countries (third row), weighted by each country's aggregate expenditure in 2012 and normalized by the shock size of 0.01. The other columns display the average of each component of welfare responses in equation (24) divided by the overall change reported in the second column.

The first row shows that, if trade costs were to decrease by 1% for all countries, then average global welfare would increase by 0.032%. In line with the discussion in Section 3.2, the average global welfare response is entirely given by the technology term, since the efficiency of the equilibrium implies that all other terms represent redistribution across countries and tend to cancel each other, with only minor differences as trade is not balanced.

The remaining rows indicate that the response of welfare to a uniform reduction in trade costs is larger for developed than developing countries. The primary reason for this difference is the larger technology term for developed countries, which follows from their higher trade openness in 2012. The other two neoclassical terms are small for both groups of countries; terms of trade because relative wages change little for a uniform shock across all countries, and demand substitution because it is second-order for small shocks.

The difference in welfare responses for developed and developing countries is further amplified by the firm components, which increase welfare for developed countries but reduce welfare for developing countries. This is a direct consequence of the estimates in Figure 4.

Consider first the firm entry component, whose response follows the intuition in Section 3.1. For developed countries, we estimate a trade elasticity with other developed countries that does not vary much with the exporter firm share. As a result, since developed countries mainly trade with other developed countries, they experience small changes in firm entry; that is, $\hat{N}_i$ is close to zero for developed countries (see Appendix Figure OA.30.a). In contrast, among developing countries, the increasing trade elasticity in Figure 4 leads to domestic firm entry following the increase in the number of exporters caused by the reduction in trade costs. Thus, firm entry has a positive contribution to welfare in developing countries, but this contribution only accounts for 37% of the average welfare gains.

The last column of the table reports the contribution of firm selection for welfare. It is positive and equivalent to 10% of gains for developed countries, but it is negative and equivalent to -63% of gains for developing countries. The trade cost reduction causes an increase in the average number of foreign varieties in all countries, but the increasing extensive margin elasticity in developing countries induces a larger drop in domestic firm selection. This leads to a decrease in the expenditure-weighted average number of firms operating in developing countries (see Appendix Figure OA.30).

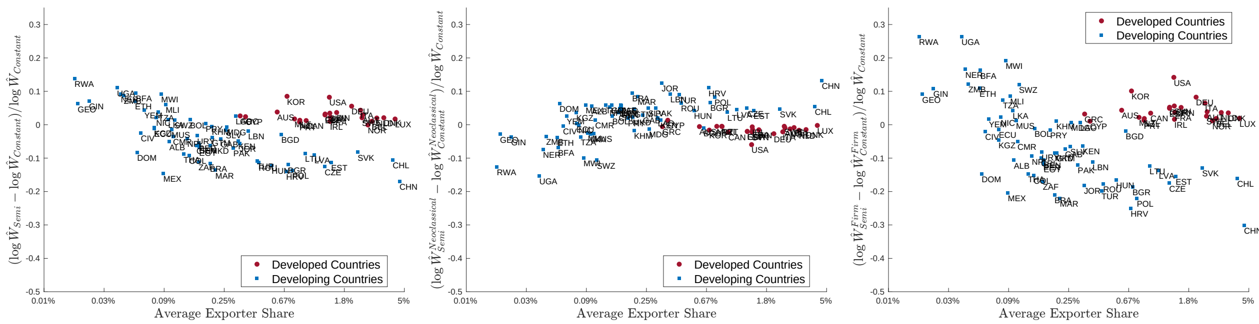
We then compare the predictions obtained using elasticity estimates from semiparametric and constant-elasticity specifications, as reported in Figures 4 and OA.9, respectively. In Figure 6 panel (a), the vertical axis is the difference in welfare responses predicted by the semiparametric and constant-elasticity specifications for each country, divided by the welfare response implied by the constant-elasticity benchmark. The horizontal axis is the log of the average exporter share of that country in 2012. In the other panels, the vertical axis is instead the difference in a component of welfare, divided by the overall welfare response implied by the constant-elasticity benchmark.

Panel (a) shows that parametric assumptions may have substantial effects on welfare predictions for different groups of countries. Relative to the constant-elasticity benchmark, our semiparametric estimates yield predicted welfare responses that are typically lower for developing countries. The average difference is 8% across developing countries, but it is as high as 16% and 18% for Mexico and China, respectively. In contrast, our semiparametric estimates yield larger welfare gains for developed countries, with an average of 4% and largest differences of 9% for South Korea and the United States. The deviations are systematically related with the country's average exporter firm share, with correlations of -0.72 and -0.20 for developing and developed countries, respectively.

Panel (b) shows that only a small fraction of the deviation comes from different predictions for the sum of neoclassical components, which is mainly driven by distinct terms of trade predictions, as the two specifications have identical technology terms and small substitution terms. Panel (c) indicates that the firm component is the main force behind the deviation between the two specifications and its correlation with the average exporter firm share. This follows from our estimates of the trade elasticity, which are systematically related with the

Figure 6: Impact of a Uniform Reduction in Bilateral Trade Costs on Welfare and its Components: The Role of Parametric Assumptions

(a) Welfare Change (b) Neoclassical Components (c) Firm Components



Note. We consider the impact of a reduction of 1% in bilateral trade costs between all countries starting from the observed equilibrium in 2012, i.e. $\tau_{ij} = 0.99$ for all $i \neq j$,. Panel (a) reports in the vertical axis is the difference in welfare responses predicted by the semiparametric and constant-elasticity specifications for each country, divided by the welfare response implied by the constant-elasticity benchmark, and the horizontal axis is the log of the average exporter share of that country in 2012. The other two panels report analogous scatter plots, but the vertical axis is instead the difference in components of predicted welfare responses, divided by the overall welfare response implied by constant-elasticity benchmark. Panel (b) does this for the sum of the neoclassical components associated with technology, terms of trade, and demand substitution in (24), and panel (c) for the sum of the firm components associated with entry and selection in (24). Semiparametric and constant-elasticity predictions use the elasticity estimates in Figures 4 and OA.9, respectively.

firm exporter shares through the export decisions of heterogeneous firms.[42]

Heterogeneous Changes in Bilateral Trade Costs. In Appendix C.2, we consider a counterfactual exercise inspired by the Generalized System of Preferences (GSP), in which developed countries lower import barriers on goods from developing countries. This exercise leverages a realistic policy setting to illustrate how our estimated heterogeneous trade elasticities interact with heterogeneous changes in bilateral trade costs. The main conclusions are similar to those under a uniform reduction in trade costs. However, in this case, the heterogeneous shock interacts with our elasticity estimates, giving firm export decisions a substantial role in driving changes in the terms of trade across countries.

7 Conclusion

We propose a new way to measure the aggregate implications of firm heterogeneity in monopolistic competition models with CES demand. We show that firm heterogeneity affects the extensive and intensive margins of firm exports through two nonparametric elasticity functions, which summarize all the key partial and general equilibrium predictions of the

[42]In fact, Appendix Figure OA.31 shows that heterogeneity in initial exporter firm share induces heterogeneity in trade elasticities and, as a result, large differences between the two specifications in predicted responses of trade flows, along both the extensive and intensive margins.

model. We estimate our model's semiparametric gravity equations for firm export margins, which indicate that trade elasticities vary with the number of exporters and the country's development level.

We show that flexibly allowing for firm heterogeneity is essential to accurately measure the impact of endogenous export decisions on the gains from trade. Compared with a constant-elasticity gravity model, this approach implies larger gains from trade in developed countries but smaller gains in developing ones. This pattern directly reflects our estimated heterogeneity in trade elasticities, which rise with exporter firm shares in developing countries and thus indicate a relatively higher mass of firms with low entry and revenue potentials. We note, however, that our analysis abstracts from additional mechanisms that may amplify gains from trade in developing countries—such as higher expenditure shares on traded goods due to non-homothetic preferences (Fajgelbaum and Khandelwal, 2016), knowledge diffusion through trade (Buera and Oberfield, 2020), and access to higher-quality intermediate and capital goods (Caselli and Wilson, 2004; Amiti and Konings, 2007; Halpern et al., 2015).

We view our work as a step toward moving beyond constant-elasticity gravity models in international trade. Within monopolistically competitive environments with CES preferences, our framework can be applied in two complementary ways: (i) by directly using the estimated trade elasticity functions instead of assuming a particular distribution of firm fundamentals, and (ii) by requiring that model-implied elasticity functions align with their empirical counterparts. In this sense, our estimated elasticity functions serve in heterogeneous-firm models the same purpose that the trade elasticity parameter serves in constant-elasticity gravity models following Arkolakis et al. (2012): they provide sufficient statistics that can be directly incorporated into quantitative analyses of the impact of trade shocks on welfare.

References

Adão, Rodrigo, "Worker heterogeneity, wage inequality, and international trade: Theory and evidence from Brazil," *Unpublished paper, MIT*, 2015.

Adao, Rodrigo, Arnaud Costinot, and Dave Donaldson, "Nonparametric counterfactual predictions in neoclassical models of international trade," *The American Economic Review*, 2017, *107* (3), 633--689.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi, "Universal gravity," *NBER Working Paper*, 2014, (w20787).

Amiti, Mary and Jozef Konings, "Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia," *American economic review*, 2007, *97* (5), 1611--1638.

__ , Sang Hoon Kong, and David E Weinstein, "Trade protection, stock-market returns, and welfare," Technical Report, National Bureau of Economic Research 2021.

Andrews, Donald WK, "Examples of l2-complete and boundedly-complete distributions," 2011.

Antras, Pol, Teresa C Fort, and Felix Tintelnot, "The margins of global sourcing: Theory and evidence from us firms," *American Economic Review*, 2017, *107* (9), 2514--2564.

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151--1199.

—, Arnaud Costinot, and Andres Rodríguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94--130.

—, —, Dave Donaldson, and Andrés Rodríguez-Clare, "The elusive pro-competitive effects of trade," *The Review of Economic Studies*, 2019, *86* (1), 46--80.

—, Sharat Ganapati, and Marc-Andreas Muendler, "The extensive margin of exporting products: A firm-level analysis," *American Economic Journal: Macroeconomics*, 2021, *13* (4), 182--245.

Atkeson, Andrew and Ariel Burstein, "Pricing to Market, Trade Costs, and International Relative Prices," *American Economic Review*, 2008, *98* (5), 1998--2031.

— and —, "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 2010, *118* (3), 433--489.

Atkin, David and Dave Donaldson, "The role of trade in economic development," in "Handbook of International Economics," Vol. 5, Elsevier, 2022, pp. 1--59.

Bai, Yan, Keyu Jin, and Dan Lu, "Misallocation under trade liberalization," *American Economic Review*, 2024, *114* (7), 1949--1985.

Baqaee, David and Emmanuel Farhi, "Networks, barriers, and trade," Technical Report, National Bureau of Economic Research 2019.

Baqaee, David Rezza, Emmanuel Farhi, and Kunal Sangani, "The darwinian returns to scale," *Review of Economic Studies*, 2024, *91* (3), 1373--1405.

Bartelme, Dominick, Ting Lan, and Andrei Levchenko, "Specialization, Market Access and Medium-Term Growth," 2019.

Bas, Maria, Thierry Mayer, and Mathias Thoenig, "From micro to macro: Demand, supply, and heterogeneity in the trade elasticity," *Journal of International Economics*, 2017, *108*, 1--19.

Berman, Nicolas, Philippe Martin, and Thierry Mayer, "How do different exporters react to exchange rate changes?," *The Quarterly Journal of Economics*, 2012, *127* (1), 437--492.

Bernard, Andrew B. and J. Bradford Jensen, "Exporting and Productivity in the USA," *Oxford Review of Economic Policy*, 2004, *20* (3), 343--357.

—, Stephen J. Redding, and Peter Schott, "Multi-Product Firms and Trade Liberalization," *Quarterly Journal of Economics*, 2011, *126* (3), 1271--1318.

Bernard, Andrew, Bradford Jensen, Steve Redding, and Peter J. Schott, "FIrms in International Trade," *Journal of Economic Perspectives*, 2007, *21* (3), 105--130.

Berry, Steven, "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25* (2), 242--262.

Berry, Steven T and Philip A Haile, "Identification in differentiated products markets using market level data," *Econometrica*, 2014, *82* (5), 1749--1797.

Blaum, Joaquin, Claire Lelarge, and Michael Peters, "The Gains From Input Trade in Firm-Based Models of Importing," 2015.

Blundell, Richard, Xiaohong Chen, and Dennis Kristensen, "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 2007, *75* (6), 1613--1669.

Boehm, Christoph E, Andrei A Levchenko, and Nitya Pandalai-Nayar, "The long and short (run) of trade elasticities," *American Economic Review*, 2023, *113* (4), 861--905.

Broda, Cristian and David Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121* (2), 541--585.

Buera, Francisco J and Ezra Oberfield, "The global diffusion of ideas," *Econometrica*, 2020, *88* (1), 83--114.

Burstein, Ariel and Jonathan Vogel, "International trade, technology, and the skill premium," *Journal of Political Economy*, 2017, *125* (5), 1356--1412.

Bustos, Paula, "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms," *American Economic Review*, 2011, *101* (1), 304--340.

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 2014.

Canay, Ivan A, Andres Santos, and Azeem M Shaikh, "On the testability of identification in some nonparametric models with endogeneity," *Econometrica*, 2013, *81* (6), 2535--2559.

Caselli, Francesco and Daniel J Wilson, "Importing technology," *Journal of monetary Economics*, 2004, *51* (1), 1--32.

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

Chen, Xiaohong and Timothy M Christensen, "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 2018, *9* (1), 39--84.

__ **and Yin Jia Jeff Qiu**, "Methods for nonparametric and semiparametric regressions with endogeneity: A gentle guide," *Annual Review of Economics*, 2016, *8*, 259--290.

__ **, Timothy Christensen, and Sid Kankanala**, "Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities," *Review of Economic Studies*, March 2024.

Compiani, Giovanni, “Market counterfactuals and the specification of multi-product demand: a nonparametric approach,” Technical Report, Working paper. 3, 43 2019.

Costinot, Arnaud and Andres Rodriguez-Clare, “Trade theory with numbers: Quantifying the consequences of globalization,” Technical Report, National Bureau of Economic Research 2013.

— and Andrés Rodríguez-Clare, “The US gains from trade: Valuation using the demand for foreign factor services,” *Journal of Economic Perspectives*, 2018, *32* (2), 3--24.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, “Global Rebalancing with Gravity: Measuring the Burden of Adjustment,” *IMF Staff Papers*, 2008, *55* (3), 511--540.

Dhingra, Swati and John Morrow, “The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition,” *mimeo, LSE*, 2012.

Dickstein, Michael J and Eduardo Morales, “What do exporters know?,” *The Quarterly Journal of Economics*, 2018, *133* (4), 1753--1801.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, “Dissecting Trade: Firms, Industries, and Export Destinations,” *American Economic Review, Papers and Proceedings*, 2004, *94* (2), 150--154.

— , — , and — , “An Anatomy of International Trade: Evidence from French Firms,” *Econometrica*, 2011, *79* (5), 1453--1498.

Egger, Peter H, Katharina Erhardt, and Sergey Nigai, “Empirical productivity distributions and international trade,” 2023.

Fajgelbaum, Pablo D and Amit K Khandelwal, “Measuring the unequal gains from trade,” *The Quarterly Journal of Economics*, 2016, *131* (3), 1113--1180.

Feenstra, Robert C and David E Weinstein, “Globalization, markups, and US welfare,” *Journal of Political Economy*, 2017, *125* (4), 1040--1074.

Fernandes, Ana M, Caroline Freund, and Martha Denisse Pierola, “Exporter behavior, country size and stage of development: Evidence from the exporter dynamics database,” *Journal of Development Economics*, 2016, *119*, 121--137.

— , Peter J Klenow, Sergii Meleshchuk, Martha Denisse Pierola, and Andrés Rodríguez-Clare, “The intensive margin in trade: How big and how important?,” *American Economic Journal: Macroeconomics*, 2023, *15* (3), 320--354.

Florens, Jean-Pierre, Jan Johannes, and Sébastien Van Bellegem, “Instrumental regression in partially linear models,” *The Econometrics Journal*, 2012, *15* (2), 304--324.

Halpern, László, Miklós Koren, and Adam Szeidl, “Imported inputs and productivity,” *American economic review*, 2015, *105* (12), 3660--3703.

Harrell Jr, Frank E, *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, Springer, 2001.

Head, Keith and Thierry Mayer, "Chapter 3 - Gravity Equations: Workhorse,Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, Vol. 4 of *Handbook of International Economics*, Elsevier, 2014, pp. 131 -- 195.

__ , __ , and Mathias Thoenig, "Welfare and trade without Pareto," *American Economic Review*, 2014, *104* (5), 310--316.

Heckman, James J and Bo E Honore, "The empirical content of the Roy model," *Econometrica: Journal of the Econometric Society*, 1990, pp. 1121--1149.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 2008, *2* (5), 441--487.

Hsieh, Chang-Tai and Benjamin A Olken, "The missing "missing middle"," *Journal of Economic Perspectives*, 2014, *28* (3), 89--108.

__ and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403--1448.

__ , Nicholas Li, Ralph Ossa, and Mu-Jeung Yang, "Accounting for the new gains from trade liberalization," *Journal of International Economics*, 2020, *127*, 103370.

Khandelwal, Amit, "The long and short (of) quality ladders," *The Review of Economic Studies*, 2010, *77* (4), 1450--1476.

Kleinman, Benny, Ernest Liu, and Stephen J Redding, "International friends and enemies," *American Economic Journal: Macroeconomics*, 2024, *16* (4), 350--385.

Kugler, Maurice and Eric Verhoogen, "The Quality-Complementarity Hypothesis: Theory and Evidence from Colombia," 2008. Manuscript, Columbia University.

Lind, Nelson and Natalia Ramondo, "Trade with correlation," Technical Report, National Bureau of Economic Research 2018.

Matsuyama, Kiminori and Philip Ushchev, "Beyond CES: Three Alternative Cases of Flexible Homothetic Demand Systems," *Buffett Institute Global Poverty Research Lab Working Paper*, 2017, (17-109).

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

Melitz, Marc J and Stephen J Redding, "Heterogeneous firms and trade," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 1--54.

__ and __ , "New trade models, new welfare implications," *American Economic Review*, 2015, *105* (3), 1105--46.

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended gravity," *The Review of economic studies*, 2019, *86* (6), 2668--2712.

Negishi, Takashi, "Welfare economics and existence of an equilibrium for a competitive economy," *Metroeconomica*, 1960, *12* (2-3), 92--97.

Newey, Whitney K and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 1994, *4*, 2111--2245.

— **and James L Powell**, "Instrumental variable estimation of nonparametric models," *Econometrica*, 2003, *71* (5), 1565--1578.

Novy, Dennis, "International trade without CES: Estimating translog gravity," *Journal of International Economics*, 2013, *89* (2), 271--282.

Pavcnik, Nina, "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *The Review of Economic Studies*, 2002, *69* (1), 245--276.

Perla, Jesse, Christopher Tonetti, and Michael E Waugh, "Equilibrium technology diffusion, trade, and growth," *American Economic Review*, 2021, *111* (1), 73--128.

Redding, Stephen J., "Theories of Heterogeneous FIrms and Trade," *Annual Review of Economics*, 2011, *3*, 77--105.

Redding, Stephen J and David E Weinstein, "Measuring aggregate price indices with taste shocks: Theory and evidence for CES preferences," *The Quarterly Journal of Economics*, 2020, *135* (1), 503--560.

— **and** — , "Accounting for trade patterns," *Journal of International Economics*, 2024, *150*, 103910.

Ryan, Stephen P, "The costs of environmental regulation in a concentrated industry," *Econometrica*, 2012, *80* (3), 1019--1061.

Simonovska, Ina and Michael E Waugh, "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 2014, *92* (1), 34--50.

Stone, Charles J, "Additive regression and other nonparametric models," *The annals of Statistics*, 1985, pp. 689--705.

— , "Large-sample inference for log-spline models," *The Annals of Statistics*, 1990, *18* (2), 717--741.

Teti, Feodora, "Missing Tariffs," *CESifo Working Paper*, 2024, (11590).

Tintelnot, Felix, "Global production with export platforms," *The Quarterly Journal of Economics*, 2017, *132* (1), 157--209.

Trefler, Daniel, "The Long and Short of the Canada-US Free Trade Agreement," *American Economic Review*, 2004, *94* (4), 870--895.

Waugh, Michael E, "International trade and income differences," *American Economic Review*, 2010, *100* (5), 2093--2124.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse, "Monopolistic Competition in General Equilibrium: Beyond the CES," *mimeo*, 2011.

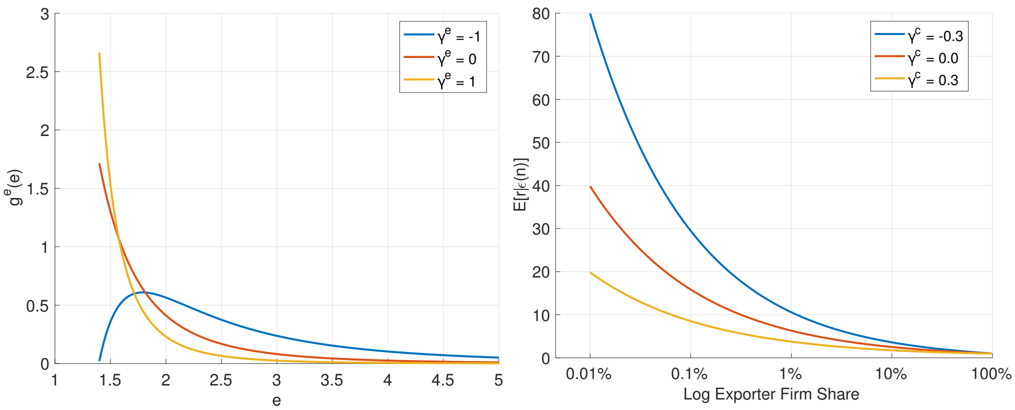
Online Appendix

A Theory Appendix: Proofs and Additional Results

A.1 Examples

Figure OA.1: Distributional Assumptions and the Firm Export Margins

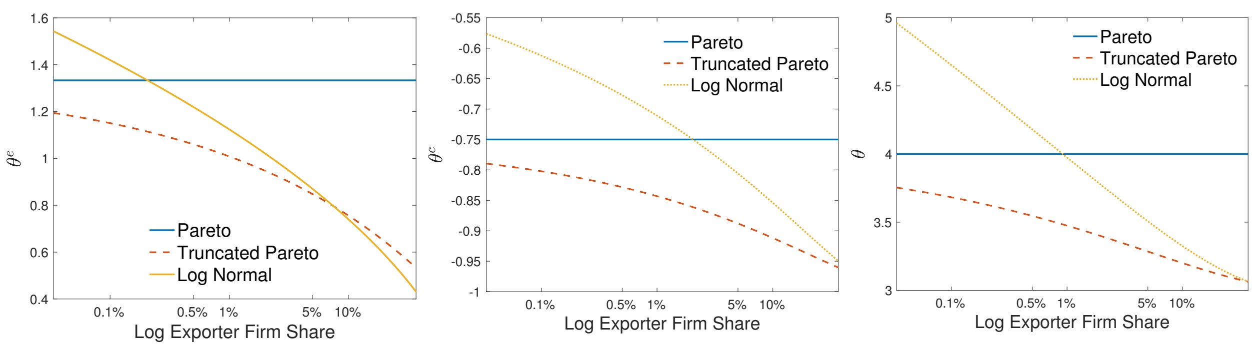
(a) Density of Entry Potentials   (b) Conditional Mean of Revenue Potentials



Note. Panel (a) reports the density of a modified Pareto distribution, $G^e(e) = 1 - (e/\underline{e})^{-\alpha^e}(\ln e/\ln \underline{e})^{-\gamma^e}$ with $e > \underline{e} > 1$, $\alpha^e = 3$ and $\underline{e} = \exp(1)$. Panel (b) reports the conditional mean revenue potential, $\mathbb{E}[r|e = \epsilon_{ij}(n)]$, for a modified power function, $\mathbb{E}[r|e = \epsilon_{ij}(n)] = n^{-\alpha^c}(1 - \ln(n))^{-\gamma^c}$ with $\alpha^c = 0.4$.

Figure OA.2: Distributional Assumptions and the Elasticity of Firm Export Margins

(a) Extensive Margin Elasticity   (b) Firm Composition Elasticity   (c) Bilateral Trade Elasticity



Note. Panel (a) reports the extensive margin elasticity, $\theta^e_{ij}(n)$ defined in (11), Panel (b) reports the firm composition elasticity, $\theta^c_{ij}(n)$ defined in (13), and panel (c) reports trade elasticity, $(\sigma - 1)\theta_{ij}(n)$ with $\theta_{ij}(n)$ defined in (14). We report the elasticity functions obtained when the productivity distribution is Pareto with shape parameter of four (Chaney, 2008), truncated Pareto with cutoff parameter of $H = 2.85$ (Melitz and Redding, 2015), or log-normal with dispersion parameter of 0.79 (Head et al., 2014).

A.2 Proofs

This Appendix contains proofs of all results in Sections 2 and 3. To simplify exposition, we introduce new notation for the elasticity of the functions $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$: $\varepsilon_{ij}(n) \equiv \partial \ln \epsilon_{ij}(n)/\partial \ln n$ and $\varrho_{ij}(n) \equiv \partial \ln \rho_{ij}(n)/\partial \ln n$, such that $\theta^e_{ij}(n) = -1/\varepsilon_{ij}(n)$, $\theta^c_{ij}(n) = \varrho_{ij}(n)$, $\theta^i_{ij}(n) = 1 - \varrho_{ij}(n)/\varepsilon_{ij}(n)$, and $\theta_{ij}(n) = 1 - (1 + \varrho_{ij}(n))/\varepsilon_{ij}(n)$.

A.2.1 Proof of Part b of Proposition 2

Equilibrium Efficiency. To prove the efficiency of the equilibrium, we show that we can find positive weights for the social planner problem so that its outcomes are the same as the competitive equilibrium.

