Not-Parametric Gravity:

Measuring the Aggregate Implications of Firm Heterogeneity*

Rodrigo Adão
Chicago Booth & NBER
rodrigo.adao@chicagobooth.edu

Costas Arkolakis
Yale & NBER
costas.arkolakis@yale.edu

Sharat Ganapati
Georgetown
sg1390@georgetown.edu

April 23, 2020

Abstract

How does firm heterogeneity affect the aggregate consequences of international trade shocks? In the workhorse monopolistic competition model, we show that the distribution of firm fundamentals affects aggregate equilibrium outcomes only through the shape of two univariate functions of the exporter firm share. These functions determine semiparametric gravity equations for the extensive and intensive margins of firm exports, yielding bilateral elasticities of trade flows to trade costs that vary with the exporter firm share. We show that the shape of these elasticity functions is sufficient to compute (i) counterfactual changes in aggregate outcomes and (ii) expressions for welfare gains. We estimate these elasticity functions using the model-implied semiparametric gravity equations of firm exports. Our estimates imply that bilateral trade is less sensitive to trade shocks when the exporter firm share is high. Firm heterogeneity leads to a 15% change in the gains from trade (compared to the constant elasticity gravity benchmark) that are higher in countries with a higher exporter firm share.

*We thank Chang-Tai Hsieh, Sam Kortum, and Steve Redding for valuable comments. Part of this research was conducted while Ganapati was a Fellow at Dartmouth College. All errors are our own.

1 Introduction

The international trade field has been transformed by the study of firm heterogeneity. The cross-firm correlation between observable attributes and international trade performance is one of the most robust and celebrated findings in the field. It became the cornerstone of the workhorse monopolistic competition model of firm heterogeneity that links firm-level decisions to aggregate outcomes – see Melitz (2003) and, for a review, Melitz and Redding (2014). In contrast to the flexibility of the workhorse model, its use for both empirical and counterfactual analyses in general equilibrium relies on restrictive parametric assumptions about the distribution of firm fundamentals. These assumptions yield a tractable and parsimonious framework, but have the cost of strongly restricting how firm heterogeneity affects the predictions of the model (Arkolakis et al., 2012; Melitz and Redding, 2014, 2015; Bas et al., 2017; Fernandes et al., 2017). In this paper, we show that parametric distributional assumptions are not essential for empirical and counterfactual analyses based on the workhorse model of firm heterogeneity. We then theoretically and empirically assess the importance of firm heterogeneity for the impact of trade shocks on aggregate outcomes.

We consider a monopolistic competition model with constant elasticity of substitution (CES) preferences. Firms are heterogeneous with respect to shifters of productivity, demand, and variable and fixed trade costs across destinations. An extensive literature has imposed different parametric distributional assumptions on these dimensions of firm heterogeneity to flexibly match observed cross-firm variation in productivity, sales, and entry in different markets – e.g., Chaney (2008), Helpman et al. (2008), Eaton et al. (2011), Redding (2011), Head et al. (2014).[1] We instead do not impose any restrictions on the distribution of these firm-level fundamentals. We thus only acknowledge that such heterogeneity exists and evaluate how it affects the response of aggregate outcomes to trade shocks.

We show that all these sources of firm heterogeneity can be folded into two semiparametric gravity equations for firm-level entry (i.e., the exporter firm share) and per-firm sales (i.e., the average firm exports). First, the extensive margin gravity equation links a function of the exporter firm share to bilateral trade costs, and origin and destination fixed-effects. Second, the intensive margin gravity equation connects the average firm exports to a function of the exporter firm share, as well as bilateral trade costs and origin and destination fixed-effects. As in existing gravity models of trade, the fixed-effects contain endogenous variables (i.e., wages and price indices). Furthermore, we establish that the two functions in the gravity equations can be used to summarize all dimensions of firm heterogeneity in the conditions

[1] As in Eaton et al. (2011), the multiple firm-specific shifters that vary across destinations allow the model to flexibly rationalize imperfect correlation between sales and entry across markets.

determining aggregate outcomes in equilibrium.[2]

In this environment, the elasticity of bilateral trade flows to changes in bilateral trade costs, or simply the aggregate trade elasticity, is not constant and may vary arbitrarily with the exporter firm share. This result follows from the fact that the trade elasticity is the sum of the extensive and intensive margin elasticities of firm exports (as determined by the two semiparametric gravity equations). The only role of the distribution of firm fundamentals is determining how the different margins of the trade elasticity vary with the exporter firm share. In fact, parametric assumptions in the literature only affect equilibrium outcomes insofar they restrict how the trade elasticity varies with the exporter firm share. For instance, all trade elasticity margins are constant if the productivity distribution is Pareto (Chaney, 2008) and are decreasing if the productivity distribution is Truncated Pareto (Melitz and Redding, 2015) or Log-normal (Head et al., 2014).