Denote $v_i(\omega) = \{a_i(\omega), b_{ij}(\omega), \tau_{ij}(\omega), f_{ij}(\omega)\}_j$ with distribution $v_i(\omega) \sim \bar{G}_i(v)$. The Planner's problem is

$$\max_{\{q_{ij}(v), D_{ij}(v), N^p_i\}} \sum_j \chi_j \left[\sum_i N^p_i \int (\bar{b}_{ij} b_{ij})^{\frac{1}{\sigma}} (q_{ij}(v))^{\frac{\sigma-1}{\sigma}} D_{ij}(v) d\bar{G}_i(v) \right]$$

subject to

$$N^p_i \bar{F}_i + \sum_j N^p_i \int \left(\frac{\bar{\tau}_{ij}}{\bar{a}_i} \frac{\tau_{ij}}{a_i} q_{ij}(v) + \bar{f}_{ij} f_{ij} \right) D_{ij}(v) d\bar{G}_i(v) = \overline{L}_i$$

$$D_{ij}(v) \in \{0, 1\}$$

We use the definitions in (5) and (6) to re-write the problem in terms of revenue and entry potentials: by defining $\tilde{q}_{ij} = (\bar{b}_{ij} b_{ij})^{\frac{1}{\sigma-1}} q_{ij}$,

$$\max_{\{\tilde{q}_{ij}(r,e), D_{ij}(r,e), N^p_i\}} \sum_j \chi_j \left[\sum_i N^p_i \int (\tilde{q}_{ij}(r,e))^{\frac{\sigma-1}{\sigma}} D_{ij}(r,e) dG_{ij}(r,e) \right]$$

subject to

$$N^p_i \bar{F}_i + \sum_j N^p_i \int \left(\frac{\sigma - 1}{\sigma} (\bar{r}_{ij} r)^{\frac{1}{1-\sigma}} \tilde{q}_{ij}(r,e) + \bar{f}_{ij} \frac{r}{e} \right) D_{ij}(r,e) dG_{ij}(r,e) = \overline{L}_i$$

$$D_{ij}(v) \in \{0, 1\}$$

Thus, the Lagrangean is

$$\begin{aligned} \mathcal{L} \;=\;& \textstyle\sum_j \chi_j \left[\sum_i N^p_i \int (\tilde{q}_{ij}(r,e))^{\frac{\sigma-1}{\sigma}} D_{ij}(r,e) dG_{ij}(r,e) \right] \\ +\;& \textstyle\sum_i \lambda^p_i \left[\overline{L}_i - N^p_i \bar{F}_i - \sum_j N^p_i \int \left(\frac{\sigma-1}{\sigma} (\bar{r}_{ij} r)^{\frac{1}{1-\sigma}} \tilde{q}_{ij}(r,e) + \bar{f}_{ij} \frac{r}{e} \right) D_{ij}(r,e) dG_{ij}(r,e) \right] \end{aligned}$$

The first-order conditions of the problem imply that any solution must satisfy

$$\tilde{q}_{ij}(r,e) = (\bar{r}_{ij} r)^{\frac{\sigma}{\sigma-1}} \left(\frac{\lambda^p_i}{\chi_j} \right)^{-\sigma} \qquad \text{(OA.1)}$$

$$D_{ij}(r,e) = 1 \Leftrightarrow \chi_j (\tilde{q}_{ij}(r,e))^{\frac{\sigma-1}{\sigma}} \geq \lambda^p_i \left(\frac{\sigma - 1}{\sigma} (\bar{r}_{ij} r)^{\frac{1}{1-\sigma}} \tilde{q}_{ij}(r,e) + \bar{f}_{ij} \frac{r}{e} \right) \qquad \text{(OA.2)}$$

$$\sum_j \chi_j \int (\tilde{q}_{ij}(r,e))^{\frac{\sigma-1}{\sigma}} D_{ij}(r,e) dG_{ij}(r,e) = \lambda_i^p \left[\bar{F}_i + \sum_j \int \left(\frac{\sigma-1}{\sigma} (\bar{r}_{ij} r)^{\frac{1}{1-\sigma}} \tilde{q}_{ij}(r,e) + \bar{f}_{ij} \frac{r}{e} \right) D_{ij}(r,e) dG_{ij}(r,e) \right] \tag{OA.3}$$

$$N_i^p \bar{F}_i + \sum_j N_i^p \int \left(\frac{\sigma-1}{\sigma} (\bar{r}_{ij} r)^{\frac{1}{1-\sigma}} \tilde{q}_{ij}(r,e) + \bar{f}_{ij} \frac{r}{e} \right) D_{ij}(r,e) dG_{ij}(r,e) = \overline{L}_i \tag{OA.4}$$

Substituting (OA.1) into (OA.2),

$$D_{ij}(r,e) = 1 \Leftrightarrow e \geq e_{ij}^p \equiv \frac{\sigma \bar{f}_{ij}}{\bar{r}_{ij}} \left(\frac{\lambda_i^p}{\chi_j} \right)^\sigma . \tag{OA.5}$$

Substituting (OA.1) and (OA.5) into (OA.3),

$$\sum_j \frac{1}{\sigma} \bar{r}_{ij} \left((\lambda_i^p)^{1-\sigma} (\chi_j)^\sigma \right) \int_{e_{ij}^p}^{\infty} r dG_{ij}(r,e) = \lambda_i^p \left[\bar{F}_i + \sum_j \bar{f}_{ij} \int_{e_{ij}^p}^{\infty} \frac{r}{e} dG_{ij}(r,e) \right] \tag{OA.6}$$

$$\Rightarrow \overline{L}_i - N_i^p \bar{F}_i = \sum_j N_i^p \int \left(\frac{\sigma-1}{\sigma} (\bar{r}_{ij} r)^{\frac{1}{1-\sigma}} (\bar{r}_{ij} r)^{\frac{\sigma}{\sigma-1}} \left(\frac{\lambda_i^p}{\chi_j} \right)^{-\sigma} + \bar{f}_{ij} \frac{r}{e} \right) D_{ij}(r,e) dG_{ij}(r,e)$$

Substituting (OA.1), (OA.5) and (OA.3) into (OA.4)

$$\sum_j \bar{r}_{ij} \left((\lambda_i^p)^{1-\sigma} (\chi_j)^\sigma \right) \int_{e_{ij}^p}^{\infty} r dG_{ij}(r,e) = \frac{\lambda_i^p \overline{L_i}}{N_i^p} . \tag{OA.7}$$

Using the change of variable $n = 1 - G_{ij}^e(e)$ in (OA.6)–(OA.7),

$$\epsilon_{ij}(n_{ij}^p) = \frac{\sigma \bar{f}_{ij}}{\bar{r}_{ij}} \frac{(\lambda_i^p)^\sigma}{(\chi_j)^\sigma} \tag{OA.8}$$

$$\sum_j \frac{1}{\sigma} n_{ij}^p \bar{x}_{ij}^p = \lambda_i^p \left[\bar{F}_i + \sum_j \bar{f}_{ij} \int_0^{n_{ij}^p} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n)} (1 + \varrho_{ij}(n)) dn \right] \tag{OA.9}$$

$$\sum_j N_i^p n_{ij}^p \bar{x}_{ij}^p = \lambda_i^p \overline{L}_i \tag{OA.10}$$

with

$$\bar{x}_{ij}^p \equiv \bar{r}_{ij} \left((\lambda_i^p)^{1-\sigma} (\chi_j)^\sigma \right) \rho_{ij}(n_{ij}^p). \tag{OA.11}$$

Thus, given a set of positive weights $\{\chi_j\}$, the system (OA.8)–(OA.11) must be solved by any efficient allocation with firm export share n_{ij}^p, average firm exports $\bar{x}_{ij}^p$, mass of firms N_i^p, and multipliers λ_i^p. We note that, if we set the weight to be equal to the destination shifter of trade flows, $(\chi_j)^\sigma = P_j^{\sigma-1} E_j$, the system above becomes

$$\epsilon_{ij}(n_{ij}^p) = \frac{\sigma \bar{f}_{ij}}{\bar{r}_{ij}} \frac{(\lambda_i^p)^\sigma}{P_j^{\sigma-1} E_j} \tag{OA.12}$$

$$\sum_j \frac{1}{\sigma} n_{ij}^p \bar{x}_{ij}^p = \lambda_i^p \left[\bar{F}_i + \sum_j \bar{f}_{ij} \int_0^{n_{ij}^p} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n)} (1 + \varrho_{ij}(n)) dn \right] \tag{OA.13}$$

$$\sum_j N_i^p n_{ij}^p \bar{x}_{ij}^p = \lambda_i^p \overline{L_i} \tag{OA.14}$$

$$\bar{x}_{ij}^p \equiv \bar{r}_{ij}\left((\lambda_i^p)^{1-\sigma} P_j^{\sigma-1} E_j\right) \rho_{ij}(n_{ij}^p). \tag{OA.15}$$

Given the equilibrium conditions in Proposition 2, one solution of the system in (OA.12)–(OA.15) is $\lambda_i^p = w_i$, $N_i^p = N_i$, $n_{ij}^p = n_{ij}$ and $\bar{x}_{ij}^p = \bar{x}_{ij}$, where the efficient set of varieties from i available in j determined by (OA.5) is identical to the equilibrium set given by (6)–(7). This implies that the equilibrium is a solution of the planner's problem for $\chi_j = (P_j^{\sigma-1} E_j)^{1/\sigma}$ and thus it is efficient.

A.2.2 Proof of Section 3.1

Proposition 3.a: Small shocks. We start by totally differentiating the equilibrium equations for the extensive and intensive margins of firm-level exports. We simplify the notation by defining $\varepsilon_{ij}^0 \equiv \varepsilon_{ij}(n_{ij}^0)$, $\varrho_{ij}^0 \equiv \varrho_{ij}(n_{ij}^0)$, and $\theta_{ij}^0 \equiv \theta_{ij}(n_{ij}^0)$, such that $\theta_{ij}^0 = 1 - (1+\varrho_{ij}^0)/\varepsilon_{ij}^0$. Equations (10) and (12) respectively imply that

$$\varepsilon_{ij}^0 d\ln n_{ij} = d\ln \bar{f}_{ij} - d\ln \bar{r}_{ij} + \sigma d\ln w_i - (\sigma-1) d\ln P_j - d\ln E_j \tag{OA.16}$$

$$d\ln \bar{x}_{ij} = \varrho_{ij}^0 d\ln n_{ij} + d\ln \bar{r}_{ij} + (1-\sigma) d\ln w_i + (\sigma-1) d\ln P_j + d\ln E_j. \tag{OA.17}$$

We can then use these equations to obtain an expression for the change in bilateral trade flows, $d\ln X_{ij} = d\ln n_{ij} + d\ln \bar{x}_{ij} + d\ln N_i$:

$$d\ln X_{ij} = \theta_{ij}^0 d\ln \bar{r}_{ij} + (1-\theta_{ij}^0) d\ln \bar{f}_{ij} + (1-\theta_{ij}^0\sigma) d\ln w_i + \theta_{ij}^0(\sigma-1) d\ln P_j + \theta_{ij}^0 d\ln E_j + d\ln N_i, \tag{OA.18}$$

where we use $\theta_{ij}^0 = 1 - (1+\varrho_{ij}^0)/\varepsilon_{ij}^0$.

We now turn to the free entry condition in (15). When combined with the labor market clearing condition in (19), equation (15) is equivalent to

$$\frac{1}{N_i} = \frac{\sigma \bar{F}_i}{\bar{L}_i} + \sum_j \frac{\sigma \bar{f}_{ij}}{\bar{L}_i} \int_0^{n_{ij}} \frac{\rho_{ij}(n')}{\epsilon_{ij}(n')} (1+\varrho_{ij}(n')) dn' \tag{OA.19}$$

where we use $\theta_{ij}^c(n) = \varrho_{ij}(n)$.

By totally differentiating this expression, we get that

$$\begin{aligned} -\tfrac{1}{N_i^0} d\ln N_i \quad &= \quad \tfrac{\sigma \bar{F}_i^0}{\bar{L}_i^0} d\ln(\bar{F}_i/\bar{L}_i) + \textstyle\sum_j \tfrac{\sigma \bar{f}_{ij}^0}{\bar{L}_i^0} \left[\int_0^{n_{ij}^0} \tfrac{\rho_{ij}(n')}{\epsilon_{ij}(n')} (1+\varrho_{ij}(n')) dn'\right] d\ln(\bar{f}_{ij}/\bar{L}_i) \\ &+ \quad \textstyle\sum_j \tfrac{\sigma \bar{f}_{ij}^0}{\bar{L}_i^0} \tfrac{n_{ij}^0 \rho_{ij}(n_{ij}^0)}{\epsilon_{ij}(n_{ij}^0)} (1+\varrho_{ij}^0) d\ln n_{ij}. \end{aligned}$$

Note that, by adding (10) and (12), $\sigma \bar{f}_{ij} w_i = \bar{x}_{ij}\epsilon_{ij}(n_{ij})/\rho_{ij}(n_{ij}) = X_{ij}\epsilon_{ij}(n_{ij})/N_i n_{ij}\rho_{ij}(n_{ij})$. Thus, the expression above can be written as

$$-d\ln N_i = (1 - \sum_j y_{ij}^0 \gamma_{ij}(n_{ij}^0)) d\ln(\bar{F}_i/\bar{L}_i) + \sum_j y_{ij}^0 \gamma_{ij}(n_{ij}^0) d\ln(\bar{f}_{ij}/\bar{L}_i) + \sum_j y_{ij}^0 (1+\varrho_{ij}^0) d\ln n_{ij} \tag{OA.20}$$

with $y_{ij} \equiv X_{ij}/\sum_{j'} X_{ij'}$ the share of sales to destination j in the output of origin i, and

$$\gamma_{ij}(n) \equiv \frac{\epsilon_{ij}(n)}{n\rho_{ij}(n)} \int_0^n \frac{\rho_{ij}(n')}{\epsilon_{ij}(n')} (1+\varrho_{ij}(n')) dn. \tag{OA.21}$$

Finally, using the definition of θ_{ij}^0, we get that

$$-d\ln N_i = (1-\sum_j y_{ij}^0\gamma_{ij}(n_{ij}^0))d\ln(\bar{F}_i/\bar{L}_i) + \sum_j y_{ij}^0\gamma_{ij}(n_{ij}^0)d\ln(\bar{f}_{ij}/\bar{L}_i) + \sum_j y_{ij}^0(1-\theta_{ij}^0)\varepsilon_{ij}d\ln n_{ij} \quad \text{(OA.22)}$$

which in combination with (OA.16) implies that

$$\begin{aligned} d\ln N_i &= d\ln(\bar{L}_i) - (1-\textstyle\sum_j y_{ij}^0\gamma_{ij}(n_{ij}^0))d\ln\bar{F}_i - \sum_j y_{ij}^0\gamma_{ij}(n_{ij}^0)d\ln\bar{f}_{ij} \\ &- \textstyle\sum_j y_{ij}^0(1-\theta_{ij}^0)\left(d\ln\bar{f}_{ij} - d\ln\bar{r}_{ij} + \sigma d\ln w_i - (\sigma-1)d\ln P_j - d\ln E_j\right). \end{aligned} \quad \text{(OA.23)}$$

The budget balance condition in (18) implies that

$$d\ln E_j = \iota_j^0(d\ln w_j + d\ln\bar{L}_j) + (1-\iota_j^0)d\ln\bar{T}_j = \sum_i x_{ij}^0 d\ln X_{ij} \quad \text{(OA.24)}$$

with $\iota_j = \sum_{j'} X_{jj'}/\sum_{j'} X_{j'j}$ the income-to-spending ratio and x_{ij} the share of origin i in destination j's spending.

The labor market clearing condition in (19) implies that

$$\sum_j y_{ij}^0 d\ln X_{ij} = d\ln w_i + d\ln\bar{L}_i. \quad \text{(OA.25)}$$

The system of equations (OA.18), (OA.23), (OA.24) and (OA.25) determines $\{d\ln X_{ij}, d\ln P_i, d\ln N_i, d\ln w_i\}$ as a function of shocks in exogenous fundamentals, $\{d\ln\bar{r}_{ij}, d\ln\bar{f}_{ij}, d\ln\bar{L}_i, d\ln\bar{T}_i, d\ln\bar{F}_i\}$. To establish the proposition, consider the special case of this system for shocks in bilateral revenue shifters:

$$d\ln X_{ij} = \theta_{ij}^0 d\ln\bar{r}_{ij} + (1-\theta_{ij}^0\sigma)d\ln w_i + \theta_{ij}^0(\sigma-1)d\ln P_j + \theta_{ij}^0\iota_j^0 d\ln w_j + d\ln N_i, \quad \text{(OA.26)}$$

$$d\ln N_i = \textstyle\sum_j y_{ij}^0(\theta_{ij}^0-1)\left(-d\ln\bar{r}_{ij} + \sigma d\ln w_i - (\sigma-1)d\ln P_j - \iota_j^0 d\ln w_j\right), \quad \text{(OA.27)}$$

$$\iota_j^0 d\ln w_j = \sum_i x_{ij}^0 d\ln X_{ij}, \quad \text{(OA.28)}$$

$$\sum_j y_{ij}^0 d\ln X_{ij} = d\ln w_i. \quad \text{(OA.29)}$$

The proposition follows from the observation that, given any shock $\{d\ln\bar{r}_{ij}\}$, the system (OA.26)–(OA.29) can be solved only with knowledge of the (i) the demand elasticity of substitution σ, (ii) the bilateral trade matrix at the initial equilibrium $\{X_{ij}^0\}$ (since it implies $\{y_{ij}^0, x_{ij}^0, \iota_j^0\}$ by definition), and (iii) the bilateral trade elasticity matrix at the initial equilibrium $\{\theta_{ij}^0\}$.

To further establish the expression in (21), we note that

$$\begin{aligned} d\ln w_i &= \textstyle\sum_j y_{ij}^0\left(d\ln N_i + d\ln n_{ij} + d\ln\bar{x}_{ij}\right) \\ d\ln w_i &= \textstyle\sum_j y_{ij}^0\left(d\ln N_i + d\ln n_{ij} + d\ln w_i + (\varrho_{ij}^0-\varepsilon_{ij}^0)d\ln n_{ij}\right) \\ d\ln N_i &= -\textstyle\sum_j y_{ij}^0\left((1+\varrho_{ij}^0)/\varepsilon_{ij}^0 - 1\right)\varepsilon_{ij}^0 d\ln n_{ij} \\ d\ln N_i &= \textstyle\sum_j y_{ij}^0\theta_{ij}^0\varepsilon_{ij}^0 d\ln n_{ij} \end{aligned}$$

where the first equality follows from (OA.29), the second equality follows from $\sigma\bar{f}_{ij}w_i = \bar{x}_{ij}\epsilon_{ij}(n_{ij})/\rho_{ij}(n_{ij})$ (as implied by the sum of (10) and (12)), the third equality from $\sum_j y_{ij}^0 = 1$, and the last equality from the definition of θ^0_{ij}. This expression implies that

$$\sum_j y^0_{ij}\varepsilon^0_{ij} d\ln n_{ij} = 0, \tag{OA.30}$$

since equation (OA.27) is equivalent to

$$d\ln N_i = \sum_j y^0_{ij}(\theta^0_{ij} - 1)\varepsilon^0_{ij} d\ln n_{ij}. \tag{OA.31}$$

Thus, (OA.30) implies that

$$\varepsilon^0_{ii} d\ln n_{ii} = -\sum_{j\neq i} \frac{y^0_{ij}}{y^0_{ii}} \varepsilon^0_{ij} d\ln n_{ij}$$

By substituting this expression into (21),

$$d\ln N_i = \sum_{j\neq i} (\theta^0_{ij} - \theta^0_{ii}) y^0_{ij}\varepsilon^0_{ij} d\ln n_{ij}. \tag{OA.32}$$

Proposition 3.b: Large shocks. Let a variable with a "hat" ($\hat{y}_i \equiv y'_i/y^0_i$) denote the ratio between that variable at the initial equilibrium, y^0_i, and the counterfactual equilibrium, y'_i. We now characterize the system that determines changes in equilibrium outcomes for any arbitrary change in fundamentals, $\{\hat{\bar{r}}_{ij}, \hat{\bar{f}}_{ij}, \hat{\bar{L}}_i, \hat{\bar{T}}_i, \hat{\bar{F}}_i\}$. Equations (10) and (12) respectively imply that

$$\frac{\epsilon_{ij}(n^0_{ij}\hat{n}_{ij})}{\epsilon_{ij}(n^0_{ij})} = \frac{\hat{\bar{f}}_{ij}}{\hat{\bar{r}}_{ij}} \frac{\hat{w}^{\sigma}_i}{\hat{P}^{\sigma-1}_j \hat{E}_j}, \tag{OA.33}$$

$$\hat{\bar{x}}_{ij} = \frac{\rho_{ij}(n^0_{ij}\hat{n}_{ij})}{\rho_{ij}(n^0_{ij})} \hat{\bar{r}}_{ij} \hat{w}^{1-\sigma}_i \hat{P}^{\sigma-1}_j \hat{E}_j. \tag{OA.34}$$

By definition, changes in bilateral trade flows are given by

$$\hat{X}_{ij} = \hat{N}_i \hat{n}_{ij} \hat{\bar{x}}_{ij}. \tag{OA.35}$$

Using the fact that $\sigma \bar{f}_{ij} w_i = \bar{x}_{ij}\epsilon_{ij}(n_{ij})/\rho_{ij}(n_{ij})$ and the definition of $\gamma_{ij}(n)$ in (OA.21), the version of the free entry condition in (OA.19) is equivalent to

$$\frac{1}{N_i} = \frac{\sigma \bar{F}_i}{\bar{L}_i} + \sum_j \frac{n_{ij}\bar{x}_{ij}}{w_i \bar{L}_i} \gamma_{ij}(n_{ij}),$$

which, in combination with the fact that $\sigma N^0_i \bar{F}^0_i / \bar{L}^0_i = 1 - \sum_j y^0_{ij}\gamma_{ij}(n^0_{ij})$, implies that

$$\frac{1}{\hat{N}_i} = (1 - \sum_j y^0_{ij}\gamma_{ij}(n^0_{ij})) \frac{\hat{\bar{F}}_i}{\hat{\bar{L}}_i} + \sum_j y^0_{ij} \frac{\hat{n}_{ij}\hat{\bar{x}}_{ij}}{\hat{w}_i \hat{\bar{L}}_i} \gamma_{ij}(n_{ij}\hat{n}_{ij}). \tag{OA.36}$$

Finally, the equations for budget balance in (18) and market clearing in (19) immediately imply that

$$\hat{E}_j = \sum_i x^0_{ij} \hat{X}_{ij} = \iota^0_j \hat{w}_j \hat{\bar{L}}_j + (1 - \iota^0_j)\hat{\bar{T}}_j, \tag{OA.37}$$

$$\sum_j y_{ij}^0 \hat{X}_{ij} = \hat{w}_i \hat{\bar{L}}_i. \tag{OA.38}$$

Part b of the proposition follows from the fact that, for any shock in fundamentals $\{\hat{\bar{r}}_{ij}, \hat{\bar{f}}_{ij}, \hat{\bar{L}}_i, \hat{\bar{T}}_i, \hat{\bar{F}}_i\}$, counterfactual changes $\{\hat{n}_{ij}, \hat{\bar{x}}_{ij}, \hat{X}_{ij}, \hat{P}_i, \hat{N}_i, \hat{w}_i\}$ are given by the solution of the system (OA.33)–(OA.38), which depends on the (i) the demand elasticity of substitution σ, (ii) the elasticity functions $\rho_{ij}(n)$ and $\epsilon_{ij}(n)$ (since they imply $\gamma_{ij}(n)$ by definition), and (iii) the exporter firm share and bilateral trade matrices at the initial equilibrium $\{n_{ij}^0, X_{ij}^0\}$ (since $\{X_{ij}^0\}$ implies $\{y_{ij}^0, x_{ij}^0, \iota_j^0\}$ by definition).

Proof of Equation (25). Consider small changes in revenue shifters under trade balance ($\sum_i X_{ij}^0 = \sum_i X_{ji}^0$ for all j). We now derive an expression for the change in the average real wage across countries, weighted by their initial spending:

$$d\ln W \equiv \sum_j \frac{E_j^0}{E^0} d\ln \frac{w_j}{P_j}$$

with $E^0 \equiv \sum_j E_j^0$.

Note that, up to a first-order approximation, $\sum_i x_{ij}^0 \ln \hat{x}_{ij} = \sum_i x_{ij}^0 d\ln x_{ij} \approx \sum_i dx_{ij} = 0$. Thus, up to a first-order approximation, equation (24) becomes

$$d\ln \frac{w_j}{P_j} \approx \sum_i \frac{x_{ij}^0}{\sigma - 1} d\ln \bar{r}_{ij} + \sum_i x_{ij}^0 d\ln \frac{w_j}{w_i} + \sum_i \frac{x_{ij}^0}{\sigma - 1} d\ln N_i + \sum_i \frac{x_{ij}^0}{\sigma - 1}(1 + \varrho_{ij}^0) d\ln n_{ij}. \tag{OA.39}$$

Thus,

$$d\ln W \approx \sum_{i,j} \frac{X_{ij}^0}{E^0} \frac{d\ln \bar{r}_{ij}}{\sigma - 1} + \sum_{i,j} \frac{X_{ij}^0}{E^0} d\ln(w_j/w_i) + \frac{1}{\sigma - 1} \sum_{i,j} \frac{X_{ij}^0}{E^0} \left(d\ln N_i + (1 + \varrho_{ij}^0) d\ln n_{ij}\right). \tag{OA.40}$$

We now establish that the second and third terms in this expression are equal to zero. Consider the second term:

$$\begin{aligned}
\textstyle\sum_{i,j} X_{ij}^0 d\ln(w_j/w_i) &= \textstyle\sum_j \left(\sum_i X_{ij}^0\right) d\ln w_j - \sum_i \left(\sum_j X_{ij}^0\right) d\ln w_i \\
&= \textstyle\sum_j \left(\sum_i X_{ij}^0\right) d\ln w_j - \sum_j \left(\sum_i X_{ji}^0\right) d\ln w_j \\
&= \textstyle\sum_j \left(\sum_i X_{ij}^0 - \sum_i X_{ji}^0\right) d\ln w_j \\
&= 0
\end{aligned}$$

where the last equality follows from trade balance.

Turning to the third term, note that

$$\begin{aligned}
\textstyle\sum_{i,j} X_{ij}^0 \left(d\ln N_i + (1 + \varrho_{ij}^0) d\ln n_{ij}\right) &= \textstyle\sum_i (\sum_j X_{ij}^0) d\ln N_i - \sum_i \left(\sum_j X_{ij}^0 (\theta_{ij}^0 - 1)\varepsilon_{ij}^0 d\ln n_{ij}\right) \\
&= \textstyle\sum_i (\sum_j X_{ij}^0) d\ln N_i - \sum_i (\sum_j X_{ij}^0) d\ln N_i \\
&= 0
\end{aligned}$$

where the first equality uses the definition of θ_{ij}^0, and the second equality uses (OA.31).

Thus,

$$d\ln W \approx \sum_{i,j} \frac{X_{ij}^0}{E^0} \frac{d\ln \bar{r}_{ij}}{\sigma - 1}.$$

Special Case of Symmetric Countries. Consider small changes in trade costs in a world economy with symmetric countries such that

$$E_i^0 = E^0, \quad X_{ij}^0 = X_{ji}^0, \quad \theta_{ij}^0 = \theta_{ji}^0, \quad d\ln N_i = d\ln N, \quad d\ln \epsilon_{ij}(n_{ij}) = d\ln \epsilon_{ji}(n_{ji}). \tag{OA.41}$$

From equation (OA.27),

$$d\ln N = d\ln N_i = \frac{1}{E^0}\sum_j X_{ij}^0(\theta_{ij}^0 - 1)d\ln \epsilon_{ij}(n_{ij}). \tag{OA.42}$$

This implies that the firm entry and firm selection terms cancel each other for every country:

$$\begin{aligned}
\tfrac{1}{\sigma-1}\textstyle\sum_i x_{ij}^0\left[d\ln N_i + (1+\varrho_{ij}^0)d\ln n_{ij}\right] &= \tfrac{1}{\sigma-1}\textstyle\sum_i \frac{X_{ij}^0}{E_j^0}\left[d\ln N_i - (\theta_{ij}^0 - 1)d\ln \epsilon_{ij}(n_{ij})\right] \\
&= \tfrac{1}{\sigma-1}\left[d\ln N - \tfrac{1}{E^0}\textstyle\sum_i X_{ji}^0(\theta_{ji}^0 - 1)d\ln \epsilon_{ji}(n_{ji})\right] \\
&= \tfrac{1}{\sigma-1}\left[d\ln N - d\ln N\right] \\
&= 0
\end{aligned}$$

where the first equality uses the definition of θ_{ij}^0, the second equality uses the symmetry assumption in (OA.41), and the third equality uses (OA.42).

Note that, in this case, the terms of trade term is also equal to zero, since $d\ln w_j = d\ln w$ for all j. Thus, the first-order approximation for welfare in (OA.39) only contains the technology term.

Constant-Elasticity Benchmark. Consider small changes in trade costs under trade balance. We assume that the economy is given by the constant-elasticity benchmark for which the elasticities of $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ are identical across n and markets: for all n and ij,

$$\varrho_{ij}(n) = \varepsilon_{ij}(n) = -1/\theta \quad \text{and} \quad \theta_{ij}(n) = \theta.$$

The resource constraint in (OA.30) implies that $\sum_j y_{ij}^0 d\ln n_{ij} = 0$. Thus, the free entry condition in (OA.31) implies that

$$d\ln N_i = \sum_j y_{ij}^0\theta_{ij}^0\varepsilon_{ij}^0 d\ln n_{ij} = -\sum_j y_{ij}^0 d\ln n_{ij} = 0.$$

From (OA.18) and (OA.28), we obtain the following expression for the price index:

$$\theta(\sigma-1)d\ln P_j = (1-\theta)d\ln w_j - \sum_i x_{ij}^0\left(\theta d\ln \bar{r}_{ij} + (1-\theta\sigma)d\ln w_i\right) \tag{OA.43}$$

By combining this expression with equation (OA.18), we can re-write the market clearing condition in (OA.29) to obtain the following system of equations determining wages:

$$\theta\sigma d\ln w_i - \sum_j\left[y_{ij}^0 + (\theta\sigma-1)\sum_d y_{id}^0 x_{jd}^0\right]d\ln w_j = \sum_j y_{ij}^0\left(\theta d\ln \bar{r}_{ij} - \sum_o x_{oj}^0\theta d\ln \bar{r}_{oj}\right). \tag{OA.44}$$

Constant-Elasticity Benchmark with Two Countries. We now focus on the special case with two countries, Home ($i = H$) and Foreign ($i = F$), where $d\ln \bar{r}_{HH} = d\ln \bar{r}_{FF} = 0$. We define Foreign's wage as the numeraire, $d\ln w_F = 0$, and denote Home's wage change as $d\ln w_H = d\ln w$. Home's labor market clearing condition determines the equilibrium change in relative wages:

$$d\ln w = \frac{-y^0_{HH}x^0_{FH}\theta d\ln \bar{r}_{FH} + y^0_{HF}\left(1 - x^0_{HF}\right)\theta d\ln \bar{r}_{HF}}{\left(\theta\sigma - y^0_{HH} - (\theta\sigma - 1)\sum_{d=H,F} y^0_{Hd}x^0_{Hd}\right)}. \tag{OA.45}$$

Using (OA.43), we solve for Foreign's price index change,

$$\theta(\sigma - 1)d\ln P_F = -x^0_{HF}\left(\theta d\ln \bar{r}_{HF} + (1 - \theta\sigma)d\ln w\right),$$

which we plug into the extensive margin expression in (OA.16) to characterize firm selection:

$$d\ln n_{FF} = -x^0_{HF}\left(\theta d\ln \bar{r}_{HF} - \theta\sigma d\ln w\right) - x^0_{HF}d\ln w$$

$$d\ln n_{HF} = x^0_{FF}(\theta d\ln \bar{r}_{HF} - \theta\sigma d\ln w) - x^0_{HF}d\ln w$$

Thus, up to a first-order approximation, the decomposition in (24) becomes

$$d\ln\frac{w_F}{P_F} = \underbrace{\left(\frac{x^0_{HF}}{\sigma - 1}d\ln \bar{r}_{HF}\right)}_{\text{Technology}} + \underbrace{\left(-x^0_{HF}d\ln w\right)}_{\text{Terms of trade}} + \underbrace{0}_{\text{Firm entry}} + \underbrace{\left(\frac{1-\theta}{\theta(\sigma-1)}x^0_{HF}d\ln w\right)}_{\text{Firm selection}}. \tag{OA.46}$$

In combination with (OA.45), (OA.46) implies that, when countries are asymmetric, responses in terms of trade and firm selection have first-order impacts on welfare. Note however that, when countries are symmetric as defined in (OA.41), we have that $d\ln w = 0$ and thus both terms are second-order.

A.2.3 Proofs for Section 3.3

Proof of Corollary 1. We consider a counterfactual exercise in which an economy without international transfers moves to autarky. Specifically, we assume that $\hat{\bar{r}}_{ij} \to 0$ for all $i \neq j$, that $\iota^0_i = 1$ for all i, and that $\hat{\bar{F}}_i = \hat{\bar{f}}_{ij} = \hat{\bar{L}}_i = \hat{\bar{r}}_{ii} = 1$ for all i and j. We set the wage of country j to be the numerarie, $w_j \equiv 1$, so that $\hat{w}_j = 1$ and $\hat{E}_j = 1$.

By noticing that $\hat{x}_{jj} = 1/x^0_{jj}$, equation (23) implies that

$$\hat{P}^{1-\sigma}_j = x^0_{jj}\hat{N}_j\hat{n}_{jj}\frac{\rho_{jj}(n^0_{jj}\hat{n}_{jj})}{\rho_{jj}(n^0_{jj})}. \tag{OA.47}$$

We then use (OA.33) to substitute for $\hat{P}^{1-\sigma}_j$:

$$\frac{\epsilon_{jj}(n^0_{jj}\hat{n}_{jj})}{\epsilon_{jj}(n^0_{jj})} = x^0_{jj}\hat{n}_{jj}\hat{N}_j\frac{\rho_{jj}(n^0_{jj}\hat{n}_{jj})}{\rho_{jj}(n^0_{jj})}. \tag{OA.48}$$

Finally, since $\hat{N}_i\hat{n}_{ij}\hat{\bar{x}}_{ij} = 0$ for all $i \neq j$ and $y^0_{jj}\hat{y}_{jj} = y^0_{jj}\hat{N}_j\hat{n}_{jj}\hat{\bar{x}}_{jj} = 1$, the free entry condition in (OA.36) becomes

$$\hat{N}_j = \frac{1 - \gamma_{jj}(n^0_{jj}\hat{n}_{jj})}{1 - \sum_d y^0_{jd}\gamma_{jd}(n^0_{jd})}. \tag{OA.49}$$

The system (OA.47)–(OA.49) determines $\{\hat{n}_{jj}, \hat{N}_j, \hat{P}_j\}$ with $\hat{w}_j = 1$.

Proof of Equation (29). Equation (23) implies that

$$\begin{aligned}(\sigma-1)d\ln\tfrac{w_j}{P_j} &= -d\ln x_{jj}/N_j+(1+\varrho_{jj}^0)d\ln n_{jj}\\ &= -d\ln x_{jj}/N_j+(1-\theta_{jj}^0)\varepsilon_{jj}^0 d\ln n_{jj}\end{aligned}\tag{OA.50}$$

where the second equality uses the definition of $\theta_{jj}^0=1-(1+\varrho_{jj}^0)/\varepsilon_{jj}^0$.

Now note that, under trade balance ($d\ln E_j=d\ln w_j$) and no domestic shocks ($\ln\bar{f}_{jj}=d\ln\bar{r}_{jj}=0$), the extensive margin equation in (OA.16) implies that

$$\varepsilon_{jj}^0 d\ln n_{jj}=(\sigma-1)d\ln\frac{w_j}{P_j}.$$

Equation (29) immediately follows from the two expressions above.

A.2.4 Responses in Aggregate Profits

First, consider the share of profits in output, which is proportional to the mass of firms in a country due to free entry:

$$\pi_i\equiv\frac{N_i\sum_j\mathbb{E}[\max\{\pi_{ij}(\omega),0\}]}{w_i\bar{L}_i}=\frac{N_i\bar{F}_i}{\bar{L}_i},\tag{OA.51}$$

which implies that

$$d\ln s_i^\pi=d\ln N_i.$$

Second, consider the share of profits from a specific market j in aggregate output, π_{ij}:

$$\begin{aligned}\pi_{ij} &= \tfrac{N_i\mathbb{E}[\max\{\pi_{ij}(\omega),0\}]}{w_i\bar{L}_i}\\ &= \tfrac{N_i}{\bar{L}_i}\left(\tfrac{1}{\sigma}\tfrac{n_{ij}\bar{x}_{ij}}{w_i}-\bar{f}_{ij}\int_0^{n_{ij}}\tfrac{\rho_{ij}(n)}{\epsilon_{ij}(n)}(1+\theta_{ij}^c(n))dn\right)\\ &= \tfrac{N_i}{\bar{L}_i}\left(\tfrac{1}{\sigma}\tfrac{n_{ij}\bar{x}_{ij}}{w_i}-\tfrac{1}{\sigma}\tfrac{n_{ij}\bar{x}_{ij}}{w_i}\tfrac{\epsilon_{ij}(n_{ij})}{n_{ij}\rho_{ij}(n_{ij})}\int_0^{n_{ij}}\tfrac{\rho_{ij}(n)}{\epsilon_{ij}(n)}(1+\theta_{ij}^c(n))dn\right)\\ &= \tfrac{N_i}{\bar{L}_i}\tfrac{1}{\sigma}\tfrac{n_{ij}\bar{x}_{ij}}{w_i}\left(1-\gamma_{ij}(n_{ij})\right)\\ &= \tfrac{1}{\sigma}y_{ij}\left(1-\gamma_{ij}(n_{ij})\right)\end{aligned}$$

where the second row follows from the expressions for expected profits and costs in Section 2.3, the third row follows from $\sigma\bar{f}_{ij}w_i=\bar{x}_{ij}\epsilon_{ij}(n_{ij})/\rho_{ij}(n_{ij})$, the fourth row from the definition of $\gamma_{ij}(n_{ij})$ in (OA.21), the fifth row from $X_{ij}=N_i n_{ij}\bar{x}_{ij}$ and $w_iL_i=\sum_j X_{ij}$.

By definition, $\pi_i=\sum_j\pi_{ij}$ and, thus, $\pi_i=\sum_j\frac{1}{\sigma}y_{ij}\left(1-\gamma_{ij}(n_{ij})\right)$. Thus, the share of profits from domestic sales in aggregate profits, $s_{ii}^\pi\equiv\pi_{ii}/\pi_i$, is given by:

$$s_{ii}^\pi=\frac{y_{ii}\left[1-\gamma_{ii}(n_{ii})\right]}{\sum_j y_{ij}\left[1-\gamma_{ij}(n_{ij})\right]}.\tag{OA.52}$$

To understand how the domestic profit share changes, it is useful to consider a different way of writing this expression:

$$\begin{aligned}s_{ii}^\pi &= \tfrac{N_i\mathbb{E}[\max\{\pi_{ii}(\omega),0\}]}{\sum_j N_i\mathbb{E}[\max\{\pi_{ij}(\omega),0\}]}\\ &= \tfrac{1}{w_i\bar{F}_i}\left(\tfrac{1}{\sigma}n_{ii}\bar{x}_{ii}-w_i\bar{f}_{ii}\int_0^{n_{ii}}\tfrac{\rho_{ii}(n)}{\epsilon_{ii}(n)}(1+\theta_{ii}^c(n))dn\right)\\ &= \tfrac{\bar{L}_i}{\bar{F}_i}\left(\tfrac{1}{\sigma}\tfrac{x_{ii}}{N_i}-\tfrac{\bar{f}_{ii}}{\bar{L}_i}\int_0^{n_{ii}}\tfrac{\rho_{ii}(n)}{\epsilon_{ii}(n)}(1+\theta_{ii}^c(n))dn\right)\end{aligned}$$

where the second row follows from the expressions for expected profits and costs in Section 2.3, and the third row from $X_{ij} = N_i n_{ij} \bar{x}_{ij}$ and $w_i L_i = E_i$ under trade balance. For small shocks, this expression implies that

$$\begin{aligned} ds_{ii}^{\pi} &= \tfrac{1}{\sigma}\tfrac{\bar{L}_i}{\bar{F}_i}\tfrac{x_{ii}}{N_i}\left(d\ln x_{ii} - d\ln N_i\right) - \tfrac{1}{\sigma}\tfrac{\bar{L}_i}{\bar{F}_i}\tfrac{\sigma w_i \bar{f}_{ii}}{w_i \bar{L}_i}\tfrac{n_{ii}\rho_{ii}(n_{ii})}{\epsilon_{ii}(n_{ii})}(1+\theta_{ii}^{c}(n_{ii}))d\ln n_{ii} \\ &= \tfrac{1}{\sigma}\tfrac{\bar{L}_i}{\bar{F}_i}\tfrac{x_{ii}}{N_i}\left(d\ln x_{ii} - d\ln N_i - (1+\theta_{ii}^{c}(n_{ii}))d\ln n_{ii}\right) \\ &= -\tfrac{\sigma-1}{\sigma}\tfrac{x_{ii}}{\pi_i}d\ln\tfrac{w_i}{P_i} \end{aligned}$$

where the second row follows from $\sigma \bar{f}_{ij} w_i = X_{ij}\epsilon_{ij}(n_{ij})/N_i n_{ij}\rho_{ij}(n_{ij})$, the third row from OA.50, and the last row from the free entryOA.51. Thus,

$$d\ln s_{ii}^{\pi} = -(\sigma-1)\frac{\frac{1}{\sigma}x_{ii}}{\pi_{ii}}d\ln\frac{w_i}{P_i}$$

Note that we can define a profit share multiplier α_i^{π}:

$$\alpha_i^{\pi} \equiv \frac{\frac{1}{\sigma}x_{ii}}{\pi_{ii}} = \frac{\frac{1}{\sigma}X_{ii}}{\frac{1}{\sigma}X_{ii} - N_i n_{ii}\bar{c}_{ii}} > 1,$$

which measures the ratio of variable profits to profits net of fixed entry costs. Thus,

$$d\ln s_{ii}^{\pi} = -(\sigma-1)\alpha_i^{\pi} d\ln\frac{w_i}{P_i}. \tag{OA.53}$$

A.3 Extensions

This appendix presents extensions of our baseline framework. Section A.3.1 relaxes the assumption of CES demand in our baseline framework by allowing for a general class of demand functions with a single aggregator. In Section A.3.2, we extend our model to include import tariffs that generate government revenue, as well as heterogeneous firms in multiple sectors whose production function uses multiple factors and sector-specific inputs. In Section A.3.3, we relax the assumption of full support in the distribution of entry potentials to allow for zero trade flows between countries. Section A.3.4 extends our baseline framework to allow firms to produce multiple products.

A.3.1 Non-CES Demand and Variable Markups

Our baseline model considers a nonparametric distribution of firm fundamentals, while maintaining the typical parametric assumption of CES demand. We now show how our insights generalize for a class of single aggregator demand functions that allow for variable markups.

Environment

We maintain the same environment of Section 2.1, except that preferences are now given by

$$q_{ij}(\omega) = \frac{1}{b_{ij}(\omega)} q_j\left(\frac{1}{b_{ij}(\omega)}\frac{p_{ij}(\omega)}{D_j}\right), \tag{OA.54}$$

where D_j is a demand aggregator that is implicitly defined by the budget constraint in (2). CES demand is the special case in which $q_j(x) = x^{-\sigma}$ and $(D_j)^{\sigma} = P_j^{\sigma-1}E_j$.

To simplify notation, we drop the components of bilateral shifters that are common to all firms, and introduce them below when deriving the expressions for the margins of firm-level exports.

Entry and Revenue Potentials. We consider a monopolistic competitive environment in which firms take w_i and D_j as given. The firm's profit maximization problem conditional on entering market j is:

$$\max_p \left(p - w_i \frac{\tau_{ij}(\omega)}{a_i(\omega)}\right) \frac{1}{b_{ij}(\omega)} q_j \left(\frac{1}{b_{ij}(\omega)} \frac{p}{D_j}\right) - w_i f_{ij}(\omega),$$

with an associated FOC of

$$q_j \left(\frac{1}{b_{ij}(\omega)} \frac{p}{D_j}\right) + \left(\frac{1}{b_{ij}(\omega)} \frac{p}{D_j} - \frac{w_i/D_j}{r_{ij}(\omega)}\right) q_j' \left(\frac{1}{b_{ij}(\omega)} \frac{p}{D_j}\right) = 0, \tag{OA.55}$$

where we define the revenue potential in j of firm ω from i as

$$r_{ij}(\omega) \equiv \frac{b_{ij}(\omega) a_i(\omega)}{\tau_{ij}(\omega)}. \tag{OA.56}$$

The equilibrium condition in (OA.55) implicitly defines the optimal price of firm ω:

$$\frac{1}{b_{ij}(\omega)} \frac{p_{ij}(\omega)}{D_j} = \mathcal{P}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right).$$

This implies that, conditional on selling in j, firm ω from i has revenue, variable cost and variable profits given by

$$\frac{R_{ij}(\omega)}{D_j} = \mathcal{R}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) \equiv \mathcal{P}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) q_j \left(\mathcal{P}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right)\right) \tag{OA.57}$$

$$\frac{C_{ij}(\omega)}{D_j} = \mathcal{C}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) \equiv \frac{w_i/D_j}{r_{ij}(\omega)} q_j \left(\mathcal{P}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right)\right) \tag{OA.58}$$

$$\frac{\Pi_{ij}(\omega)}{D_j} = \Pi_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) \equiv \mathcal{R}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) - \mathcal{C}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) \tag{OA.59}$$

We assume that the demand function in (OA.54) implies that firms with a higher marginal cost have lower revenue and variable profit,

$$\mathcal{R}_j' < 0 \quad \text{and} \quad \Pi_j' < 0, \tag{OA.60}$$

with $\lim_{x\to 0} \mathcal{R}_j(x) = \infty$ and $\lim_{x\to\infty} \mathcal{R}_j(x) = 0$.[43]

Firm ω from i decides to sell in j if, and only if, $\Pi_{ij}(\omega) \geq w_i f_{ij}(\omega)$ which is equivalent to

$$\bar{\Pi}_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right) \equiv \frac{\Pi_j \left(\frac{w_i/D_j}{r_{ij}(\omega)}\right)}{\frac{w_i/D_j}{r_{ij}(\omega)}} > r_{ij}(\omega) f_{ij}(\omega)$$

Note that $\bar{\Pi}_j' < 0$ since $\Pi_j' < 0$. Thus,

$$\Omega_{ij} \equiv \{\omega : e_{ij}(\omega) > w_i/D_j\} \quad \text{such that} \quad e_{ij}(\omega) \equiv r_{ij}(\omega) \bar{\Pi}_j^{-1} \left(r_{ij}(\omega) f_{ij}(\omega)\right). \tag{OA.61}$$

[43] This is a mild restriction that arises from assumptions about the second derivative of the demand function.

Extensive and Intensive Margins of Firm Exports

As in Section 2.2, we consider the distribution of $(r_{ij}(\omega), e_{ij}(\omega))$ for each origin i and destination j. We now explicitly introduce shifters of entry and revenue potentials:

$$r_{ij}(\omega) \sim G^r_{ij}\left(r/\bar{r}_{ij}|e\right) \qquad \text{and} \qquad e_{ij}(\omega) \sim G^e_{ij}(e/\bar{r}_{ij}), \tag{OA.62}$$

where G^e_{ij} satisfies Assumption 1.

Extensive margin of firm-level exports. The entry condition in (OA.61) implies that $1 - n_{ij} = Pr(e_{ij}(\omega) < w_i/D_j) = G^e_{ij}(d_{ij})$ with $d_{ij} \equiv w_i/D_j\bar{r}_{ij}$. Let us define again the extensive margin elasticity function as $\epsilon_{ij}(n) \equiv (G^e_{ij})^{-1}(1-n)$ such that $\epsilon_{ij}(n)$ is strictly decreasing, $\epsilon_{ij}(1) = 0$, and $\lim_{n\to 0}\epsilon_{ij}(n) = \infty$. Thus,

$$\ln \epsilon_{ij}(n_{ij}) = -\ln \bar{r}_{ij} + \ln w_i - \ln D_j. \tag{OA.63}$$

Intensive margin of firm-level exports. Given the profit maximization problem above, average firm exports are given by $\bar{x}_{ij} = D_j\mathbb{E}[\mathcal{R}_j(w_i/D_{ij}r_{ij}(\omega))|\omega \in \Omega_{ij}]$. The entry decision in (OA.61) implies that

$$\bar{x}_{ij} = \frac{D_j}{n_{ij}}\int_{d_{ij}}^{\infty}\mathbb{E}\left[\mathcal{R}_j(d_{ij}/r)|e\right]dG^e_{ij}(e)$$

with $\mathbb{E}\left[\mathcal{R}_j(d_{ij}/r)|e\right] \equiv \int \mathcal{R}_j(d_{ij}/r)dG^r_{ij}(r|e)$. Let us define $\tilde{\rho}_{ij}(d) \equiv \int_d^\infty \mathbb{E}\left[\mathcal{R}_j(d/r)|e\right]dG^e_{ij}(e)$. Since $\tilde{\rho}'_{ij}(d) < 0$, $\lim_{d\to 0}\tilde{\rho}_{ij}(d) = \infty$ and $\lim_{d\to\infty}\tilde{\rho}_{ij}(d) = 0$, $\tilde{\rho}_{ij}(d)$ is invertible and we can define $\rho_{ij}(x) \equiv \tilde{\rho}^{-1}_{ij}(x)$ such that

$$\ln \rho_{ij}(\bar{x}_{ij}n_{ij}/D_j) = -\ln \bar{r}_{ij} + \ln w_i - \ln D_j. \tag{OA.64}$$

We can now extend Proposition 1 for our setting with non-CES demand of the form in equation (OA.54).

Proposition 1 (non-CES demand)**.** *Consider the monopolistic competition model with non-CES demand in the environment of Appendix Section A.3.1 under (OA.62). Then, for any origin i and destination j, the exporter firm share, n_{ij}, and the average firm exports, $\bar{x}_{ij}$, are given by equations (OA.63) and (OA.64), which depend on country-level endogenous variables, exogenous bilateral shifters, and two elasticity functions of the exporter firm share $n \in [0,1]$, $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$.*

Intensive margin of firm-level exports across percentiles of the distribution of firm exports. We now characterize an expression for percentile π of the distribution of firm-level exports from i to j. We start by deriving the distribution of firm-level exports from i to j:

$$\begin{aligned}
G^R_{ij}(\bar{R}) &\equiv Pr(R_{ij}(\omega) < \bar{R}|\omega \in \Omega_{ij}) \\
&= \tfrac{1}{n_{ij}}Pr(\mathcal{R}_j(w_i/D_jr_{ij}(\omega)) < \bar{R}/D_j,\ e(\omega) > d_{ij}) \\
&= \tfrac{1}{n_{ij}}Pr(w_i/D_jr_{ij}(\omega) > \mathcal{R}_j^{-1}\left(\bar{R}/D_j\right),\ e(\omega) > d_{ij}) \\
&= \tfrac{1}{n_{ij}}\textstyle\int_{d_{ij}}^\infty H^r_{ij}\left(d_{ij}/\mathcal{R}_j^{-1}\left(\bar{R}/D_j\right)|e\right)dG^e_{ij}(e) \\
&= \tfrac{1}{n_{ij}}\textstyle\int_0^{n_{ij}} G^r_{ij}\left(\epsilon_{ij}(n_{ij})/\mathcal{R}_j^{-1}\left(\bar{R}/D_j\right)|\epsilon_{ij}(n)\right)dn
\end{aligned}$$

where the second row uses (OA.57), the third row uses $\mathcal{R}_j(.)$ invertible with $\mathcal{R}_j' < 0$, the fourth row uses (OA.62), and the last row uses change of variables $n = 1 - G_{ij}^e(e)$.