We then show how to compute nonparametric counterfactual predictions in response to shocks in bilateral trade costs. We first extend the "exact hat algebra" technique in Dekle et al. (2008) to compute counterfactual changes in aggregate outcomes using (i) bilateral data on exporter firm shares and aggregate trade flows in the initial equilibrium, (ii) the CES elasticity of substitution, and (iii) the elasticity functions of the extensive and intensive margins of firm exports. This result establishes that, given the initial equilibrium, the different sources of firm heterogeneity, and any associated parametric assumption imposed, only matter for the model's counterfactual predictions through the shape of the two elasticity functions of firm exports.

We further establish that the importance of firm heterogeneity for the model's counterfactual predictions depends on how the aggregate trade elasticity varies with the exporter firm share. For small changes in trade costs, responses of aggregate outcomes depend solely on the elasticity matrix of bilateral trade flows. The elasticity functions of the firm export margins simply determine the bilateral trade elasticities given the initial exporter firm shares. For large changes in trade costs, bilateral trade elasticities change along the adjustment path to the new equilibrium due to changes in exporter firm shares. Thus, in this case, counterfactual predictions depend on the two elasticity functions of firm exports as they determine the changes in exporter firm shares and, therefore, the changes in aggregate trade elasticities along the adjustment path. In fact, when both elasticity functions are constant, the bilateral trade elasticity is also constant and, therefore, the model's predictions are isomorphic to those

[2] In deriving the semiparametric gravity equations, we rely on the type of inversion argument used by Berry and Haile (2014) and Adão (2015). All variables can be written in terms of the set of firms operating in each country pair, which has a one-to-one mapping with the exporter firm share through the entry condition. We then re-write all variables in the model as a function of the exporter firm share, with functions determined by the economy's distribution of firm fundamentals.

of the class of constant-elasticity gravity trade models analyzed by Arkolakis et al. (2012).

To conclude our theoretical analysis of the aggregate implications of firm heterogeneity, we derive a set of nonparametric ex-post sufficient statistics that link welfare gains to changes in domestic trade outcomes and domestic trade elasticities (as determined by the initial share of active domestic firms). Our derivations are extensions of related sufficient statistics in the literature – e.g., Arkolakis et al. (2012) and Melitz and Redding (2015). In particular, we derive an extension of the formula in Arkolakis et al. (2012) for small shocks in which the welfare gain combines the initial trade elasticity and changes in the domestic spending share and number of domestic entrants. For large shocks though, welfare gains must account for the correlation between changes in the trade elasticity and changes in the domestic spending share and firm entry. In the model, this correlation is a function of the two elasticity functions of firm exports.

Overall, our theoretical results synthesizes the debate for the role of firm heterogeneity. For small changes in trade costs, our results indicate the aggregate trade elasticity is sufficient to characterize aggregate outcomes, in line with Arkolakis et al. (2012). However, for large changes in trade costs, changes in the trade elasticity and the number of entrants are key, in line with Melitz and Redding (2015) and Feenstra (2018). Our results allow us to compute such changes conditional on knowledge of the two elasticity functions governing the extensive and intensive margins of firm exports.

Given their importance for counterfactual predictions, the second part of the paper develops a methodology for estimating the elasticity functions of bilateral trade and firm exports. We do so by directly estimating how observed trade shocks affect the different margins of trade flows between countries. Formally, we estimate the elasticity functions using the semiparametric gravity equations implied by our model for the extensive and intensive margins of firm exports. Accordingly, our estimation strategy does not impose any parametric restriction on the distribution of firm heterogeneity in each country.

Our methodology relies on cross-country variation in exporter firm share and average firm revenue induced by observed shifters of bilateral trade costs (conditional on origin and destination fixed-effects). It requires two main assumptions. First, different groups of country pairs must have the same trade elasticity functions. Second, observed cost shifters must satisfy the same set of assumptions necessary for the consistent estimation of constant-elasticity gravity models – for a review, see Head and Mayer (2013). Following our theoretical results, the empirical methodology extends the estimation of standard gravity equations by specifying each margin's elasticity to trade costs as a flexible function of the observed exporter firm share.

We implement our estimation methodology using a sample of exporter-importer pairs for which we have exporter firm shares and average firm exports in 2012. We find that the "average" aggregate trade elasticity is five – similar to estimates reviewed by Costinot and Rodriguez-Clare (2013). This average trade elasticity combines responses in both the extensive and the intensive margins of firm exports. Importantly, our estimates indicate that the different margins of bilateral trade responses vary with the exporter firm share. The extensive margin becomes more responsive as more firms serve a market – in line with the evidence in Kehoe and Ruhl (2013a). The intensive margin elasticity has an inverted "U" shape on the exporter firm share. Together, these estimates imply that the bilateral trade elasticity varies between three and eight depending on the level of the exporter firm share, being typically higher when the exporter firm share is lower.