We now define $F_{ij}(R|n) \equiv \frac{1}{n}\int_0^n G_{ij}^r\left(R|\epsilon_{ij}(n')\right)dn'$. Note that $F_{ij}(.|n)$ is invertible for any n since $\partial F_{ij}(R|n)/\partial R > 0$, $F_{ij}(0|n) = 0$ and $\lim_{R\to\infty} F_{ij}(R|n) = 1$. The distribution of firm-level exports from i to j can be written as

$$G_{ij}^R(\bar{R}) = F_{ij}\left(\epsilon_{ij}(n_{ij})/\mathcal{R}_j^{-1}\left(\bar{R}/D_j\right)|n_{ij}\right).$$

We denote the revenue of firms in percentile π of the distribution of firm-level exports from i to j as x_{ij}^π, which is implicitly given by $\pi = G_{ij}^R(x_{ij}^\pi)$. Since $F_{ij}(.|n)$ is invertible for any n, we have that

$$\epsilon_{ij}(n_{ij})/\mathcal{R}_j^{-1}\left(x_{ij}^\pi/D_j\right) = F_{ij}^{-1}(\pi|n_{ij})$$

which implies that

$$x_{ij}^\pi = D_j\rho_{ij}^\pi(n_{ij}) \quad \text{with} \quad \rho_{ij}^\pi(n) \equiv \mathcal{R}_j\left(\epsilon_{ij}(n)/F_{ij}^{-1}(\pi|n)\right). \tag{OA.65}$$

We note that, by definition, when we know the functions $\rho_{ij}^\pi(n)$, $\epsilon_{ij}(n)$ and $\mathcal{R}_j(.)$, we can define the following function of π: $\tilde{F}_{ij}(\pi|n) \equiv \epsilon_{ij}(n)/\mathcal{R}_j^{-1}\left(\rho_{ij}^\pi(n)\right)$. Note that $\rho_{ij}^\pi(n)$ is increasing in π given n, which allows us to write $F_{ij}(R|n) = \tilde{F}_{ij}^{-1}(R|n)$. Thus, since $nF_{ij}(R|n) = \int_0^n G_{ij}^r\left(R|\epsilon_{ij}(n')\right)dn'$ by definition, $n\tilde{F}_{ij}^{-1}(R|n) = \int_0^n G_{ij}^r\left(R|\epsilon_{ij}(n')\right)dn'$ and

$$\frac{\partial\left[n\tilde{F}_{ij}^{-1}(R|n)\right]}{\partial n} = G_{ij}^r\left(R|\epsilon_{ij}(n)\right). \tag{OA.66}$$

Sufficient Statistics in General Equilibrium

We now outline the conditions that determine $\{w_i, D_i, N_i\}$ in general equilibrium. As in the baseline model, budget balance and labor market clearing are given by

$$\sum_i N_i n_{ij}\bar{x}_{ij} = w_j\bar{L}_j + \bar{T}_j, \tag{OA.67}$$

$$\sum_j N_i n_{ij}\bar{x}_{ij} = w_i\bar{L}_i. \tag{OA.68}$$

Thus, these two conditions can be written in terms of $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$ in equations (OA.63)-(OA.64), respectively.

We now turn to the free entry condition, which is still given by

$$\sum_j n_{ij}(\bar{x}_{ij} - \bar{c}_{ij}) = w_i\bar{F}_i, \tag{OA.69}$$

where $\bar{c}_{ij} \equiv \mathbb{E}[C_{ij}(\omega)|\omega \in \Omega_{ij}]$ is the mean cost of firms from i selling in j.

As in our baseline model, expressions (OA.63) and (OA.64) characterize n_{ij} and $\bar{x}_{ij}$ using the elasticity functions for the extensive and intensive margins of firm-level exports. Thus, it suffices to characterize the mean cost $\bar{c}_{ij}$, which can be written in terms of variable and fixed costs:

$$\bar{c}_{ij} = \mathbb{E}\left[D_j\mathcal{C}_j\left(w_i/D_j r_{ij}(\omega)\right)|\omega \in \Omega_{ij}\right] + \mathbb{E}\left[w_i f_{ij}(\omega)|\omega \in \Omega_{ij}\right]$$

Consider first the expected variable cost of firms from i operating in j:

$$\begin{aligned}\mathbb{E}\left[D_j\mathcal{C}_j\left(\tfrac{w_i/D_j}{r_{ij}(\omega)}\right)|\omega\in\Omega_{ij}\right] &= D_j\tfrac{1}{n_{ij}}\int_{d_{ij}}^{\infty}\int\mathcal{C}_j\left(d_{ij}/r\right)dG_{ij}^r\left(r|e\right)dG_{ij}^e(e)\\ &= D_j\tfrac{1}{n_{ij}}\int_0^{n_{ij}}\int\mathcal{C}_j\left(\epsilon_{ij}(n_{ij})/r\right)dG_{ij}^r\left(r|\epsilon_{ij}(n)\right)dn\end{aligned}$$

where the first equality uses (OA.62) and the second equality the change of variables $n=1-G_{ij}^e(e)$.

Turning to the mean fixed cost, we have that

$$\begin{aligned}\mathbb{E}\left[w_if_{ij}(\omega)|\omega\in\Omega_{ij}\right] &= w_i\mathbb{E}\left[\bar{\Pi}_j\left(e_{ij}(\omega)/r_{ij}(\omega)\right)/r_{ij}(\omega)|\omega\in\Omega_{ij}\right]\\ &= D_jd_{ij}\tfrac{1}{n_{ij}}\int_{d_{ij}}^{\infty}\int\bar{\Pi}_j\left(e/r\right)/rdG_{ij}^r\left(r|e\right)dG_{ij}^e(e)\\ &= D_j\epsilon_{ij}(n_{ij})\tfrac{1}{n_{ij}}\int_0^{n_{ij}}\int\bar{\Pi}_j\left(\epsilon_{ij}(n)/r\right)/rdG_{ij}^r\left(r|\epsilon_{ij}(n)\right)dn\end{aligned}$$

where the first equality uses (OA.61), the second equality uses (OA.62) and the third equality uses the change of variables $n=1-G_{ij}^e(e)$.

Combining these expressions, we get that

$$\bar{c}_{ij}=D_j\kappa_{ij}(n_{ij}) \tag{OA.70}$$

with

$$\kappa_{ij}(n_{ij})\equiv\frac{1}{n_{ij}}\int_0^{n_{ij}}\int\left[\mathcal{C}_j\left(\epsilon_{ij}(n_{ij})/r\right)+\frac{\epsilon_{ij}(n_{ij})}{r}\bar{\Pi}_j\left(\epsilon_{ij}(n)/r\right)\right]dG_{ij}^r\left(r|\epsilon_{ij}(n)\right)dn. \tag{OA.71}$$

In order to compute $\kappa_{ij}(n)$ using (OA.71), one needs to know $G_{ij}^r(r|\epsilon_{ij}(n))$, $\epsilon_{ij}(n)$, $\mathcal{C}_j(.)$ and $\bar{\Pi}_j(.)$. Note that knowledge of the demand function $q_j(.)$ in (OA.54) implies that we can compute $\mathcal{C}_j(.)$, $\mathcal{R}_j(.)$ and $\bar{\Pi}_j(.)$ using (OA.57)–(OA.59). Thus, it only remains to show how we can recover $G_{ij}^r(r|\epsilon_{ij}(n))$. We consider two cases. First, without dispersion in fixed costs, there is a one-to-one mapping between $r_{ij}(\omega)$ and $e_{ij}(\omega)$, given the definition in (OA.61). This implies that $G_{ij}^r(r|\epsilon_{ij}(n))$ is degenerate at a known value determined by $\epsilon_{ij}(n)$. Second, when there is dispersion in fixed costs, expression (OA.66) yields $G_{ij}^r(r|\epsilon_{ij}(n))$ from $\rho_{ij}^\pi(n)$ and $\epsilon_{ij}(n)$. Thus, in this case, knowledge of $q_j(.)$, $\rho_{ij}^\pi(n)$ and $\epsilon_{ij}(n)$ implies that we can compute $\kappa_{ij}(n)$ using (OA.71).

We can now extent our proposition outlining the sufficient statistics for computing aggregate variables in general equilibrium.

Proposition 2 (non-CES demand)**.** *Consider the monopolistic competition model with the demand function in (OA.54) described in the environment of Appendix Section A.3.1 under (OA.62). Assume knowledge of the exogenous fundamentals $\{\bar{r}_{ij},\bar{L}_i,\bar{T}_i,\bar{F}_i\}$, the demand function in (OA.54), and the elasticity functions $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$. Then:*

a. For a given $\kappa_{ij}(n)$, the equilibrium vector $\{D_i,N_i,w_i\}$ solves the system of equations (OA.67)–(OA.69) with n_{ij}, $\bar{x}_{ij}$ and $\bar{c}_{ij}$ respectively given by (OA.63), (OA.64), and (OA.70).

b. The function $\kappa_{ij}(n)$ is identified (i) from $\epsilon_{ij}(n)$ without fixed cost dispersion and (ii) from $\epsilon_{ij}(n)$ and $\rho_{ij}^\pi(n)$ with fixed cost dispersion.

A.3.2 Multi-Sector, Multi-Factor Heterogeneous Firm Model with Input-Output Links and Import Tariffs

In this section, we extend our baseline framework to allow for firm heterogeneity in a model with multiple sectors, multiple factors of production, input-output linkages, and import tariffs. Our specification of the model can be seen as a generalization of the formulation in Costinot and Rodriguez-Clare (2013).

Environment

The world economy is constituted of countries with multiple sectors indexed by s. Each country has a representative household that inelastically supplies $\bar{L}_i^v$ units of multiple factors of production indexed by v.

Preferences. The representative household in country j has CES preferences over the composite good of multiple sectors, $s = 1, ...S$:

$$U_j = \left[\sum_s \gamma_j^s \left(Q_j^s\right)^{\frac{\lambda_j - 1}{\lambda_j}} \right]^{\frac{\lambda_j - 1}{\lambda_j}}.$$

Given the price of the sectoral composite goods, the share of spending on sector s is

$$c_j^s = \gamma_j^s \left(\frac{P_j^s}{P_j} \right)^{1-\lambda_j} \tag{OA.72}$$

where the consumption price index is

$$P_j = \left[\sum_s \gamma_j^s \left(P_j^s\right)^{1-\lambda_j} \right]^{\frac{1}{1-\lambda_j}}. \tag{OA.73}$$

Sectoral final composite good. In each sector s of country j, there is a perfectly competitive market for a non-tradable final good whose production uses different varieties of the tradable varieties $\omega \in \Omega^s$ in sector s:

$$Q_j^s = \left(\sum_i \int_{\Omega_{ij}^s} \left(\bar{b}_{ij}^s b_{ij}^s(\omega)\right)^{\frac{1}{\sigma^s}} \left(q_{ij}^s(\omega)\right)^{\frac{\sigma^s - 1}{\sigma^s}} \, d\omega \right)^{\frac{\sigma^s}{\sigma^s - 1}}$$

where $\sigma^s > 1$ and Ω_{ij}^s is the set of sector s's varieties of intermediate goods produced in country i available in country j.

The demand of country j by variety ω of sector s in country i is

$$q_{ij}^s(\omega) = \left(\bar{b}_{ij}^s b_{ij}^s(\omega)\right) \left(\frac{p_{ij}^s(\omega)}{P_j^s} \right)^{-\sigma^s} \frac{E_j^s}{P_j^s}$$

where E_j^s is the total spending of country j in sector s.

Because the market for the composite sectoral good is competitive, its price is the CES price index of intermediate inputs:

$$\left(P_j^s\right)^{1-\sigma^s} = \sum_i \int_{\Omega_{ij}^s} \left(\bar{b}_{ij}^s b_{ij}^s(\omega)\right) \left(p_{ij}^s(\omega)\right)^{1-\sigma^s} \, d\omega. \tag{OA.74}$$

Sectoral intermediate good. In sector s of country i, there is a representative competitive firm that produces a non-traded sectoral intermediate good using different factors and the non-traded composite final good of different sectors. The production function is

$$q_i^s = \left[\alpha_i^s \left(L_i^s\right)^{\frac{\mu_i^s-1}{\mu_i^s}} + \left(1-\alpha_i^s\right)\left(M_i^s\right)^{\frac{\mu_i^s-1}{\mu_i^s}}\right]^{\frac{\mu_i^s}{\mu_i^s-1}},$$

where

$$L_i^s = \left[\sum_v \beta_i^{s,v}\left(L_i^{s,v}\right)^{\frac{\eta_i^s-1}{\eta_i^s}}\right]^{\frac{\eta_i^s}{\eta_i^s-1}} \quad \text{and} \quad M_i^s = \left[\sum_k \theta_i^{ks}\left(Q_i^k\right)^{\frac{\kappa_i^s-1}{\kappa_i^s}}\right]^{\frac{\kappa_i^s-1}{\kappa_i^s}}.$$

Zero profit implies that the price of the sectoral intermediate good is

$$p_i^s = \left[\alpha_i^s \left(W_i^s\right)^{1-\mu_i^s} + \left(1-\alpha_i^s\right)\left(J_i^s\right)^{1-\mu_i^s}\right]^{\frac{1}{1-\mu_i^s}}, \tag{OA.75}$$

where

$$W_i^s = \left[\sum_v \beta_i^{s,v}\left(w_i^v\right)^{1-\eta_i^s}\right]^{\frac{1}{1-\eta_i^s}} \quad \text{and} \quad J_i^s = \left[\sum_k \theta_i^{ks}\left(P_i^k\right)^{1-\kappa_i^s}\right]^{\frac{1}{1-\kappa_i^s}}. \tag{OA.76}$$

The share of total production cost in sector s spent on factor f and input k are given by

$$l_i^{s,v} = \beta_i^{s,v}\left(\frac{w_i^v}{W_i^s}\right)^{1-\eta_i^s}\alpha_i^s\left(\frac{W_i^s}{p_i^s}\right)^{1-\mu_i^s} \quad \text{and} \quad m_i^{ks} = \theta_i^{ks}\left(\frac{P_i^k}{J_i^s}\right)^{1-\kappa_i^s}\left(1-\alpha_i^s\right)\left(\frac{J_i^s}{p_i^s}\right)^{1-\mu_i^s}. \tag{OA.77}$$

Production of traded intermediate varieties ω. Assume that sector s has a continuum of monopolistic firms that produce output using only a non-tradable input q_i^s. We also assume that country j imposes an ad-valorem tariff of t_{ij}^s on goods of sector s from country i. In order to sell q in market j, variety ω of country i faces a cost function given by

$$C_{ij}(\omega, q) = p_i^s(1+t_{ij}^s)\frac{\bar{\tau}_{ij}^s}{\bar{a}_i^s}\frac{\tau_{ij}^s(\omega)}{a_i^s(\omega)}q + p_i^s\bar{f}_{ij}^s f_{ij}^s(\omega)$$

where p_i^s is the price of the non-tradable input q_i^s in country i.

Entry and Revenue Potentials. We now define the two variables that determine firm-level revenue and entry in each sector. Given this production technology, the optimal price is $p_{ij}^s(\omega) = \frac{\sigma^s}{\sigma^s-1}\frac{\tau_{ij}^s(\omega)}{a_i^s(\omega)}\frac{(1+t_{ij}^s)\bar{\tau}_{ij}^s}{\bar{a}_i^s}p_i^s$ and the associated revenue is

$$R_{ij}^s(\omega) = \left(\left(p_j^s\right)^{1-\sigma^s}\left(P_j^s\right)^{\sigma^s-1}E_j^s\right)\bar{r}_{ij}^s r_{ij}^s(\omega) \tag{OA.78}$$

where

$$r_{ij}^s(\omega) \equiv b_{ij}^s(\omega)\left(\frac{\tau_{ij}^s(\omega)}{a_i^s(\omega)}\right)^{1-\sigma^s} \quad \text{and} \quad \bar{r}_{ij}^s \equiv \bar{b}_{ij}^s\left(\frac{\sigma^s}{\sigma^s-1}\frac{(1+t_{ij}^s)\bar{\tau}_{ij}^s}{\bar{a}_i^s}\right)^{1-\sigma^s}. \tag{OA.79}$$

Firm ω of country i chooses to enter market j if, and only if, $\pi_{ij}^s(\omega) = (1/\sigma^s)R_{ij}^s(\omega) - p_i^s\bar{f}_{ij}^s f_{ij}^s(\omega) \geq 0$. This condition determines the set of firms from country i that operate in sector s of country j:

$$\Omega_{ij}^s = \{\omega : e_{ij}^s(\omega) \geq e_{ij}^{s,*}\} \tag{OA.80}$$

where

$$e_{ij}^s(\omega) \equiv \frac{r_{ij}^s(\omega)}{f_{ij}^s(\omega)}, \quad \text{and} \quad e_{ij}^{s,*} \equiv \frac{\bar{r}_{ij}^s}{\sigma^s \bar{f}_{ij}^s} \left[\left(\frac{p_i^s}{P_j^s} \right)^{\sigma^s} \frac{P_j^s}{E_j^s} \right]. \tag{OA.81}$$

Extensive and Intensive Margins of Firm Exports

We now turn to the characterization of entry and sales in each sector. We consider the distribution of $(r_{ij}^s(\omega), e_{ij}^s(\omega))$:

$$r_{ij}^s(\omega) \sim G_{ij}^{r,s}(r|e) \qquad \text{and} \qquad e_{ij}^s(\omega) \sim G_{ij}^{e,s}(e), \tag{OA.82}$$

where $G_{ij}^{e,s}$ has full support in $\mathbb{R}_+$.

Extensive margin of firm-level exports. The share of firms in sector s of country i serving market j is $n_{ij}^s = Pr\left[\omega \in \Omega_{ij}^s\right]$. We define $\epsilon_{ij}^s(n) \equiv \left(G_{ij}^{e,s}\right)^{-1}(1-n)$ such that

$$\ln \epsilon_{ij}^s(n_{ij}^s) = \ln \sigma^s \bar{f}_{ij}^s / \bar{r}_{ij}^s + \ln \left(p_i^s\right)^{\sigma^s} - \ln E_j^s \left(P_j^s\right)^{\sigma^s - 1}. \tag{OA.83}$$

Thus, we obtain a sector-specific version of the relationship between the function of the share of firms from i selling in j and the linear combination of exogenous bilateral trade shifters and endogenous outcomes in the origin and destination markets.

Intensive margin of firm-level exports. The average revenue of firms from country i in country j is $\bar{x}_{ij}^s \equiv \mathbb{E}\left[R_{ij}^s(\omega)|\omega \in \Omega_{ij}^s\right]$. Define the mean revenue potential of exporters when $n\%$ of $i's$ firms in sector s export to j as $\rho_{ij}^s(n) \equiv \frac{1}{n}\int_0^n \mathbb{E}[r|e = \epsilon_{ij}^s(n)]\, dn$. The change of variable $n = 1 - G_{ij}^{e,s}(e)$ implies that

$$\ln \bar{x}_{ij}^s - \ln \rho_{ij}^s(n_{ij}^s) = \ln \left(\bar{r}_{ij}^s\right) + \ln \left(p_i^s\right)^{1-\sigma^s} + \ln E_j^s \left(P_j^s\right)^{\sigma^s - 1}. \tag{OA.84}$$

Thus, we obtain a sector-specific version of the relationship between the composition-adjusted per-firm sales and a linear combination of exogenous bilateral revenue shifters and endogenous outcomes in the origin and destination markets.

We can now extend Proposition 1.

Proposition 1 (multi-sector, multi-factor, import tariffs)**.** *Consider the monopolistic competition model with multiple factors, multiple sectors, input-output linkages and import tariffs described in the environment of Appendix Section A.3.2 under (OA.82). Then, for any origin i, destination j and sector s, the exporter firm share, n_{ij}^s, and the average firm exports, $\bar{x}_{ij}^s$, are given by equations (OA.83) and (OA.84), which are separable on country-level endogenous variables, exogenous bilateral shifters, and two elasticity functions of the exporter firm share $n \in [0,1]$, $\epsilon_{ij}^s(n)$ and $\rho_{ij}^s(n)$.*

Sufficient Statistics in General Equilibrium

We now describe the conditions establishing free entry, budget balance and factor market clearing.

Firms in sector s of country i can create a new variety by spending $\bar{F}_i^s$ units of the non-tradable sectoral input q_i^s. In equilibrium, free entry implies that N_i^s firms pay the fixed cost of entry in exchange for an ex-ante expected profit of zero, $\sum_j \mathbb{E}\left[\max\left\{\pi^s_{ij}(\omega);\ 0\right\}\right] = p^s_i \bar{F}^s_i$. Following the same steps described in Section 2.3, we can show that

$$\frac{1}{\sigma^s}\sum_j \frac{n^s_{ij}\bar{x}^s_{ij}}{1+t^s_{ij}} = p^s_i \bar{F}_i + p^s_i \sum_j \bar{f}^s_{ij}\int_0^{n^s_{ij}} \frac{\rho^s_{ij}(n)}{\epsilon^s_{ij}(n)}(1+\varrho^s_{ij}(n))dn, \tag{OA.85}$$

with $\varrho^s_{ij}(n) = \partial \ln \rho^s_{ij}(n)/\partial \ln n$.

Thus, the free entry condition can be written as a function of the elasticity functions $\rho^s_{ij}(n)$ and $\epsilon^s_{ij}(n)$ (recall that we argued above that this is true also for $\bar{x}^s_{ij}$ and n^s_{ij}).

We now turn to the budget balance condition that determines the sectoral price index P^s_j in (OA.74). Using the expression for $p^s_{ij}(\omega)$ and (OA.74), we have that $(P^s_j)^{1-\sigma^s} = \sum_i \bar{r}^s_{ij}\left(p^s_i\right)^{1-\sigma^s}\int_{\Omega^s_{ij}} r^s_{ij}(\omega)\ d\omega$. Since $\int_{\Omega^s_{ij}} r^s_{ij}(\omega)\ d\omega = N^s_i Pr[\omega \in \Omega^s_{ij}]\mathbb{E}[r|\omega \in \Omega^s_{ij}] = N^s_i n^s_{ij}\rho^s_{ij}(n^s_{ij})$, we can write P^s_j as

$$(P^s_j)^{1-\sigma^s} = \sum_i \bar{r}^s_{ij}\left(p^s_i\right)^{1-\sigma^s}\rho^s_{ij}(n^s_{ij})n^s_{ij}N^s_i. \tag{OA.86}$$

We again follow Dekle et al. (2008) by allowing for a set of exogenous transfers. Thus, the spending on goods of sector s by country i is

$$E^s_i = c^s_i\left(\sum_v w^v_i \bar{L}^v_i + \bar{T}_i + R^t_i\right) + \sum_k m^{sk}_i \sum_j \frac{N^k_i n^k_{ij}\bar{x}^k_{ij}}{1+t^k_{ij}}, \tag{OA.87}$$

with m^{sk}_i the intermediate spending share given by (OA.77), and R^t_i is the import tariff revenue that is given by

$$R^t_i = \sum_j\sum_s \frac{t^s_{ji}}{1+t^s_{ji}} N^s_j n^s_{ji}\bar{x}^s_{ji}.$$

Finally, the market clearing conditions for factor v in country i is

$$w^v_i \bar{L}^v_i = \sum_s l^{s,v}_i \sum_j \frac{N^s_i n^s_{ij}\bar{x}^s_{ij}}{1+t^s_{ij}}, \tag{OA.88}$$

with $l^{s,v}_i$ given by (OA.77).

Thus, because the conditions above only depend $\bar{x}^s_{ij}$ and n^s_{ij}, they can also be written as a function of the elasticity functions $\rho^s_{ij}(n)$ and $\epsilon^s_{ij}(n)$.

The following proposition summarizes the conditions that determine aggregate variables in general equilibrium.

Proposition 2 (multi-sector, multi-factor, import tariffs)**.** *Consider the monopolistic competition model with multiple factors, multiple sectors, input-output linkages and import tariffs described in the environment of Appendix Section A.3.2 under (OA.82). Assume knowledge of the exogenous fundamentals* $\{\bar{r}^s_{ij}, \bar{f}^s_{ij}, t^s_{ij}, \bar{F}^s_i, \gamma^s_i, \alpha^s_i, \theta^{ks}_i, \beta^{s,v}_i, \bar{L}^v_i, \bar{T}_i\}$*, the elasticity of substitution in consumption and production* $\{\sigma^s, \lambda_i, \mu^s_i, \eta^s_i, \kappa^s_i\}$*, and the elasticity functions* $\epsilon^s_{ij}(n)$ *and* $\rho^s_{ij}(n)$*. Then, the equilibrium vector* $\{N^s_i, P^s_i, E^s_i, w^v_i\}$ *solves the system of equations (OA.85)-(OA.88) with* n^s_{ij} *and* $\bar{x}^s_{ij}$ *given by (OA.83) and (OA.84), and the sectoral input price* p^s_i *given by (OA.75)-(OA.76).*

A.3.3 Allowing for Zero Bilateral Trade

In this section, we extend our baseline framework to allow for zero trade flows between two countries. We do so by considering a weaker version of the full support requirement for entry potentials in Assumption 1.

Environment

Consider the same environment described in Section 2.1.

Extensive and Intensive Margins of Firm Exports

As in our baseline, we consider the distribution of $(r_{ij}(\omega), e_{ij}(\omega))$:

$$r_{ij}(\omega) \sim G^r_{ij}\left(r|e\right), \qquad \text{and} \qquad e_{ij}(\omega) \sim G^e_{ij}(e). \tag{OA.89}$$

We now however consider a weaker version of Assumption 1.

Assumption 1': *$G^e_{ij}(e)$ is continuous and strictly increasing in $[0, \bar{e}_{ij}]$ with $\bar{e}_{ij} < \infty$.*

This assumption specifies that the distribution of entry potentials has full support in a bounded interval. This allows for zero trade flows, as in Helpman et al. (2008). We now use this assumption to derive the expressions for the extensive and intensive margins of firm-level exports.

Extensive margin of firm-level exports. Recall that $n_{ij} \equiv Pr[\omega \in \Omega_{ij}]$ where Ω_{ij} is given by (6). It implies that

$$n_{ij} = \begin{cases} 1 - G^e_{ij}(e^*_{ij}) & \text{if} \quad e^*_{ij} \leq \bar{e}_{ij} \\ 0 & \text{if} \quad e^*_{ij} > \bar{e}_{ij} \end{cases}$$

with e^*_{ij} defined in (7).

Let us now define

$$\tilde{\epsilon}_{ij}(n) \equiv \begin{cases} \left(G^e_{ij}\right)^{-1}(1-n) & \text{if} \quad n > 0 \\ \bar{e}_{ij} & \text{if} \quad n = 0 \end{cases}.$$

The definition of $\tilde{\epsilon}_{ij}(n)$ and the expression for n_{ij} above imply that that $\tilde{\epsilon}_{ij}(n_{ij}) = \min\left\{e^*_{ij}, \bar{e}_{ij}\right\}$. Thus, by defining $\epsilon_{ij}(n) \equiv \tilde{\epsilon}_{ij}(n_{ij})/\bar{e}_{ij}$, we get that

$$\ln \epsilon_{ij}(n_{ij}) = \min\left\{-\ln(\sigma \bar{f}_{ij}\bar{e}_{ij}/\bar{r}_{ij}) + \ln\left(w^\sigma_i\right) - \ln\left(E_j P^{\sigma-1}_j\right), 0\right\}. \tag{OA.90}$$

Intensive margin of firm-level exports. Conditional on $n_{ij} > 0$, we now compute the average revenue in j:

$$\bar{x}_{ij} = \bar{r}_{ij}\left[\left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j\right]\frac{1}{n_{ij}}\int_{e^*_{ij}}^{\bar{e}_{ij}} \mathbb{E}[r|e]\; dG^e_{ij}(e).$$

We again consider the transformation $n = 1 - G^e_{ij}(e)$ such that $e = \epsilon_{ij}(n)$ and $dG^e_{ij}(e) = -dn$. Thus,

$$\ln \bar{x}_{ij} - \ln \rho_{ij}(n_{ij}) = \ln \bar{r}_{ij} + \ln\left(w^{1-\sigma}_i\right) + \ln\left(E_j P^{\sigma-1}_j\right), \tag{OA.91}$$

where we normalize $\rho_{ij}(0) = 0$.

Proposition 1 (zero trade flows)**.** *Consider the monopolistic competition model with CES demand described in the environment of Appendix Section A.3.3 under Assumption 1'. Then, for any origin i and destination j, the exporter firm share,* n_{ij}*, and the average firm exports,* $\bar{x}_{ij}$*, are given by equations (OA.90) and (OA.91), which are separable on country-level endogenous variables, exogenous bilateral shifters, and two elasticity functions of the exporter firm share* $n \in [0,1]$*,* $\epsilon_{ij}(n)$ *and* $\rho_{ij}(n)$*.*

Sufficient Statistics in General Equilibrium

The modified assumption on the support of entry potentials does not affect any of the derivations for the conditions determining free entry, budget balance, and labor market clearing. Thus, we can immediately state the extension of Proposition 2 using the modified expression for the extensive margin of firm exports in (OA.90).

Proposition 2 (zero trade flows)**.** *Consider the monopolistic competition model with CES demand in the environment of Appendix Section A.3.3 under Assumption 1'. Assume knowledge of the exogenous fundamentals* $\{\bar{r}_{ij}, \bar{f}_{ij}, \bar{e}_{ij}, \bar{L}_i, \bar{T}_i, \bar{F}_i\}$*, the demand elasticity of substitution* σ*, and the elasticity functions* $\epsilon_{ij}(n)$ *and* $\rho_{ij}(n)$*. Then, the equilibrium vector* $\{P_i, N_i, w_i\}$ *solves the system of equations (17)-(19) with* n_{ij} *and* $\bar{x}_{ij}$ *given by (OA.90) and (OA.91).*

A.3.4 Multi-product Firms

In this section, we extend our framework to incorporate multi-product firms.

Environment

Preferences. We maintain the assumption that each country j has a representative household that inelastically supplies $\bar{L}_j$ units of labor. The demand for variety ω from country i is

$$q_{ij}(\omega) = \left(\frac{p_{ij}(\omega)}{P_j}\right)^{-\sigma} \frac{E_j}{P_j}, \tag{OA.92}$$

where, in market j, E_j is the total spending, $p_{ij}(\omega)$ is the price of variety ω of country i, and P_j is the CES price index,

$$P_j^{1-\sigma} = \sum_i \int_{\Omega_{ij}^v} (p_{ij}(\omega))^{1-\sigma} \, d\omega, \tag{OA.93}$$

and Ω_{ij}^v is the set of varieties produced in country i that are sold in country j.

Technology. We consider a monopolistic competitive environment. Each firm η can choose how many varieties to sell in each market. In order to operate in market j, the firm must pay a fixed entry cost $w_i \bar{f}_{ij} f_{ij}(\eta)$. Conditional on entry, selling N varieties entails a labor cost of $w_i \frac{1}{1+1/\alpha} N^{1+1/\alpha}$. For every variety, the firm then has a unit production cost of $w_i \frac{\tau_{ij}(\eta)}{a_i(\eta)} \frac{\bar{\tau}_{ij}}{\bar{a}_i}$.

Entry and Revenue Potentials. For each variety ω of firm η from country i, the optimal price in market j is $p_{ij}(\omega) = \frac{\sigma}{\sigma-1}\frac{\bar{\tau}_{ij} w_i}{\bar{a}_i}\frac{\tau_{ij}(\eta)}{a_i(\eta)}$ with an associated revenue of

$$R_{ij}^N(\eta) = \bar{r}_{ij}^N r_{ij}^N(\eta) \left[\left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j \right], \tag{OA.94}$$

where

$$r_{ij}^N(\eta) \equiv \left(\frac{\tau_{ij}(\eta)}{a_i(\eta)}\right)^{1-\sigma} \quad \text{and} \quad \bar{r}_{ij}^N \equiv \left(\frac{\sigma}{\sigma-1}\frac{\bar{\tau}_{ij}}{\bar{a}_i}\right)^{1-\sigma}. \tag{OA.95}$$

The firm then decides how many varieties to sell by solving the following problem:

$$\max_N \frac{1}{\sigma} R_{ij}^N(\eta) N - w_i \frac{1}{1+1/\alpha} N^{1+1/\alpha},$$

which implies that

$$N_{ij}(\eta) = \left(\frac{1}{\sigma}\frac{R_{ij}^N(\eta)}{w_i}\right)^{\alpha}. \tag{OA.96}$$

Thus, firm sales are

$$R_{ij}(\eta) = N_{ij}(\eta) R_{ij}^N(\eta) = \frac{1}{\sigma^\alpha w_i^\alpha}\left(\bar{r}_{ij}^N r_{ij}^N(\eta)\right)^{1+\alpha} \left[\left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j \right]^{1+\alpha}.$$

To simplify the notation, conditional on entering market j, the sales of firm η can be written as

$$R_{ij}(\eta) = \bar{r}_{ij} r_{ij}(\eta) w_i^{1-(1+\alpha)\sigma} \left(E_j P_j^{\sigma-1}\right)^{1+\alpha} \tag{OA.97}$$

$$r_{ij}(\eta) \equiv \left(r_{ij}^N(\eta)\right)^{1+\alpha} \quad \text{and} \quad \bar{r}_{ij} \equiv \frac{1}{\sigma^\alpha}\left(\bar{r}_{ij}^N\right)^{1+\alpha}. \tag{OA.98}$$

Conditional on entering market j, the firm's profit in that market is

$$\begin{aligned} \pi_{ij}(\eta) \quad = \quad & N_{ij}(\eta)\tfrac{1}{\sigma} R_{ij}^N(\eta) - w_i \tfrac{1}{1+1/\alpha} N_{ij}(\eta)^{1+1/\alpha} - w_i \bar{f}_{ij} f_{ij}(\eta) \\ & \left(\tfrac{1}{\sigma}\tfrac{R_{ij}^N(\eta)}{w_i}\right)^{\alpha} \tfrac{1}{\sigma} R_{ij}^N(\eta) - w_i \tfrac{1}{1+1/\alpha}\left(\tfrac{1}{\sigma}\tfrac{R_{ij}^N(\eta)}{w_i}\right)^{1+\alpha} - w_i \bar{f}_{ij} f_{ij}(\eta) \\ & \tfrac{1}{(1+\alpha)\sigma}\tfrac{1}{\sigma^\alpha w_i^\alpha}\left(R_{ij}^N(\eta)\right)^{1+\alpha} - w_i \bar{f}_{ij} f_{ij}(\eta) \end{aligned}$$

and, therefore,

$$\pi_{ij}(\eta) = \frac{1}{(1+\alpha)\sigma} \bar{r}_{ij} r_{ij}(\eta) w_i^{1-(1+\alpha)\sigma} \left(E_j P_j^{\sigma-1}\right)^{1+\alpha} - w_i \bar{f}_{ij} f_{ij}(\eta). \tag{OA.99}$$

Firm η of i sells in j if, and only if profits are positive, $\pi_{ij}(\eta) \geq 0$. This yields the set of firms of country i operating in j, Ω_{ij}:

$$\Omega_{ij} = \{\eta : e_{ij}(\eta) \geq e_{ij}^*\} \tag{OA.100}$$

where

$$e_{ij}(\eta) \equiv \frac{r_{ij}(\eta)}{f_{ij}(\eta)}, \quad \text{and} \quad e_{ij}^* \equiv \frac{\bar{r}_{ij}}{(1+\alpha)\sigma \bar{f}_{ij}} \left[\frac{w_i^{(1+\alpha)\sigma}}{\left(E_j P_j^{\sigma-1}\right)^{1+\alpha}} \right]. \tag{OA.101}$$

Extensive and Intensive Margins of Firm Exports

We use the definitions of entry and revenue potentials to characterize firm-level entry and sales in different markets in general equilibrium. We assume that

$$r_{ij}(\eta) \sim G^{r}_{ij}(r|e) \qquad \text{and} \qquad e_{ij}(\eta) \sim G^{e}_{ij}(e), \tag{OA.102}$$

where G^e_{ij} satisfies Assumption 1.

Extensive margin of firm-level exports. The share of firms of country i serving market j is $n_{ij} = Pr\left[\eta \in \Omega_{ij}\right]$. Defining $\epsilon_{ij}(n) \equiv (G^e_{ij})^{-1}(1-n)$, equation (OA.100) yields

$$\ln \epsilon_{ij}(n_{ij}) = \ln\left((1+\alpha)\sigma \bar{f}_{ij}/\bar{r}_{ij}\right) + \ln\left(w_i^{(1+\alpha)\sigma}\right) - \ln\left(E_j P_j^{\sigma-1}\right)^{1+\alpha}. \tag{OA.103}$$

Intensive margin of firm-level exports. The average revenue of firms from country i in country j is $\bar{x}_{ij} \equiv \mathbb{E}\left[R_{ij}(\eta)|\eta \in \Omega_{ij}\right]$ where $R_{ij}(\eta)$ is given by (OA.97). Define the average revenue potential of exporters when $n\%$ of $i's$ firms in sector s export to j as $\rho_{ij}(n) \equiv \frac{1}{n}\int_0^n \mathbb{E}[r|e = \epsilon_{ij}(n')]\, dn'$ where $\mathbb{E}[r|e=\epsilon_{ij}(n)]$ is the average revenue potential in quantile n of the entry potential distribution. Using the transformation $n = 1 - G^e_{ij}(e)$ such that $e = \epsilon_{ij}(n)$ and $dG^e_{ij}(e) = -dn$, we can follow the same steps as in the baseline model to show that

$$\ln \bar{x}_{ij} - \ln \rho_{ij}(n_{ij}) = \ln\left(\bar{r}_{ij}\right) + \ln\left(w_i^{1-(1+\alpha)\sigma}\right) + \ln\left(E_j P_j^{\sigma-1}\right)^{1+\alpha}. \tag{OA.104}$$

Extensive margin of products per firm.
The average number of products among firms from i operating in market j is $N^v_{ij} = \mathbb{E}\left[N_{ij}(\eta)|\eta \in \Omega_{ij}\right]$. The expression for $N_{ij}(\eta)$ in (OA.96) implies that

$$N^v_{ij} = \frac{1}{\sigma^\alpha}\left(\bar{r}^N_{ij}\right)^\alpha w_i^{-\alpha\sigma}\left(E_j P_j^{\sigma-1}\right)^\alpha \mathbb{E}\left[(r^N_{ij}(\eta))^\alpha|\eta \in \Omega_{ij}\right]$$

and, since $\bar{r}_{ij} \equiv \frac{1}{\sigma^\alpha}\left(\bar{r}^N_{ij}\right)^{1+\alpha}$,

$$N^v_{ij} = \sigma^{-\frac{\alpha}{1+\alpha}} \bar{r}_{ij}^{\frac{\alpha}{1+\alpha}} w_i^{-\alpha\sigma}\left(E_j P_j^{\sigma-1}\right)^\alpha \mathbb{E}\left[(r_{ij}(\eta))^{\frac{\alpha}{1+\alpha}}|\eta \in \Omega_{ij}\right].$$

We consider a similar transformation as the one used above. Define $\rho^v_{ij}(n) \equiv \frac{1}{n}\int_0^n \mathbb{E}[r^{\frac{\alpha}{1+\alpha}}|e = \epsilon_{ij}(n)]\, dn$. Using the transformation $n = 1 - G^e_{ij}(e)$ such that $e = \epsilon_{ij}(n)$ and $dG^e_{ij}(e) = -dn$, we can follow the same steps as in the baseline model to show that

$$\ln N^v_{ij} - \ln \rho^v_{ij}(n_{ij}) = \frac{\alpha}{1+\alpha}\ln(\bar{r}_{ij}/\sigma) + \ln w_i^{-\alpha\sigma} + \ln\left(E_j P_j^{\sigma-1}\right)^\alpha. \tag{OA.105}$$

The elasticity of the average number of varieties per firm with respect to changes in bilateral revenue shifters is $\alpha/(1+\alpha)$, conditional on the composition control function, $\rho^v_{ij}(n_{ij})$, and the origin and destination fixed-effects.

We can now extend Proposition 1 for the model with multi-product firms.

Proposition 1 (multi-product firms)**.** *Consider the monopolistic competition model with CES demand described in the environment of Appendix Section A.3.4 under (OA.102). Then, for any origin i and* *destination j, the exporter firm share, n_{ij}, the average firm exports, $\bar{x}_{ij}$, and the average products per firm, N^v_{ij}, are respectively given by equations (OA.103), (OA.104) and (OA.105), which are separable on country-level endogenous variables, exogenous bilateral shifters, and three elasticity functions of the exporter firm share $n \in [0,1]$, $\epsilon_{ij}(n)$, $\rho_{ij}(n)$, and $\rho^v_{ij}(n)$.*

Sufficient Statistics in General Equilibrium

We now turn to the conditions determining aggregate variables in general equilibrium. We first consider the free entry condition for firms. As in the baseline, we assume that an entrant firm pays a fixed labor cost $\bar{F}_i$ to draw its type. In a free entry equilibrium, N_i firms pay the fixed cost of entry in exchange for an ex-ante expected profit of zero such that $\sum_j \mathbb{E}\left[\max\left\{\pi_{ij}(\eta);\ 0\right\}\right] = w_i \bar{F}_i$. The expected profit can be written as

$$\begin{aligned}
\mathbb{E}\left[\max\left\{\pi_{ij}(\omega);\ 0\right\}\right] &= \textstyle\sum_j Pr[\eta \in \Omega_{ij}]\mathbb{E}\left[\frac{1}{(1+\alpha)\sigma} R_{ij}(\eta) - w_i \bar{f}_{ij} f_{ij}(\eta) | \eta \in \Omega_{ij}\right] \\
&= \textstyle\sum_j n_{ij} \left(\frac{1}{(1+\alpha)\sigma}\bar{x}_{ij} - w_i \bar{f}_{ij}\mathbb{E}[r_{ij}(\eta)/e_{ij}(\eta)|\eta \in \Omega_{ij}]\right). \\
&= \textstyle\sum_j n_{ij} \left(\frac{1}{(1+\alpha)\sigma}\bar{x}_{ij} - w_i \bar{f}_{ij} \int_{e^*_{ij}}^{\infty} \frac{1}{e} E[r|e] \frac{dG^e(e)}{1-G^e(e^*_{ij})}\right) \\
&= \textstyle\sum_j \frac{1}{(1+\alpha)\sigma} n_{ij}\bar{x}_{ij} - w_i \bar{f}_{ij} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n)}(1+\varrho_{ij}(n))dn
\end{aligned}$$

where the first row uses (OA.97) and (OA.99), the second row uses (OA.101), third row uses (OA.100), and the fourth row uses the change of variables $n = 1 - G_{ij}(e)$ and the definition of $\rho_{ij}(.)$ and $\varrho_{ij}(.)$.

We can then write the free entry condition as

$$\frac{1}{(1+\alpha)\sigma}\sum_j n_{ij}\bar{x}_{ij} = w_i \bar{F}_i + w_i \sum_j \bar{f}_{ij} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n)}(1+\varrho_{ij}(n))dn. \tag{OA.106}$$

We then turn to the budget balance condition that determines the CES price index. Here, we use the fact that $p_{ij}(\omega) = \frac{\sigma}{\sigma-1}\frac{\bar{\tau}_{ij} w_i}{\bar{a}_i}\frac{\tau_{ij}(\eta)}{a_i(\eta)}$ for every variety ω of firm η to write directly the CES price index as

$$P_j^{1-\sigma} = \sum_i \int_{\Omega_{ij}} N_{ij}(\eta)\left(p_{ij}(\eta)\right)^{1-\sigma}\ d\eta.$$

Using the expression for $N_{ij}(\eta)$ in (OA.96) and the definitions in (OA.98), this expression can be written as

$$P_j^{1-\sigma} = \sum_i w_i^{1-\sigma}\frac{\bar{r}_{ij}}{w_i^{\alpha}}\left[\left(\frac{w_i}{P_j}\right)^{1-\sigma}\frac{E_j}{w_i}\right]^{\alpha}\int_{\Omega_{ij}} r_{ij}(\eta)\ d\eta.$$

Notice that $\int_{\Omega_{ij}} r_{ij}(\eta)\ d\eta = N_i Pr[\eta \in \Omega_{ij}]\mathbb{E}[r|\eta \in \Omega_{ij}] = N_i n_{ij}\rho_{ij}(n_{ij})$. This immediately yields

$$P_j^{1-\sigma} = \sum_i \bar{r}_{ij} w_i^{1-\sigma}\left[\left(\frac{w_i}{P_j}\right)^{1-\sigma}\frac{E_j}{w_i}\right]^{\alpha}\rho_{ij}(n_{ij})n_{ij}N_i,$$

and, therefore,

$$P_j^{1-\sigma} = \sum_i \bar{r}_{ij} w_i^{1-(1+\alpha)\sigma}\left(E_j P_j^{\sigma-1}\right)^{\alpha}\rho_{ij}(n_{ij})n_{ij}N_i. \tag{OA.107}$$

Finally, we again follow Dekle et al. (2008) by introducing exogenous international transfers, so that spending is

$$E_i = w_i \bar{L}_i + \bar{T}_i, \quad \sum_i \bar{T}_i = 0.$$

Since labor is the only factor of production, labor income in i equals the total revenue of firms from i:

$$w_i \bar{L}_i = \sum_j N_i n_{ij} \bar{x}_{ij}. \tag{OA.108}$$

We can now extend Proposition 2 for the model with multi-product firms.

Proposition 2 (multi-product firms)**.** *Consider the monopolistic competition model with CES demand described in the environment of Appendix Section A.3.4 under (OA.102). Assume knowledge of the exogenous fundamentals* $\{\bar{r}_{ij}, \bar{f}_{ij}, \bar{L}_i, \bar{T}_i, \bar{F}_i\}$*, the elasticity of supplying new varieties in a firm* α*, the demand elasticity of substitution* σ*, and the elasticity functions* $\epsilon_{ij}(n)$ *and* $\rho_{ij}(n)$*. Then, the equilibrium vector* $\{P_i, N_i, w_i\}$ *solves the system of equations (OA.106)-(OA.108) with* n_{ij} *and* $\bar{x}_{ij}$ *given by (OA.103) and (OA.104).*

B Empirical Appendix

This appendix complements Sections 4 and 5. Appendix B.1 provides further details about the simulations in Section 4.2. Appendix B.2 describes the procedure to construct the data used for estimation. Appendix B.3 reports additional results that complement our baseline estimates in Section 5.

B.1 Monte Carlo Simulations

B.1.1 Simulations

We consider three economies that differ with respect to the calibration of the trade elasticity functions. We assume that all markets have the same distribution of entry potentials, $G^e_{ij}(e) = G^e(e)$, such that (i) $G^e(e) = 1 - (e/\underline{e})^{-\alpha^e}$ for $\underline{e} = \exp(1)$ and $\alpha^e = 5/(\sigma - 1)$ in the economy with a constant trade elasticity, (ii) $G^e(e) = \Phi(\exp(e)/\nu^e)$ for $\Phi(.)$ the standard normal CDF and $\nu^e = 1.2$ in the economy with a decreasing trade elasticity, and (iii) $G^e(e) = 1 - (e/\underline{e})^{-\alpha^e}(\ln e)^{-\gamma^e}$ for $\underline{e} = \exp(1)$, $\alpha^e = 5/(\sigma - 1) - 1$ and $\gamma^e = 1.5$ for the economy with an increasing trade elasticity. For all economies, we set the conditional distribution of revenue potentials to be log-normal with mean zero and dispersion of $\nu^r = 1.2$. This implies that the distribution of log revenue across firms in each market has a normal distribution: $G^{\ln R}_{ij,b}(x) = \Phi((x - \ln \bar{R}_{ij,b})/\nu^r)$ where $\bar{R}_{ij,b} \equiv (w^{1-\sigma}_{i,b} P^{\sigma-1}_{j,b} E_{j,b})\bar{r}_{ij,b}$.

In addition, for all three economies, we set $\sigma = 3.2$ following Redding and Weinstein (2024) and do not allow for international transfers ($\bar{T}_i = 0$). We also assume that countries are identical with $\bar{L}_i = \bar{F}_i = 1$ and $\bar{f}_{ij} = \mu^f$. We then generate a random realization b of bilateral revenue shifters such that $\ln \bar{r}_{ij,b} = z_{ij,b} + \mu^r + \eta^r_{ij,b}$, where μ^r is a constant, and $z_{ij,b}$ and $\eta^r_{ij,b}$ are independently drawn from mean-zero normal distributions with standard deviations of ϕ^z and ϕ^η. For each economy, we set $\{\mu^f, \mu^r, \phi^z, \phi^\eta\}$ to generate an equilibrium distribution of exporter firm shares that resembles the empirical distribution for 2012 depicted in Figure OA.6.

Conditional on the realization of the fundamentals for simulation b,

1. We use Proposition 2 to compute the economy's equilibrium, including bilateral trade variables, $\{n_{ij,b}, \bar{x}_{ij,b}, X_{ij,b}\}$, and the quantiles of the distribution of firm log-sales, $Q_{ij,b}(p)$ with $G^{\ln R}_{ij,b}(Q_{ij,b}(p)) = p$. We also compute the true gains from trade for each country i, $GFT_{i,b}$, using Corollary 1 given the observed trade data, $\{n_{ij,b}, X_{ij,b}\}$, and the functions, $\epsilon_{ij}(n)$ and $\rho_{ij}(n)$.

2. For Researcher SP, we implement the semiparametric estimator described in Section 4.1, which only uses $\{n_{ij,b}, \bar{x}_{ij,b}, z_{ij,b}\}$ to estimate $\epsilon^{SP}_{ij,b}(n)$ and $\rho^{SP}_{ij,b}(n)$ and the associated trade elasticity function, $\theta^{SP}_{ij,b}(n)$ in (14). We then compute the gains from trade for each country i, $GFT^{SP}_{i,b}$, using Corollary 1 given $\{n_{ij,b}, X_{ij,b}\}$, $\epsilon^{SP}_{ij,b}(n)$, and $\rho^{SP}_{ij,b}(n)$.

3. For researcher P, we implement a minimum-distance estimator that matches the theoretical and empirical quantiles of the distribution of firm-sales:

$$\min_{\delta} \sum_{p=1}^{99} (G^{\ln R}_{ij}(Q_{ij}(p)|\delta) - p)^2,$$

where δ is the parameter vector in the researcher's model. Note that, under the log-normal distribution of firm sales in Head et al. (2014) and Bas et al. (2017), this estimator is equivalent to a quantile-on-quantile log-normal regression,

$$Q_{ij,b}(p) = \alpha_{ij,b} + \nu_b \Phi^{-1}(p) + \eta_{ij,b}(p)$$

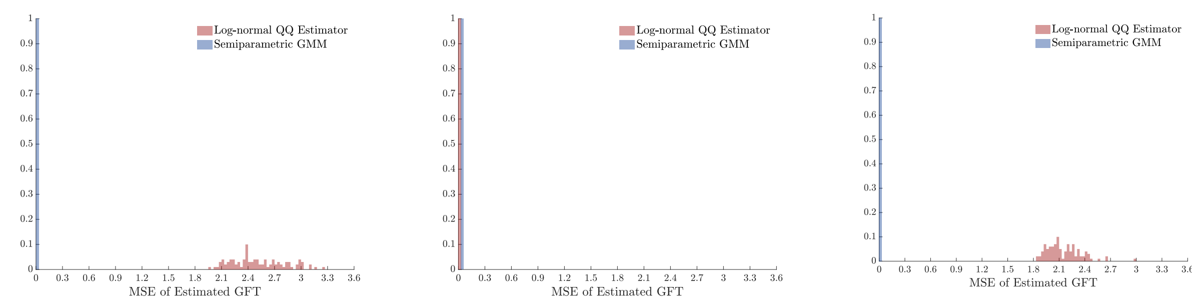
where $\alpha_{ij,b}$ is a market fixed effect, and $p \in \{1, ..., 99\}$ is a percentile of the log-revenue distribution. Given the estimate of ν_b, we obtain the calibrated elasticity functions implied by a log-normal distribution of entry potentials, $\epsilon_{ij}^P(n|\nu_b)$ and $\rho_{ij}^P(n|\nu_b)$, and the associated trade elasticity, $\theta_{ij,b}^P(n)$ in (14). We compute the the gains from trade for each country i, $GFT_{i,b}^P$, using Corollary 1 given $\{n_{ij,b}, X_{ij,b}\}$, $\epsilon_{ij}^P(n|\nu_b)$, and $\rho_{ij}^P(n|\nu_b)$.

4. We then compute the normalized mean square error for the predictions of each researchers as $nMSE_b^g = (\sum_i (GFT_{i,b}^g - GFT_{i,b})^2)^{1/2} / \sum_i GFT_{i,b}$ for $g \in \{SP, P\}$.

B.1.2 Additional Monte Carlo Results

Figure OA.3: Monte Carlo: Distributions of Gains From Trade

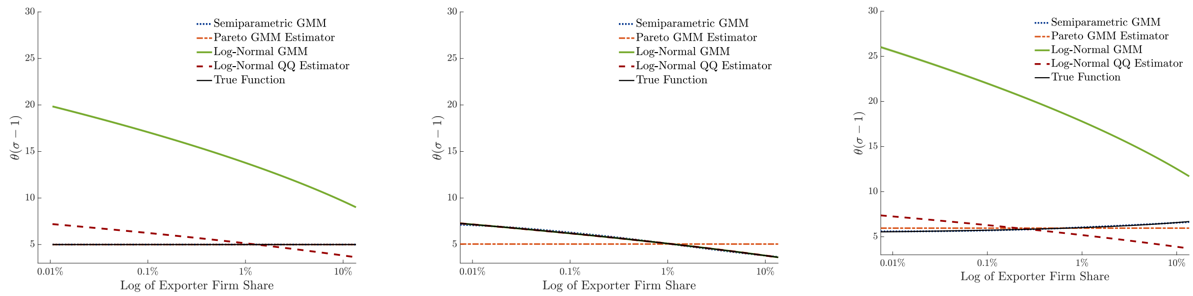
(a) Constant Trade Elasticity (b) Decreasing Trade Elasticity (c) Increasing Trade Elasticity



Note. We display the bias in predicted gains from trade for 100 simulated economies implied by either the Semiparametric GMM estimator that we propose or the QQ Log-Normal estimator that perfectly matches the ex-post distribution of firm-level exports. The blue histogram is the mean squared error of the estimated gains from trade versus the true gains from trade recovered using our semiparametric estimator (i.e., Researcher SP's approach). The red histogram shows the same mean squared error using the QQ log-normal estimator (Researcher P's approach). Panel (a) represents a model in which the trade elasticity is constant. Panel (b) represents a model in which the trade elasticity is decreasing, as implied by a log-normal distribution of entry potentials. Panel (c) represents a model in which the trade elasticity is increasing, as implied by the modified Pareto distribution of entry potentials in Section 2.2.

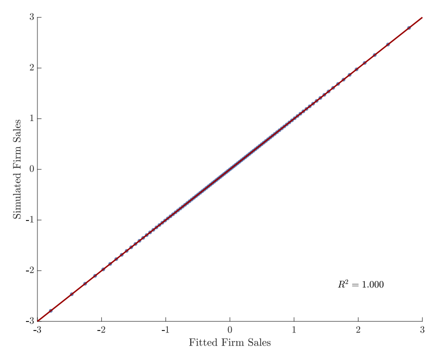
Figure OA.4: Monte Carlo: GMM Estimation with functional for restrictions

(a) Constant Trade Elasticity (b) Decreasing Trade Elasticity (c) Increasing Trade Elasticity



Note. In this figure, we show two additional estimators of the trade elasticity (in addition to the two estimators in Figure 2). The orange dashed line displays the results of our baseline semiparametric GMM estimator under the assumption that the underlying economy is Pareto. The solid green line displays the results of our semiparametric GMM estimator under the assumption that the underlying economy is log-normal. Panel (a) represents a model in which the trade elasticity is constant. Panel (b) represents a model in which the trade elasticity is decreasing, as implied by a log-normal distribution of entry potentials. Panel (c) represents a model in which the trade elasticity is increasing, as implied by the modified Pareto distribution of entry potentials in Section 2.2.

Figure OA.5: Monte Carlo: Example QQ Estimator



Note. In this figure, we show the quantiles of the QQ log-normal estimator (Researcher P's approach). The y-axis represents the log(sales) from the simulated economy (one with an increasing trade elasticity). The x-axis represents the log(sales) from an fitted and estimated log-normal distribution (one with a decreasing trade elasticity). Each dot represents a quantile of firm sales, from the 1st percentile to the 99th percentile. The quantiles perfectly match, with an R^2 of 1.

B.2 Data Construction: Sample for Estimation

The sample creation procedure combines several datasets to construct a comprehensive origin-destination database with information on exporter firm shares and average firm exports in 2010, 2012 and 2014. Table OA.1 reports the sources for each origin country. Table OA.2 presents summary statistics of the key variables.

Number of Exporters and Average Firm Exports. To obtain the number of exporters to each destination (N_{ij}), we follow a priority order. We first use data from the OECD Trade by Enterprise Characteristics (TEC). If a market is not covered in TEC, we then use the Exporter Dynamics Database (EDD) to obtain bilateral exporter counts. Finally, when neither of these sources is available, we obtain the number of exporters to each destination from Australian customs data, and China's enterprise statistics. To guarantee consistency in firm export margins, we measure average exports per firm ($\bar{x}_{ij}$) for each origin-destination using the same source used for N_{ij}. When EDD suppresses $\bar{x}_{ij}$ due to small N_{ij}, we instead construct it as $\bar{x}_{ij} = X_{ij}/N_{ij}$, where X_{ij} comes from BACI and N_{ij} from the EDD.

Number of Domestic Firms and Average Domestic Sales. We assume that all firms in a country sell domestically. Accordingly, to measure the number of domestic firms in each country (N_{ii}), we again follow a packing order. We first obtain data on the number of manufacturing firms from the OECD Structural and Demographic Business Statistics (SDBS). If a country is not covered in this database, we use the alternative OECD data (SSIS). If a country is not available in both databases, we use the World Bank Enterprise Survey data combined with the EDD's exporter counts and export probabilities. We supplement these sources with country-specific datasets for Australia and China. Finally, we measure domestic average sales as $x_{ii} = X_{ii}/N_{ii}$, where X_{ii} are total domestic sales in manufacturing from the World Input-Output Database (WIOD).

Exporter Firm Share (n_{ij}). The variable n_{ij} is the share of firms in country i that export to destination j, adjusted for firm survival. It is constructed as $n_{ij} = (N_{ij}/N_{ii}) \times \text{SurvivalRate}_i$, where N_{ij} is the number of firms from i exporting to j, N_{ii} is the total number of manufacturing firms in country i, and SurvivalRate_i is the one-year survival rate. Firm survival rates are drawn from OECD business demography indicators. When OECD survival rates are unavailable for a country, we use the sample average of 0.85. As a robustness check in Figure OA.26, we drop these origin countries and re-estimate the parameters. In Figure OA.25, we use 3-year, instead of 1-year survival rates. Appendix Figure OA.6 shows the empirical distribution of $\ln n_{ij}$ in our sample for 2012.

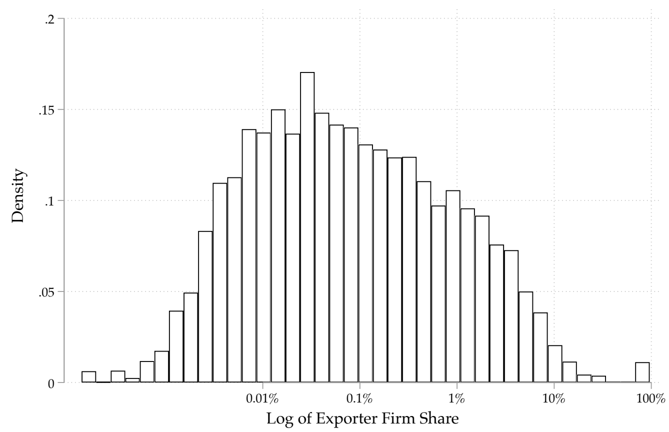
Bilateral Average Tariffs. To construct bilateral average tariffs, we rely on the Global Tariff Database (v_beta1-2024-12) from Teti (2024) for years 2010, 2012, and 2014. Our measure of the bilateral average tariff rate is $\log(1 + \text{tariff}_{ij}/100)$, where tariff_{ij} is the simple average of the import tariffs that j applies to goods from i, converted from percentages. We use the simple arithmetic average across all HS6 product lines with positive trade for an origin-destination.[44] We consider both Applied Harmonized System (AHS)

[44] We match trade flow records with tariff measures at the year-product-origin-destination level, and retain only observations that are successfully merged. This ensures that tariff variables cover only product lines with verified bilateral trade relationships, which helps to attenuate the problem highlighted by Teti (2024) of outlier values for tariffs.

tariffs—reflecting actual rates including preferential agreements—and Most Favored Nation (MFN) tariffs. We set own-country tariffs (where origin equals destination) to zero.

Sectoral Database. For sectoral analysis, we map 2-digit HS products (HS2) into distinct industry groups, as defined in Table OA.3. We then use the World Bank's Exporter Dynamics Database (EDD) to obtain bilateral firm counts ($N_{ij,h}$) and average exports per firm ($\bar{x}_{ij,h}$) at the HS2 level for years 2010 and 2012. We collapse HS2-level observations to categories using firm-count weighted averages for average exports per firm. These sector-level data are then merged with the bilateral average tariff by sector, computed as the simple average of HS6 product lines in each category. Since total manufacturing firm counts (N_{ii}) are not available by sector, we approximate $n_{ij,s}$ by assuming that firms active in each sector represent approximately 10% of the maximum exporter count for that country-sector.

Figure OA.6: Empirical Distribution of Exporter Firm Shares, 2012



Note. Empirical distribution of $\ln(n_{ij})$ in the cross-section of origin-destination pairs in 2012.

Table OA.1: Estimation Data Sources

Country	Source for N_{ii}	Source for N_{ij}	Source for $\bar{x}_{ij}$	Sectoral Data Available
ALB	EDD/WBES	EDD	EDD	1
AUS	SDBS	AUS	AUS	
AUT	SDBS	TEC	TEC	
BEL	SDBS	TEC	TEC	
BFA	EDD/WBES	EDD	EDD	1
BGD	EDD/WBES	EDD	EDD	1
BGR	SDBS	EDD	EDD	
BOL	EDD/WBES	EDD	EDD	1
BRA	SDBS	EDD	EDD	
BWA	EDD/WBES	EDD	EDD	1
CAN	SDBS	TEC	TEC	
CHL	SDBS	EDD	EDD	1
CHN	CHN	CHN	CHN	
CIV	EDD/WBES	EDD	EDD	1
CMR	EDD/WBES	EDD	EDD	1
COL	EDD/WBES	EDD	EDD	1
CRI	EDD/WBES	EDD	EDD	1
CYP	SDBS	TEC	TEC	
CZE	SDBS	TEC	TEC	
DEU	SDBS	TEC	TEC	
DNK	SDBS	EDD	EDD	1
DOM	EDD/WBES	EDD	EDD	1

continued

Table OA.1: Estimation Data Sources

Country	Source for N_{ii}	Source for N_{ij}	Source for $\bar{x}_{ij}$	Sectoral Data Available
ECU	EDD/WBES	EDD	EDD	1
EGY	EDD/WBES	EDD	EDD	1
ESP	SDBS	EDD	EDD	1
EST	SDBS	TEC	TEC	
ETH	EDD/WBES	EDD	EDD	1
FIN	SDBS	TEC	TEC	
FRA	SDBS	TEC	TEC	
GAB	EDD/WBES	EDD	EDD	
GBR	SDBS	TEC	TEC	
GEO	EDD/WBES	EDD	EDD	1
GIN	EDD/WBES	EDD	EDD	1
GRC	SDBS	TEC	TEC	
GTM	EDD/WBES	EDD	EDD	1
HRV	SDBS	EDD	EDD	1
HUN	SDBS	TEC	TEC	
IRL	SDBS	TEC	TEC	
ITA	SDBS	TEC	TEC	
JOR	EDD/WBES	EDD	EDD	1
KEN	EDD/WBES	EDD	EDD	1
KGZ	EDD/WBES	EDD	EDD	1
KHM	EDD/WBES	EDD	EDD	
KOR	SDBS	TEC*	TEC	
LAO	EDD/WBES	EDD	EDD	
LBN	EDD/WBES	EDD	EDD	1
LKA	EDD/WBES	EDD	EDD	
LTU	SDBS	TEC	TEC	
LUX	SDBS	TEC	TEC	
LVA	SDBS	TEC	TEC	
MAR	EDD/WBES	EDD	EDD	1
MDG	EDD/WBES	EDD	EDD	1
MEX	EDD/WBES	EDD	EDD	1
MKD	EDD/WBES	EDD	EDD	
MLI	EDD/WBES	EDD	EDD	
MLT	SDBS	TEC	TEC	
MMR	EDD/WBES	EDD	EDD	1
MUS	EDD/WBES	EDD	EDD	1
MWI	EDD/WBES	EDD	EDD	1
NER	EDD/WBES	EDD	EDD	
NIC	EDD/WBES	EDD	EDD	1
NLD	SDBS	TEC	TEC	
NOR	SDBS	EDD	EDD	1
NPL	EDD/WBES	EDD	EDD	1
NZL	SDBS	TEC	TEC	
PAK	EDD/WBES	EDD	EDD	
PER	EDD/WBES	EDD	EDD	1
POL	SDBS	TEC	TEC	
PRT	SDBS	EDD	EDD	1
PRY	EDD/WBES	EDD	EDD	1
ROU	SDBS	TEC	TEC	
RWA	EDD/WBES	EDD	EDD	1
SEN	EDD/WBES	EDD	EDD	1
SLV	EDD/WBES	EDD	EDD	
SVK	SDBS	TEC	TEC	
SVN	SDBS	TEC	TEC	
SWE	SDBS	TEC	TEC	
SWZ	EDD/WBES	EDD	EDD	1
THA	EDD/WBES	EDD	EDD	1
TUR	SDBS	EDD	EDD	
TZA	EDD/WBES	EDD	EDD	1
UGA	EDD/WBES	EDD	EDD	
URY	EDD/WBES	EDD	EDD	1
USA	SDBS	TEC	TEC	
YEM	EDD/WBES	EDD	EDD	1
ZAF	EDD/WBES	EDD	EDD	1
ZMB	EDD/WBES	EDD	EDD	

Notes: This table shows the most frequent use, however in certain cases when N_{ij} is small, the EDD suppresses $\bar{x}$, but not N_{ij}, so we construct $\bar{x}_{ij} = X_{ij}/N_{ij}$, where X_{ij} comes from BACI and N_{ij} comes from the EDD.

Table OA.2: Estimation Data Summary

Country	Developed	Number of	Average across j		Standard deviation across j	
	Dummy	Destinations	$\ln n_{ij}$	$\ln \bar{x}_{ij}$	$\ln n_{ij}$	$\ln \bar{x}_{ij}$
ALB	0	73	-8.51	-2.82	1.67	2.28
AUS	1	144	-6.91	-1.23	1.9	1.39
AUT	1	58	-3.44	.11	.86	.98
BEL	1	58	-3.21	.36	.74	1.08
BFA	0	53	-7.98	-1.54	1.46	2.24
BGD	0	127	-7	-2.02	2.08	1.42
BGR	0	145	-6.74	-2.26	1.98	1.98
BOL	0	61	-7.15	-1.73	1.45	1.72
BRA	0	156	-7.92	-.82	1.92	1.44
BWA	0	57	-9.25	-3.23	1.66	2.91
CAN	1	47	-4.78	-.18	1.29	1.31
CHL	0	135	-5.13	-1.3	1.9	1.59
CHN	0	155	-4.47	-1.15	1.73	.99
CIV	0	97	-8.41	-.95	1.52	1.72
CMR	0	87	-7.78	-2	1.36	1.5
COL	0	126	-8.64	-1.56	2.01	1.45
CRI	0	108	-8.36	-2.21	1.95	2.09
CYP	1	34	-5.28	-.97	1.13	1.14
CZE	0	58	-4.18	-.21	1.05	.86
DEU	1	35	-3.15	.47	.74	1.15
DNK	1	155	-4.69	-1.55	1.69	1.37
DOM	0	110	-8.7	-2.09	1.66	1.51
ECU	0	106	-8.69	-1.53	1.76	1.82
EGY	0	136	-7.61	-1.72	1.58	1.49
ESP	1	157	-5.78	-1.76	1.95	1.28
EST	0	70	-4.42	-1	1.29	1.29
ETH	0	74	-8.14	-2.68	1.26	2.34
FIN	1	35	-3.48	.3	.93	1.2
FRA	1	58	-3.99	.17	.9	.99
GAB	0	57	-6.48	-1.96	1.32	2.16
GBR	1	58	-3.96	-.2	.85	.99
GEO	0	84	-9.17	-1.43	1.6	1.77
GIN	0	60	-8.73	-2.31	1.31	3.02
GRC	1	58	-5.78	-.61	1.26	1.07
GTM	0	106	-8.21	-1.92	1.94	1.98
HRV	0	101	-7.29	-2.49	1.7	2.33
HUN	0	35	-4.58	.66	1.15	.91
IRL	1	35	-3.66	1.15	.95	1.4
ITA	1	58	-3.54	-.89	1.14	.82
JOR	0	118	-7.04	-2.07	1.72	1.71
KEN	0	112	-7.35	-2.74	1.69	1.7
KGZ	0	57	-7.58	-2.06	1.42	2.03
KHM	0	93	-7.25	-2.56	1.62	1.92
KOR	1	48	-5.24	-.04	1.32	.99
LAO	0	45	-6.55	-1.86	1.22	2.18
LBN	0	136	-6.86	-2.74	1.83	1.45
LKA	0	135	-8.34	-2.41	1.66	1.55
LTU	0	58	-4.73	-.6	1.31	1.3
LUX	1	33	-2.43	.61	.75	1.55
LVA	0	47	-4.25	-.75	1.18	1
MAR	0	117	-7.74	-1.3	1.67	1.77
MDG	0	81	-7.16	-2.34	1.57	1.71
MEX	0	149	-9.39	-1.5	2.29	1.61
MKD	0	77	-7.61	-2.7	1.85	1.96
MLI	0	41	-7.17	-2.14	1.31	2.77
MLT	1	57	-4.72	-.63	1.14	1.69
MMR	0	51	-8.12	-1.44	1.45	1.5
MUS	0	108	-8.14	-2.55	1.53	2.03
MWI	0	75	-7.95	-1.16	1.29	2.2
NER	0	27	-7.35	-2.19	.83	2.8
NIC	0	84	-8.33	-2.65	1.6	2.23
NLD	1	58	-3.09	.14	.71	1.07
NOR	1	148	-5.25	-1.62	1.9	1.93
NPL	0	78	-7.76	-4.19	1.68	1.48
NZL	1	39	-4.95	-.73	1.21	1.39
PAK	0	149	-7.52	-2.44	1.85	1.21
PER	0	122	-8.31	-1.65	1.98	1.65
POL	0	58	-4.72	-.2	1.16	.85
PRT	1	149	-6.37	-2	2.03	1.44

continued

Table OA.2: Estimation Data Summary

Country	Developed Dummy	Number of Destinations	Average across j		Standard deviation across j	
			$\ln n_{ij}$	$\ln \bar{x}_{ij}$	$\ln n_{ij}$	$\ln \bar{x}_{ij}$
PRY	0	66	-7.07	-1.26	1.43	1.94
ROU	0	27	-4.78	.28	1.38	.93
RWA	0	52	-9.13	-3.98	1.3	2.38
SEN	0	81	-7.18	-2.99	1.53	2.42
SLV	0	82	-7.49	-2.44	1.73	1.73
SVK	0	56	-4.28	-.16	1.54	.82
SVN	1	55	-4.16	-.43	1.11	.94
SWE	1	58	-3.48	.33	.9	.98
SWZ	0	58	-8.25	-1.53	1.39	2.83
THA	0	157	-8.3	-1.02	1.78	1.25
TUR	0	152	-6.82	-1.37	1.81	.92
TZA	0	101	-8.22	-1.89	1.52	1.8
UGA	0	81	-8.83	-2.92	1.45	2.97
URY	0	123	-8.02	-1.54	1.62	1.88
USA	1	47	-4.34	-.01	1.26	1.21
YEM	0	46	-6.95	-2.07	1.17	2.66
ZAF	0	153	-8.14	-1.62	1.86	1.35
ZMB	0	74	-8.77	-1.98	1.53	2.84

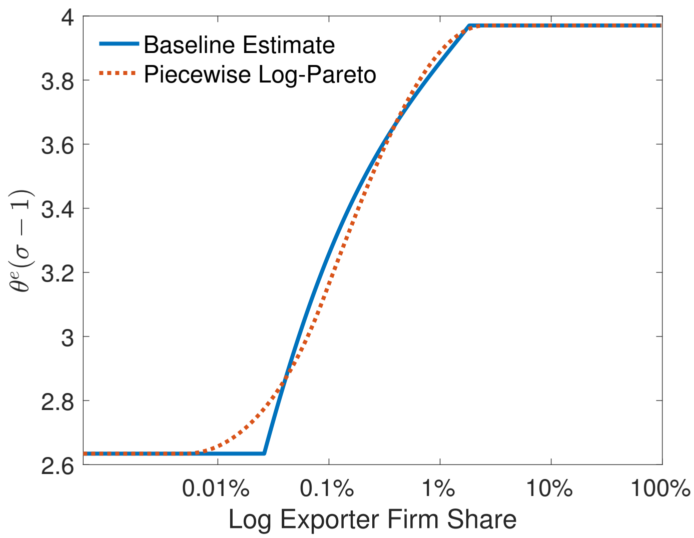
Notes: This table shows the summary statistics of the key variables used.

Table OA.3: HS Manufacturing Sectoral Aggregates

#	Sector Description	HS2 Range
1	Foodstuffs	16-24
2	Mineral Products & Stone/Glass	25-27, 68-71
3	Chemicals & Allied Industries	28-38
4	Plastics/Rubbers	39-40
5	Wood & Wood Products	44-49
6	Textiles, Leather & Footwear	41-43, 50-67
7	Metals	72-83
8	Machinery	84-85
9	Transportation	86-89

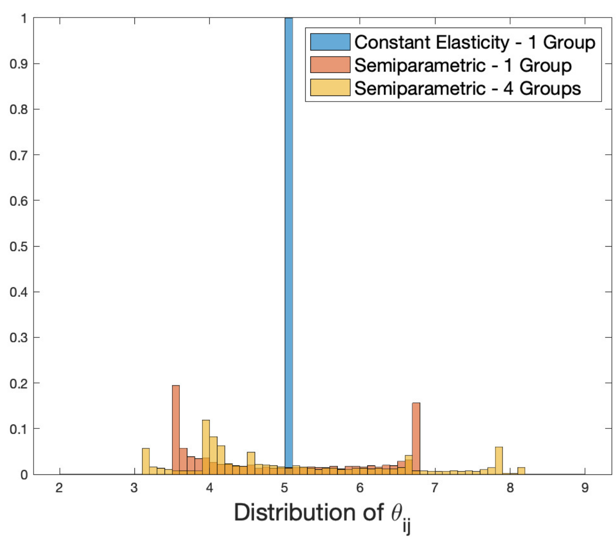
B.3 Additional Estimation Figures and Tables

Figure OA.7: Semiparametric Gravity of Firm Exports versus Log-Pareto Distribution of Entry Potentials



Note. Figure displays the baseline estimate of the extensive margin elasticity from panel (a) of Figure 3, as well as a simulation using a piecewise and shifted log-Pareto distribution, with inverse CDF: $e^{k/(1-p)^{1/\alpha}} + \mu$, except at the tails. In particular, we constrain the tails to have constant elasticities when $n_{ij} > \bar{e}$ and $n_{ij} < \underline{e}$. A minimum distance routine fits parameters $\{k, \alpha, \mu, \bar{e}, \underline{e}\}$.

Figure OA.8: Empirical Distribution of Bilateral Trade Elasticities in 2012



Note. Figure displays the histogram of the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma-1)\theta(n^0_{ij})$, in the 2012 sample of origin-destination pairs. The blue bars denote the empirical distribution implied by the constant-elasticity benchmark obtained from the estimation of (35) under (22) (as reported in panel (c) of Figure OA.9). Yellow and orange bars denote empirical distributions implied by semiparametric estimates obtained with GMM estimator in (35) for a single group with all countries and for four groups based on the income level of the origin and destination (as reported in Figures 3 and 4, respectively).

Table OA.4: Reduced-Form Gravity Specification

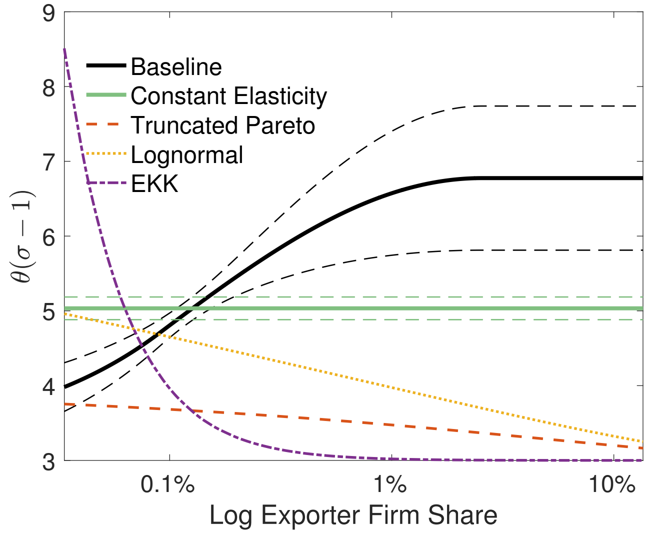
	(1)	(2)	(3)	(4)
ln(Distance_ij)	1.305	1.281		
	(0.056)	(0.056)		
ln(Distance_ij)	0.223	0.139		
× Above Median n_{ij}	(0.059)	(0.061)		
Developing Origins				
ln(Distance_ij)			1.193	1.172
			(0.064)	(0.065)
ln(Distance_ij)			0.456	0.372
× Above Median n_{ij}			(0.068)	(0.069)
Developed Origins				
ln(Distance_ij)			1.430	1.364
			(0.091)	(0.091)
ln(Distance_ij)			-0.250	-0.308
× Above Median n_{ij}			(0.097)	(0.098)
N	9738	9738	9738	9738
R^2	0.798	0.800	0.800	0.802
Gravity Controls		✓		✓
FE	i, j, $\mathbb{I}_{n_{ij}>Median}$		i, j, $\mathbb{I}_{n_{ij}>Median}$×Developed	

Note. The table reports estimates of

$$-\ln X_{ij} = \beta_g \ln D_{ij} + \beta_g^H 1[n_{ij} > med_g] \ln D_{ij} + C_{ij}\gamma + \delta_i + \zeta_i + \epsilon_{ij}$$

where g denotes a group of origin-destination pairs ij, X_{ij} denotes trade flows, D_{ij} denotes distance, $1[n_{ij} > med_g]$ denotes a dummy indicating that n_{ij} is above the sample median for group g, C_{ij} denotes controls, and δ_i and ζ_j denote origin and destination fixed effects. The control set always includes $1[n_{ij} > med_g]$ for each group. Columns (2) and (4) also include the other variables in z_{ij} described in Section 5.1. Columns (1) and (2) consider a single group that pools all origin-destination pairs, and columns (3) and (4) consider two groups of markets ($G = 2$) defined by whether the origin is developed or developing (see Table OA.2).

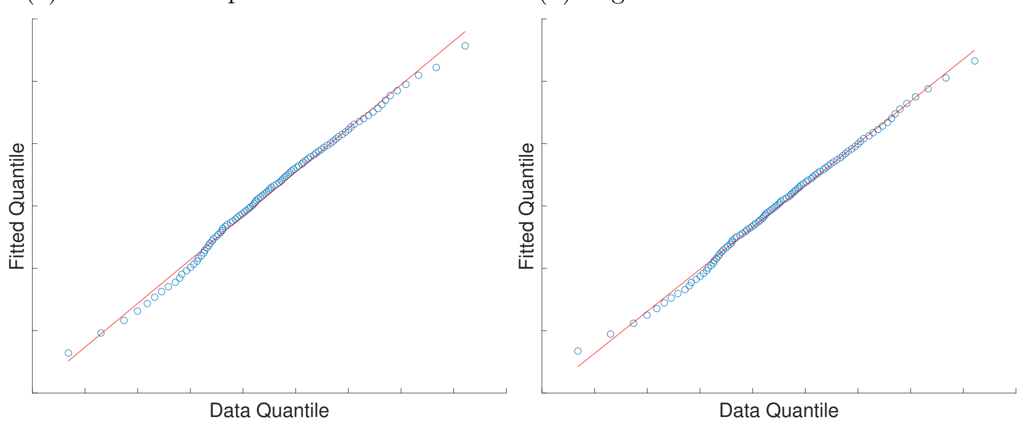
Figure OA.9: Elasticity of Firm Exports and Distributional Assumptions – Single Group



Note. This reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$. Solid black lines and dashed black lines are the estimates and associated 90% confidence intervals of the semiparametric estimates reported in Figure 3. The green solid line is the estimate of the constant-elasticity benchmark obtained from the estimation of (35) under (22) for our baseline sample of origin-destination pairs, with dashed green lines the associated 90% confidence intervals. We report the elasticity functions obtained when the productivity distribution is truncated Pareto with cutoff parameter of $H = 2.85$ (Melitz and Redding, 2015), log-normal with dispersion parameter of 0.79 (Head et al., 2014), and the mixture of Pareto and log-normal in Eaton et al. (2011).

Figure OA.10: Log-Normal QQ Estimator: Exports of Colombian Firms to the United States

(a) Baseline Semiparametric Estimator (b) Log-Normal Distributed Estimates



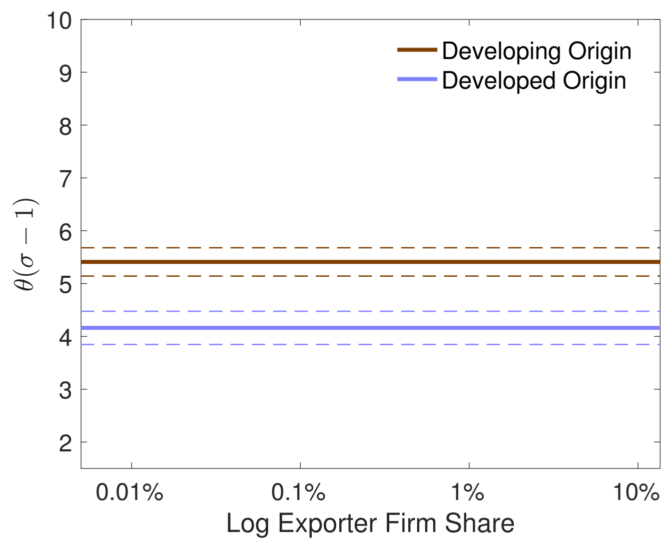
Note. Figure compares the predicted and empirical quantiles of the distribution of exports to the United States of Colombian firms in 2012. The left panel shows the QQ estimator generated from our baseline specification, and the right panel shows the QQ estimator generated from an unconditional log-normal distribution of sales.

Table OA.5: Fit of QQ Estimator: Exports of Colombian Firms by Destination

Export Destination	Baseline Model	Log-Normal
USA	0.99	1.00
CHN	0.99	0.99
ESP	0.99	0.99
PAN	1.00	1.00
VEN	1.00	1.00
NLD	0.99	0.97
CHL	0.98	0.98
ECU	0.99	0.99
PER	0.99	1.00
BRA	1.00	0.99
Mean	0.99	0.99

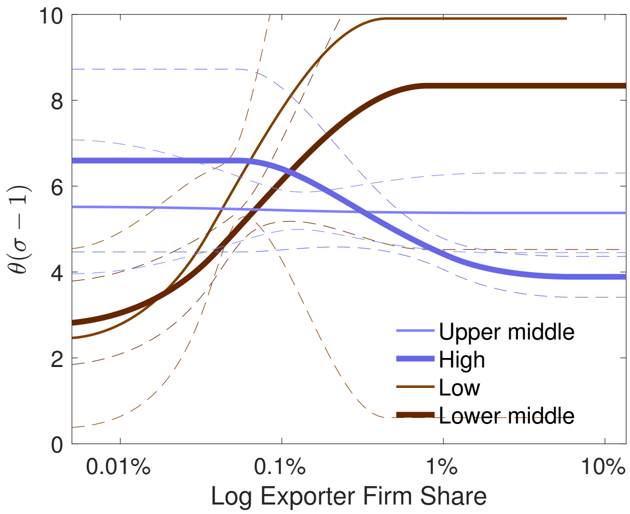
Note. This table shows the fit of the QQ Estimator measured as the R^2 of a regression of the model predicted on the actual values of the quantiles of the distribution of exports of Colombian firms in 2012, separately for each of the 10 largest export markets (by number of exporters). Column (1) reports results for the QQ estimator generated from our baseline specification, and column (2) for the QQ estimator generated from an unconditional log-normal distribution of sales.

Figure OA.11: Constant-Elasticity Gravity – Developed and Developing Origins



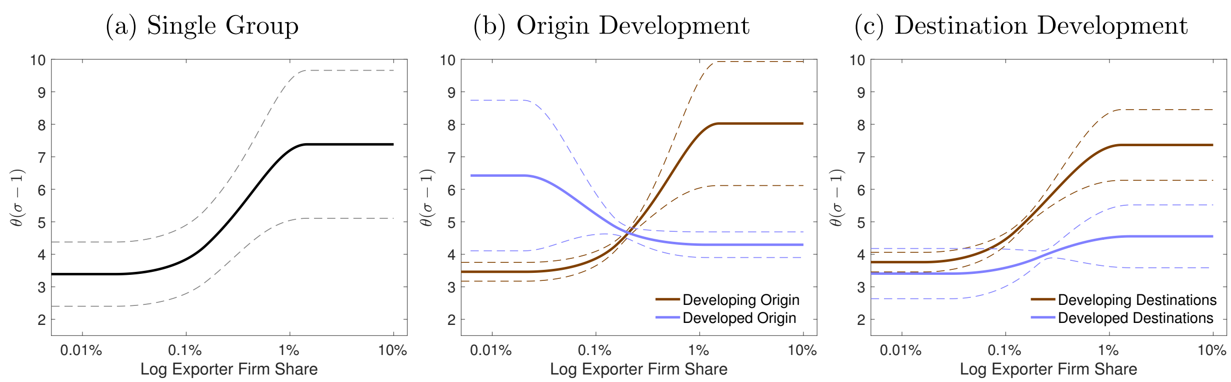
Note. Estimates obtained with GMM estimator in (35) under (22) in the 2012 sample of 7,243 origin-destination pairs. Two groups of markets ($G = 2$) defined by whether the origin country is developed (light purple) or developing (dark brown), as defined in Table OA.2. We report the elasticity of the (absolute) elasticity of bilateral trade with respect to trade costs, $(\sigma-1)\theta_g$. Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.12: Semiparametric Gravity of Firm Exports – Origin's Income Level



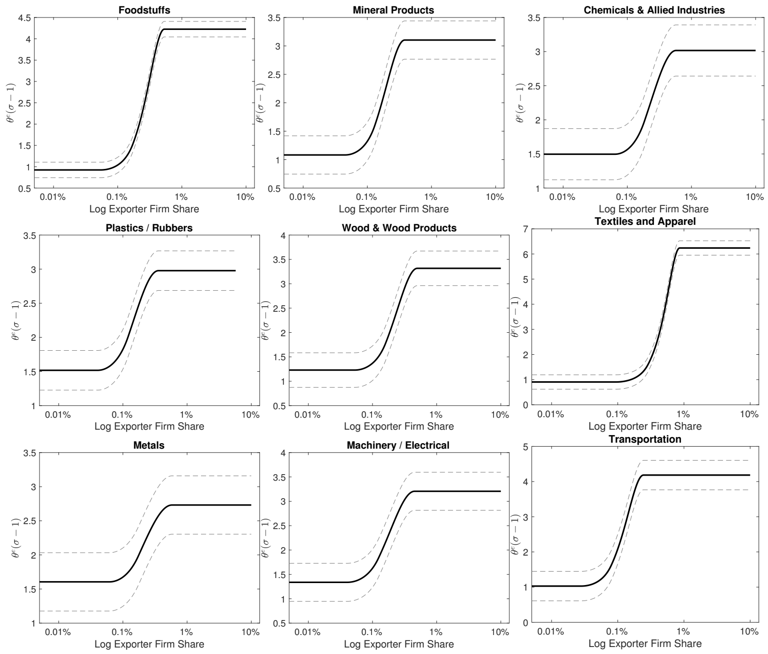
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination. We assume that there are four groups of markets ($G = 4$) defined by the origin's level of income (low, med-low, med-high, high) according to the World Bank classification in 2000. We report the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta_g(n)$. Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.13: Semiparametric Gravity of Firm Exports – Within-Sector Estimation



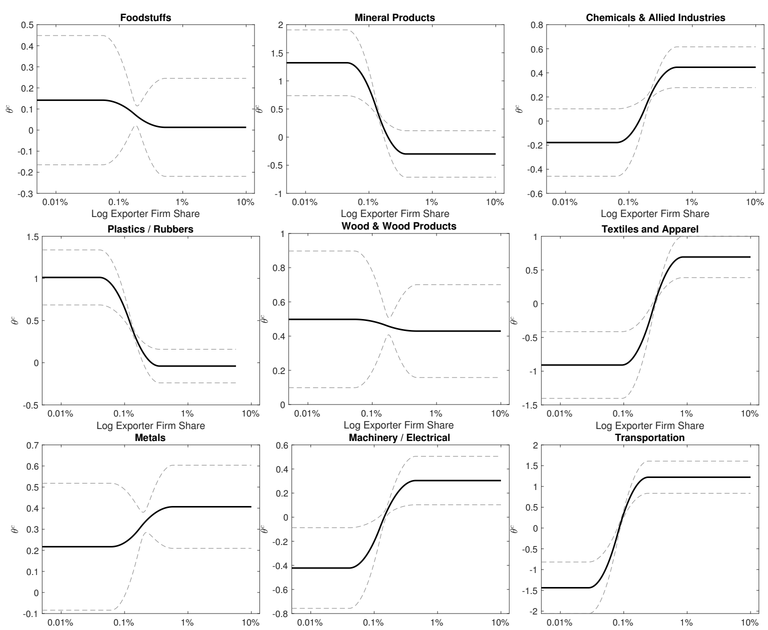
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 16,052 markets defined as origin-destination-sector triplets and fixed effects for sector-origin and sector-destination. We use the raw tariff data at the sectoral level. We report estimates for a single group with all markets in panel (a), for two groups based on whether the origin country is developed in panel (b), and for two groups based on whether the destination country is developed in panel (c). Solid lines are estimates of the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta_g(n)$, and dashed lines are the associated 90% confidence intervals computed with robust standard errors.

Figure OA.14: Semiparametric Gravity of Firm Exports – Extensive Margin Elasticity



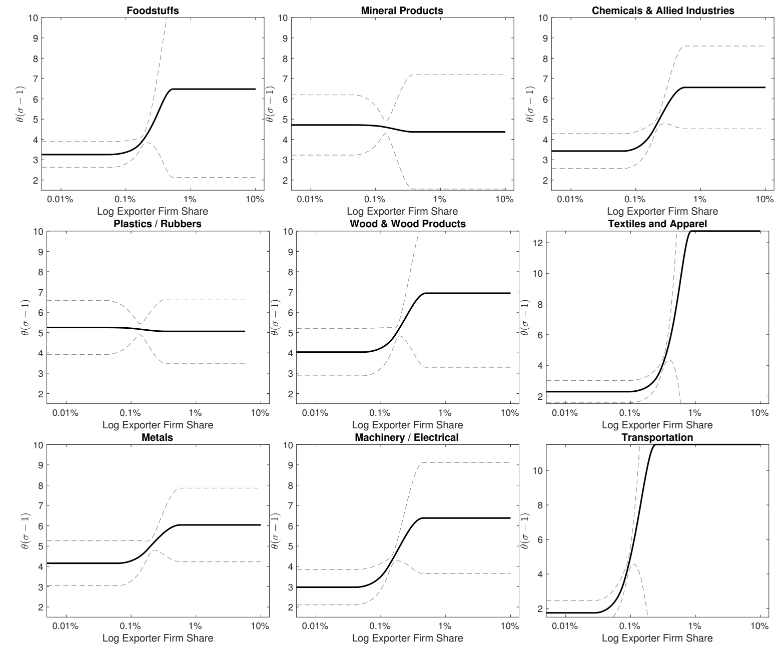
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of markets defined as origin-destination-sector triplets and fixed effects for sector-origin and sector-destination. We report estimates for a single group with all markets. Each panel reports estimates for the sector indicated in the title. Solid lines denote the extensive margin elasticity, $(\sigma - 1)\theta_s^e(n)$ with $\theta_s^e(n)$ defined in (11), and dashed lines are the associated 90% confidence intervals computed with robust standard errors.

Figure OA.15: Semiparametric Gravity of Firm Exports – Firm Composition Elasticity



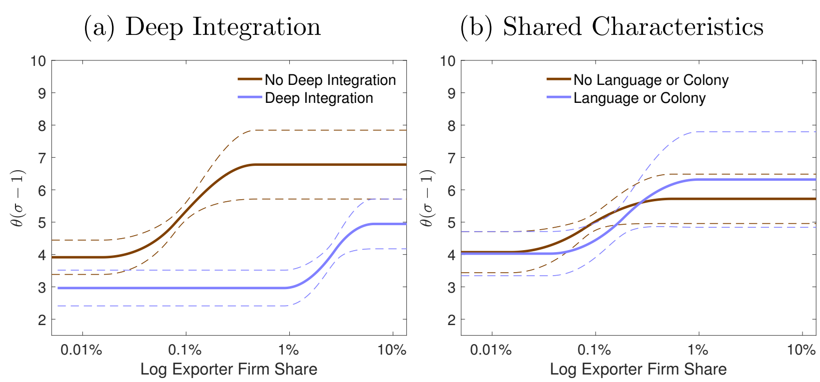
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of markets defined as origin-destination-sector triplets and fixed effects for sector-origin and sector-destination. We report estimates for a single group with all markets. Each panel reports estimates for the sector indicated in the title. Solid lines denote the firm composition elasticity, $\theta_s^c(n)$ defined in (13), and dashed lines are the associated 90% confidence intervals computed with robust standard errors.

Figure OA.16: Semiparametric Gravity of Firm Exports – Bilateral Trade Elasticity



Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of markets defined as origin-destination-sector triplets and fixed effects for sector-origin and sector-destination. We report estimates for a single group with all markets. Each panel reports estimates for the sector indicated in the title. Solid lines denote estimates of the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma-1)\theta_s(n)$, and dashed lines are the associated 90% confidence intervals computed with robust standard errors.

Figure OA.17: Semiparametric Gravity of Firm Exports – Determinants of Market Integration

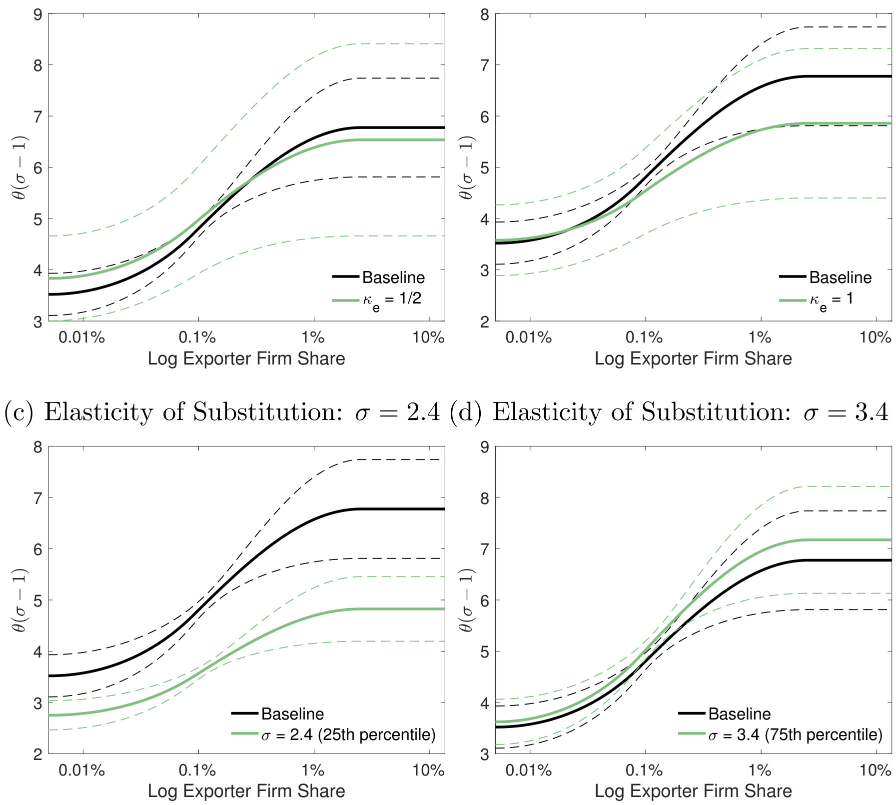


Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination. We assume that there are two groups of markets ($G = 2$) defined by whether the origin and destination have a free trade agreement and a common currency in panel (a), and the origin and destination have either a common language or colonial ties in panel (b). Solid lines are estimates of the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma-1)\theta_g(n)$, and dashed lines are the associated 90% confidence intervals computed with robust standard errors.

Figure OA.18: Semiparametric Gravity of Firm Exports – Alternative Cost Pass-Through

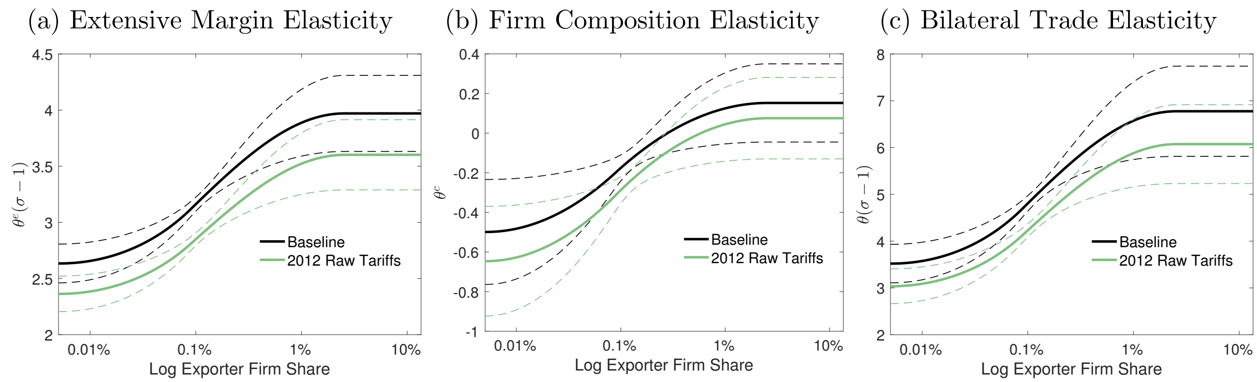
(a) Pass-through to Fixed Cost of 0.5 (b) Pass-through to Fixed Cost of 1

(c) Elasticity of Substitution: $\sigma = 2.4$ (d) Elasticity of Substitution: $\sigma = 3.4$



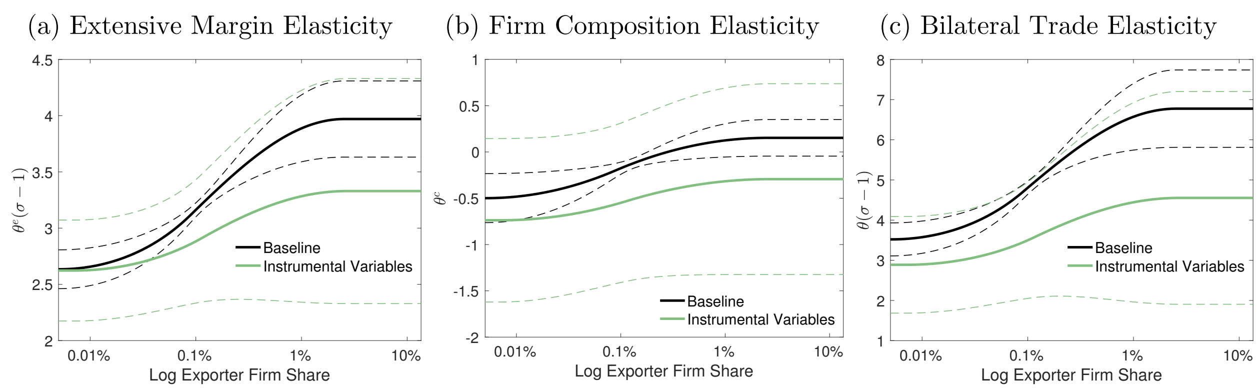
Note. Estimates obtained with GMM estimator in (35) the 2012 sample of 7,243 origin-destination pairs for a single group pooling all pairs ($G = 1$). We report estimates assuming in panel (a) that the pass-through of tariffs to fixed costs is 0.5 ($\kappa^r = 1 - \sigma$ and $\kappa^e = \sigma - 0.5$), in panel (b) that pass-through of tariffs to fixed costs is 1 ($\kappa^r = 1 - \sigma$ and $\kappa^e = \sigma$), in panels (c) and (d) that σ is respectively given by the 25th and 75th percentiles of the estimates in Redding and Weinstein (2024) (i.e., $\sigma = 2.4$ and $\sigma = 3.4$). Solid lines are estimates of the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$, and dashed lines are the associated 90% confidence intervals computed with robust standard errors.

Figure OA.19: Semiparametric Gravity of Firm Exports, Alternative Tariff Database



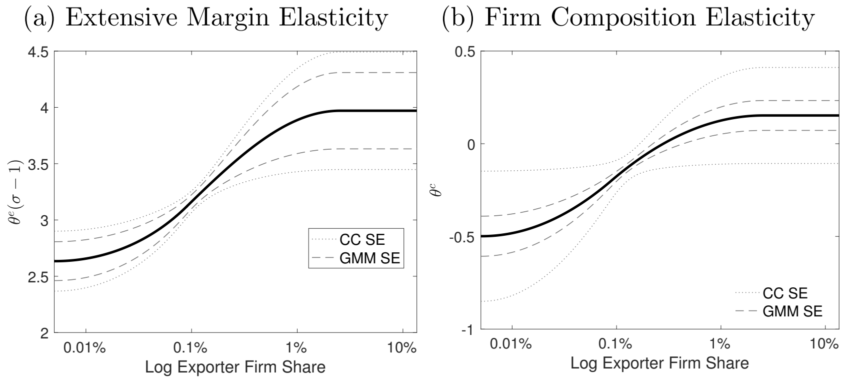
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of origin-destination pairs for a single group pooling all pairs ($G = 1$). Black lines represent baseline estimates based on the Global Tariff Database from Teti (2024). Green lines present estimates based on the raw tariff data from the TRAINS database. Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.20: Semiparametric Gravity of Firm Exports – Tariff IV



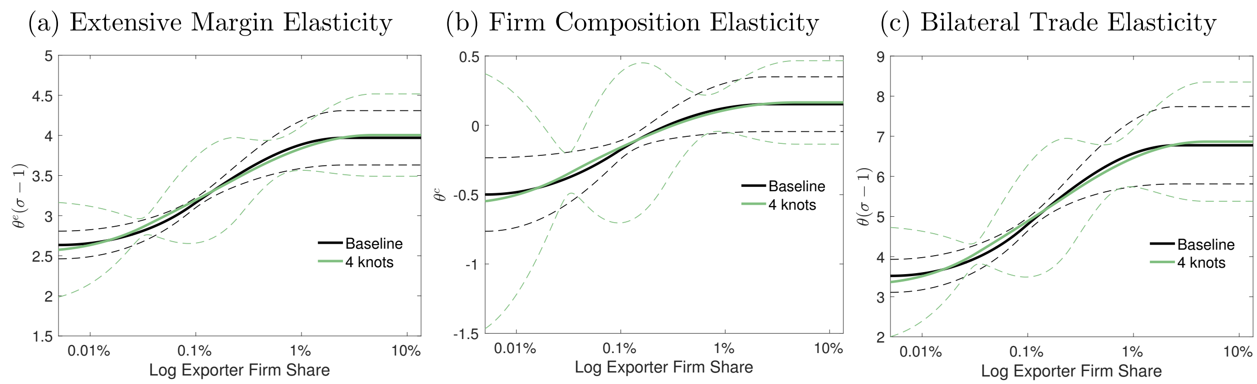
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination pairs for a single group pooling all pairs ($G = 1$). Here we use z_{ij}^{tariffIV} as instrument variable for bilateral import tariffs in z_{ij}. Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.21: Semiparametric Gravity of Firm Exports – Alternative Inference



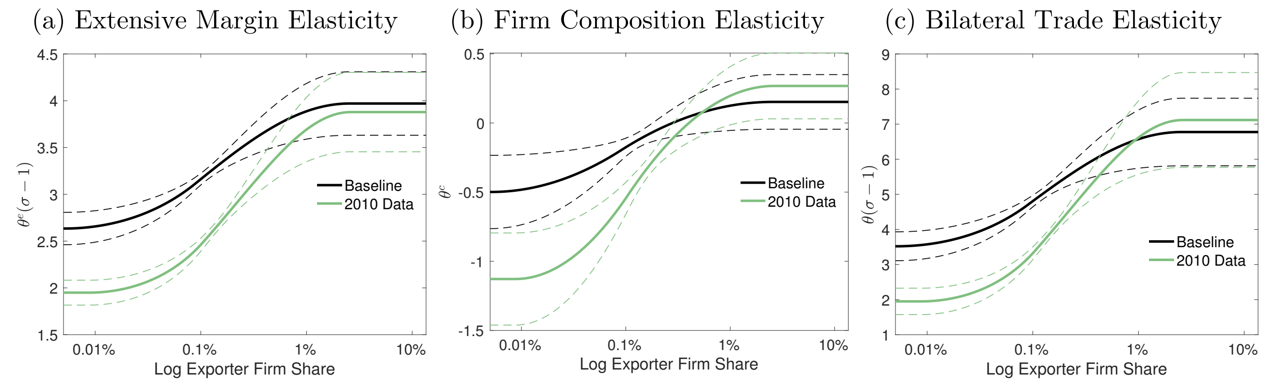
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination pairs for a single group pooling all pairs ($G = 1$). Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), and panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with the nonparametric inference procedure of Chen and Christensen (2018).

Figure OA.22: Semiparametric Gravity of Firm Exports – Alternative Functional Form



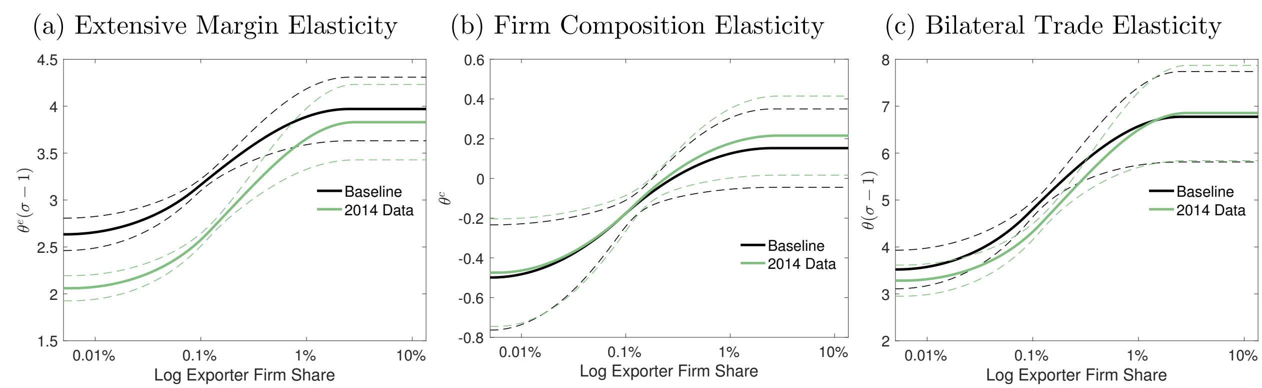
Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination pairs for a single group pooling all pairs ($G = 1$). Estimates reported in green are based on Assumption 4 over five intervals ($M = 5$) instead of the three intervals imposed in the baseline specification shown in black ($M = 3$). Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.23: Semiparametric Gravity of Firm Exports – Single Group, 2010



Note. Estimates obtained with GMM estimator in (35) in the 2010 sample of origin-destination pairs for a single group pooling all pairs ($G = 1$). Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

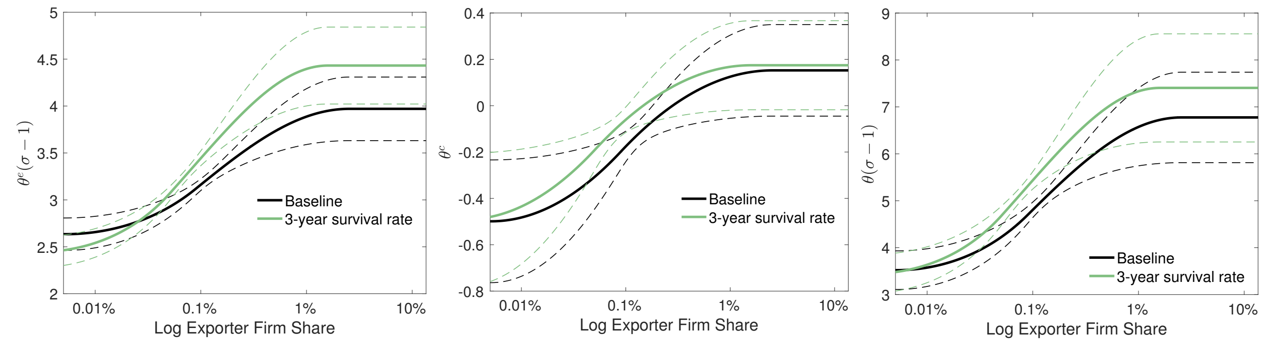
Figure OA.24: Semiparametric Gravity of Firm Exports – Single Group, 2014



Note. Estimates obtained with GMM estimator in (35) in the 2014 sample of origin-destination pairs for a single group pooling all pairs ($G = 1$). Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.25: Semiparametric Gravity of Firm Exports – 3-year Survival Rate n_{ii}

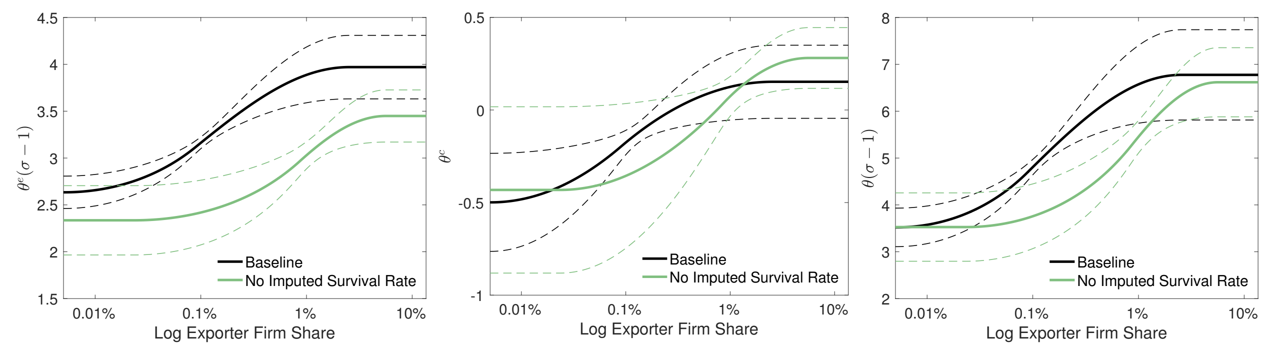
(a) Extensive Margin Elasticity (b) Firm Composition Elasticity (c) Bilateral Trade Elasticity



Note. Estimates obtained with GMM estimator in (35) in the 2012 sample of 7,243 origin-destination pairs for single group pooling all pairs ($G = 1$). Measure of n_{ii} is the survival rate over three years instead of one year used in the baseline. Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

Figure OA.26: Semiparametric Gravity of Firm Exports – Dropping Observations with Imputed n_{ii}

(a) Extensive Margin Elasticity (b) Firm Composition Elasticity (c) Bilateral Trade Elasticity

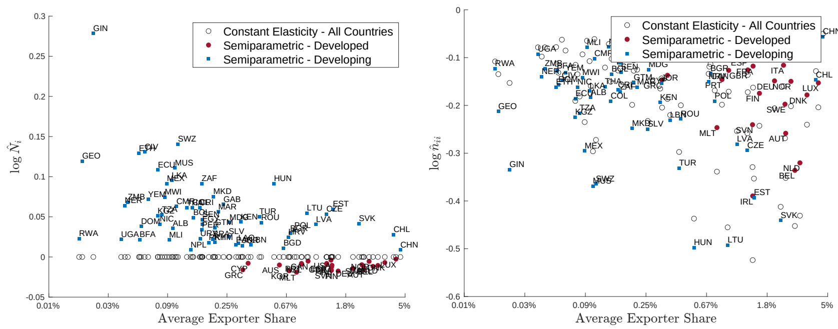


Note. Estimates obtained with GMM estimator in (35) in the the 2012 subsample of 1,844 origin-destination pairs without imputed n_{ii} for single group pooling all pairs ($G = 1$). Panel (a) reports the extensive margin elasticity, $(\sigma - 1)\theta^e(n)$ with $\theta^e(n)$ defined in (11), panel (b) reports the firm composition elasticity, $\theta^c(n)$ defined in (13), and panel (c) reports the (absolute) elasticity of bilateral trade with respect to bilateral trade costs, $(\sigma - 1)\theta(n)$ with $\theta(n)$ defined in (14). Solid lines are the point estimates and dashed lines are the 90% confidence intervals computed with robust standard errors.

C Counterfactual Analysis: Additional Results

C.1 Additional Results

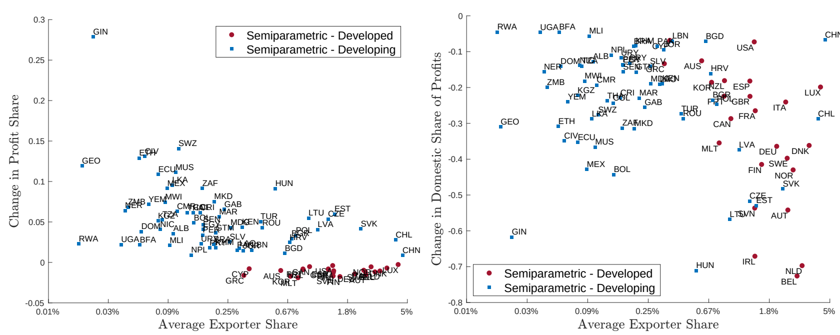
Figure OA.27: Gains from Trade: Entry and Selection of Domestic Firms



(a) Domestic Firm Entry (b) Domestic Firm Selection

Note. Figure reports the percentage change in the mass of firms $\hat{N}_i$ (panel a) and in the share of domestic firms $\hat{n}_{ii}$ (panel b) implied by moving from autarky to the observed equilibrium in 2012, computed with the formula in Corollary 1. White circles represent predictions obtained with the constant-elasticity estimates implied by (35) under (22). The blue squares and red dots represent predictions for developing and developed countries, respectively, that we obtain with the semiparametric estimates reported in Figure 4.

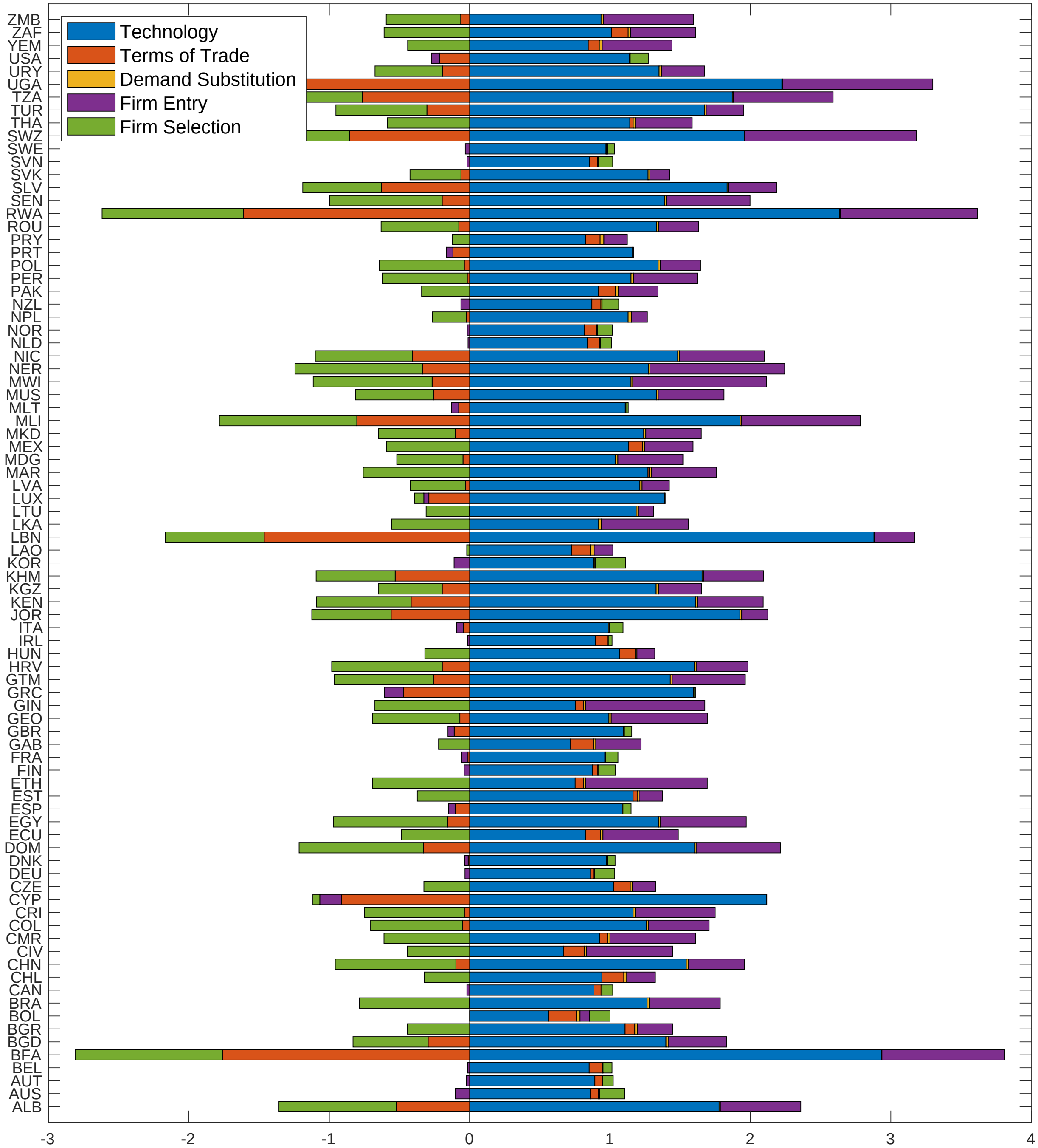
Figure OA.28: Gains from Trade: Firm Profit Margins



(a) Profit Share of Income (b) Domestic Share of Profits

Note. This figure computes the profit changes implied by moving from autarky to the observed equilibrium in 2012. Panel (a) reports the log-change in the share of income accruing to profits, $\log \hat{\pi}_i$. Panel (b) reports the log-change in the domestic share of profits, $\log \hat{s}^{\pi}_{ii}$. The blue squares and red dots represent predictions for developing and developed countries, respectively, that we obtain with the semiparametric estimates reported in Figure 4.

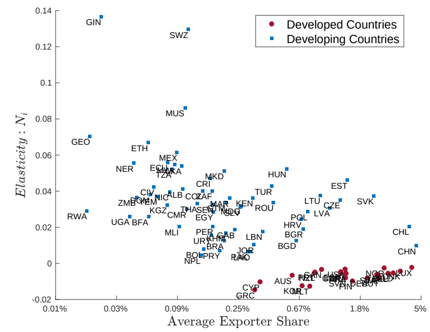
Figure OA.29: Impact of a Uniform Reduction in Trade Costs on Welfare and its Components



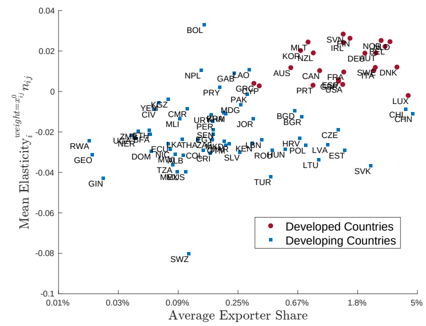
Note. Starting from the observed equilibrium in 2012, we compute the counterfactual equilibrium implied by a reduction of 1% in bilateral trade costs between all countries. For each group of countries, the size of the column denotes 100 times the log-change in real wage normalized by the shock size of 0.01. Each region of a row corresponds to a component of the welfare change in (24). Counterfactual predictions computed with Proposition 3.b, given the elasticity estimates reported in Figure 4.

Figure OA.30: Impact of a Uniform Reduction in Trade Costs on Firm Entry and Selection

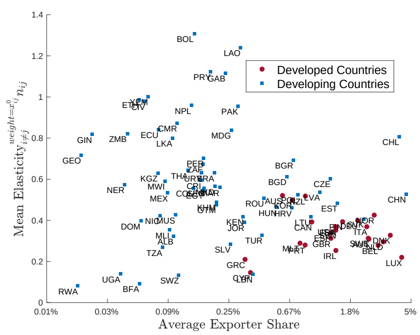
(a) Domestic Firm Entry



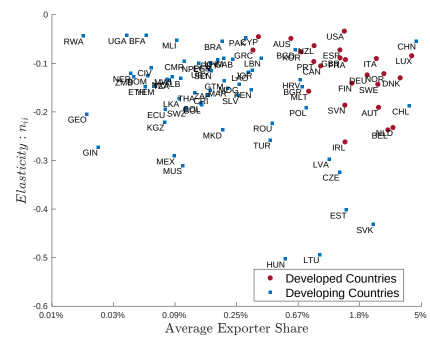
(b) Firm Selection



(c) Foreign Firm Selection



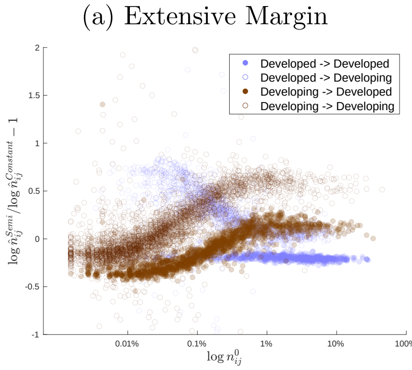
(d) Domestic Firm Selection



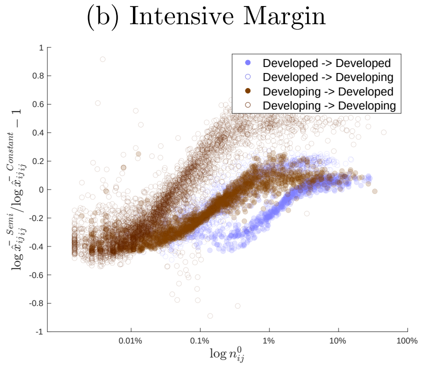
Note. Starting from the observed equilibrium in 2012, we compute the counterfactual equilibrium implied by a reduction of 1% in bilateral trade costs between all countries. Each panel reports (100 times) changes in outcomes for a country against that country's log of the average firm exporter share in 2012. We report log-change of the mass of firms, $\ln \hat{N}_i$, in panel (a), of firm selection, $\sum_j x_{ij}^0 \ln \hat{n}_{ij}$,in panel (b), of foreign firm selection, $\sum_{i\neq j} x_{ij}^0 \ln \hat{n}_{ij} / \sum_{i\neq j} x_{ij}^0$, in panel (c), and of domestic firm selection, $\ln \hat{n}_{ii}$, in panel (d). Counterfactual predictions computed with Proposition 3.b, given the elasticity estimates reported in Figure 4.

Figure OA.31: Impact of a Uniform Reduction in Trade Costs on Firm Export Margins: The Role of Parametric Assumptions

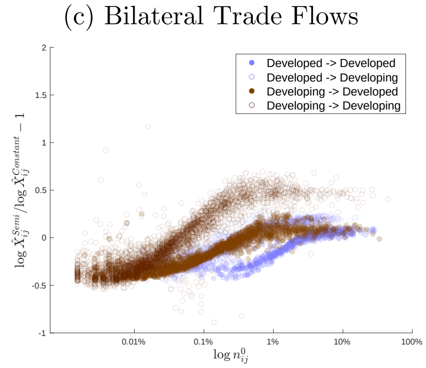
(a) Extensive Margin



(b) Intensive Margin



(c) Bilateral Trade Flows



Note. For a reduction of 1% in bilateral trade costs between all countries starting from the observed equilibrium in 2012, the figure reports in the vertical axis is the ratio of the log-change of each margin of firm exports for an origin-destination pair predicted by the semiparametric and constant-elasticity specifications, and the horizontal axis is the log of the firm exporter share in 2012 for that origin-destination. Panel (a) does this for the extensive margin ($\ln \hat{n}_{ij}$), panel (b) for the intensive margin ($\ln \hat{\bar{x}}_{ij}$), and panel (c) for bilateral trade flows ($\ln \hat{X}_{ij}$). Semiparametric and constant-elasticity predictions use the elasticity estimates in Figures 4 and OA.9, respectively.

C.2 Heterogeneous Changes in Bilateral Trade Costs

In this Appendix, we simulate an asymmetric shock to bilateral trade costs across countries motivated by the rules of the Generalized System of Preference (GSP). Under those rules, developed countries concede preferentially lower import barriers to a subset of developing countries (i.e., those in the country's GSP list). We consider a counterfactual in which developed countries reduce further barriers on imports from countries in their GSP lists. Our shock reduces bilateral trade costs by 1% only for developing origins in the GSP list of each developed destination that allows preferential treatment under GSP rules.[45]

Appendix Table OA.6 reports average welfare gains for all countries, the developed countries reducing import costs (donors), and the developing countries benefiting from the reduction (beneficiaries). The second column indicates that this shock has a smaller impact on global welfare relative to the uniform reduction in Table 1, given that it affects only a subset of the trading partners in the world. While welfare gains still are larger for developed countries, it is due to different mechanisms. The shock only reduces import costs for donor countries, so only these countries have a positive technology term. Because now the shock is heterogeneous, there is a substantial contribution of terms of trade. Donor countries experience a deterioration in their terms of trade, which reduces welfare by an equivalent of 61% of their overall gain. The opposite is true for developing countries. The contributions of the firm components in this case are qualitatively similar to those obtained with the uniform reduction in trade costs, but larger in magnitude.

In Appendix Figure OA.32, an analog of Figure 6 investigates the role of parametric assumptions for this alternative counterfactual. Panel (a) shows that semiparametric estimates yield welfare responses that can be substantially different from those implied by the constant-elasticity benchmark, with differences of more than 30% in absolute value for several developing and developed countries. Interestingly, panel (b) shows that neoclassical terms are more important for heterogeneous changes in trade costs, since they generate movements in terms of trade. However, such differences are only weakly correlated with a country's average exporter share. Panel (c) indicates that firm components also lead to substantial deviations in welfare predictions, again systematically linked to the country's average exporter firm share. This pattern reflects substantial differences in predicted responses in firm export margins, as detailed in Appendix Figure OA.33.

[45] We obtain the list of beneficiaries of GSP in 2012 for the following developed countries in our sample: the European Union, the United States, Japan, Australia, Canada, New Zealand, and South Korea. In reality, these countries reduce tariffs imposed on imports from the developing countries in their GSP lists. Our counterfactual exercise instead considers a hypothetical reduction in import barriers that does not affect tax revenue. As such, it should be seen as a reduction in non-tariff barriers, like sanitary and inspection requirements or technical barriers.

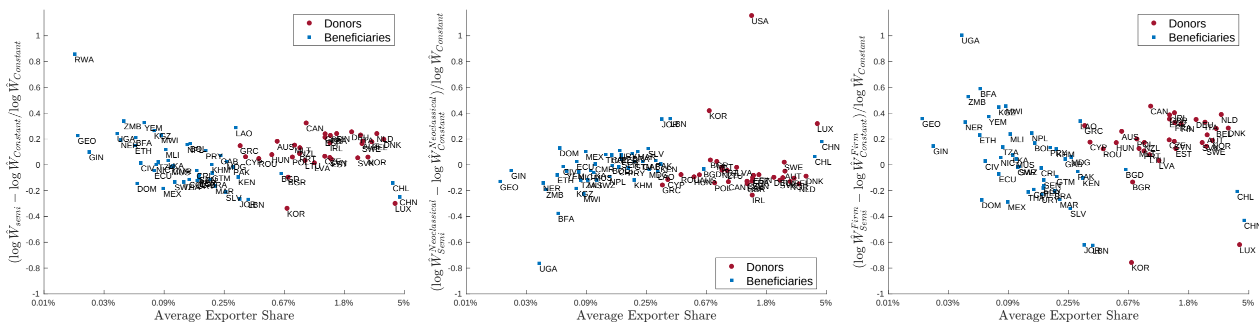
Table OA.6: Impact of Heterogeneous Reductions in Trade Costs on Welfare and its Components

Group of Countries	Welfare Elasticity (×100)	Contribution to Welfare Elasticity				
		Neoclassical Components			Firm Components	
		Technology	Terms of trade	Substitution	Entry	Selection
All	0.31	97.1 %	2.5 %	0.9 %	13.1 %	-13.6 %
Donors	0.28	157.6 %	-61.3 %	1.1 %	-16.6 %	19.1 %
Beneficiaries	0.29	0 %	120.6 %	0.6 %	68.9 %	-90.1 %

Note. Starting from the observed equilibrium in 2012, we compute the counterfactual equilibrium implied by a reduction of 1% in bilateral trade costs, i.e. $\tau_{ij} = 0.99$ for all $i \neq j$, from developing origins in the GSP list of the developed destinations in our sample conceding preferential treatment under GSP rules. For each group of countries, the second column of each panel reports 100 times the average log-change in real wage, weighted by each country's aggregate expenditure in 2012 and normalized by the shock size of 0.01. The remaining columns report the average of each component in (24) divided by the value reported in the second column. Counterfactual predictions computed with Proposition 3.b, given the elasticity estimates reported in Figure 4.

Figure OA.32: Impact of Reducing the Cost of Exporting from Developing to Developed Countries on Welfare and its Components: The Role of Parametric Assumptions

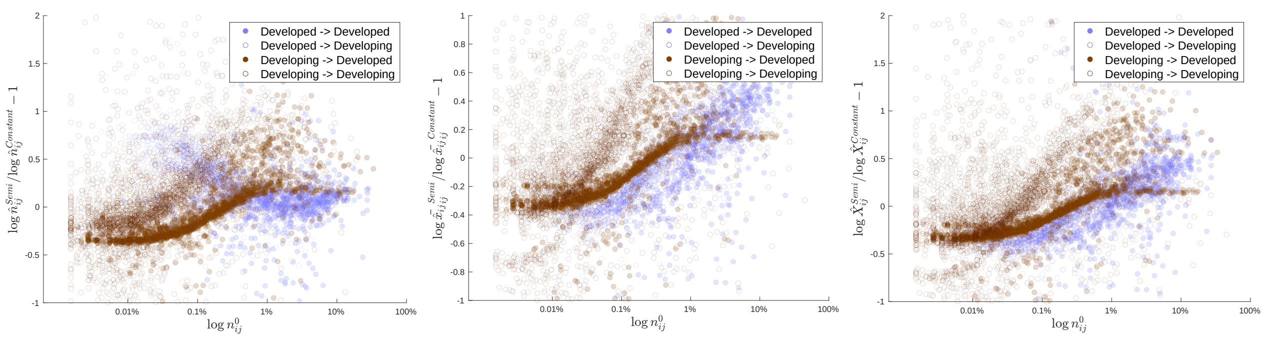
(a) Welfare Change (b) Neoclassical Components (c) Firm Components



Note. We consider the impact of a reduction of 1% in bilateral trade costs from developing countries in the GSP list to developed countries that concede preferential treatment to countries in the GSP list. Panel (a) reports in the vertical axis is the difference in welfare responses predicted by the semiparametric and constant-elasticity specifications for each country, divided by the welfare response implied by the constant-elasticity benchmark, and the horizontal axis is the log of the average exporter share of that country in 2012. The other two panels report analogous scatter plots, but the vertical axis is instead the difference in components of predicted welfare responses, divided by the overall welfare response implied by the constant-elasticity benchmark. Panel (b) does this for the sum of the neoclassical components associated with technology, terms of trade, and demand substitution in (24), and panel (c) for the sum of the firm components associated with entry and selection in (24). Semiparametric and constant-elasticity predictions use the elasticity estimates in Figures 4 and OA.9, respectively. We report the same ranges for all figures and omit the United States in Panel (a), as it extends below the displayed range as the denominator is very close to zero.

Figure OA.33: Impact of Reducing the Cost of Exporting from Developing to Developed Countries on Firm Export Margins: The Role of Parametric Assumptions

(a) Extensive Margin (b) Intensive Margin (c) Bilateral Trade Flows



Note. We consider the impact of a reduction of 1% in bilateral trade costs from developing countries in the GSP list to developed countries that concede preferential treatment to countries in the GSP list. The figure reports in the vertical axis is the ratio of the log-change of each margin of firm exports for an origin-destination pair predicted by the semiparametric and constant-elasticity specifications, and the horizontal axis is the log of the firm exporter share in 2012 for that origin-destination. Panel (a) does this for the extensive margin ($\ln \hat{n}_{ij}$), panel (b) for the intensive margin ($\ln \hat{\bar{x}}_{ij}$), and panel (c) for bilateral trade flows ($\ln \hat{X}_{ij}$). Semiparametric and constant-elasticity predictions use the elasticity estimates in Figures 4 and OA.9, respectively.