We conclude the paper by revisiting the question: How large are the gains from trade? We first measure the importance of firm heterogeneity for the gains from trade. We do so with a comparison between the gains from trade implied by our semiparametric estimates of the gravity equations of firm exports and a benchmark constant-elasticity gravity model of bilateral trade flows. As discussed above, firm heterogeneity does not matter in this constant-elasticity benchmark since the gains from trade are given by the sufficient statistic formula of Arkolakis et al. (2012). The gains implied by our baseline estimates have a 15% average absolute difference with respect to the gains implied by the constant elasticity gravity benchmark. For some countries, the impact of firm heterogeneity on the gains from trade may be substantial, generating gains that are 20% higher or lower. In fact, firm heterogeneity amplifies the gains from trade in countries with higher exporter firm shares. This additional statistic explains most of the differences in gains from trade implied by our semiparametric specification and the constant-elasticity benchmark.

We then evaluate the quantitative importance of measuring firm heterogeneity with the elasticity functions governing the extensive and intensive margins of firm exports. In this case, we compare our baseline estimates of the gains from trade to those implied by alternative parametric methodologies that estimate the distribution of firm fundamentals by matching cross-sectional dispersion in firm-level outcomes. Compared to the assumptions of Log-normal productivity in Head et al. (2014) and Truncated Pareto productivity in Melitz and Redding (2015), the average absolute differences with respect to our baseline estimates are 39% and 35%, respectively.

Our paper is related to an extensive literature using variations of the framework in Melitz (2003) together with parametric distributional assumptions to conduct empirical and counterfactual analyses. For instance, several papers impose that the firm productivity distribution belongs either to the Pareto family (Chaney, 2008; Arkolakis et al., 2008; Arkolakis, 2010; Bernard et al., 2011), the Truncated Pareto family (Helpman et al., 2008;

Melitz and Redding, 2015), or the Log-normal family (Head et al., 2014; Bas et al., 2017). Other papers impose distributional assumptions on multiple dimensions of firm heterogeneity (Eaton et al., 2011; Fernandes et al., 2017). These papers then estimate the distribution's parameters by matching cross-section variation in firm-level outcomes. Our results show that this parametric approach effectively extrapolates from observed cross-firm heterogeneity to put discipline on the two main elasticity functions that control the model's counterfactual predictions.[3] We instead propose a methodology to directly estimate these two key elasticity functions using the semiparametric gravity equations of firm exports implied by the model. Our empirical results yield welfare gains from trade that differ by 35%–39% from those obtained with different versions of the parametric approach in the literature.

Our empirical analysis relies on two semiparametric gravity equations of firm exports. It is thus related to the literature estimating extensions of the log-linear gravity equation of bilateral trade flows – e.g., Novy (2013), Fajgelbaum and Khandelwal (2016), and Lind and Ramondo (2018). Our framework yields a bilateral trade elasticity that varies with the firm export share. Trade elasticities that vary with trade openness also arise in monopolistic competitive models with variable markups – e.g., Melitz and Ottaviano (2008), Feenstra and Weinstein (2017), Feenstra (2018), and Arkolakis et al. (2019a).[4]

Finally, we contribute to a recent literature focusing on nonparametric counterfactual analysis in international trade models (Adao et al., 2017; Bartelme et al., 2019). Counterfactual predictions in these settings require knowledge of multivariate functions whose nonparametric estimation is challenging in finite samples – for example, Adao et al. (2017) must estimate a country's demand function for all factors in the world economy. Compared to these papers, we consider a different class of models featuring monopolistic competition, CES preferences, and firm heterogeneity. In this environment, we show that counterfactual analysis depends only on two univariate functions of the exporter firm share. These elasticity functions can be easily estimated using two semiparametric gravity equations of firm exports.

Our paper is organized as follows. Section 2 derives the semiparametric gravity equations for the extensive and intensive margins of firm exports. In Section 3, we present our nonparametric counterfactual analysis. Section 4 outlines the methodology to estimate the two main elasticity functions in the model. In Section 5, we report our baseline estimation results. Section 6 conducts counterfactual exercises. Section 7 concludes.

[3] Fernandes et al. (2017) show counterfactuals can be conducted for any given distribution. Parametrization of this specific distribution is ultimately needed to both run counterfactuals and estimate observed firm heterogeneity.

[4] In an extension of our baseline framework, we consider endogenous markup changes implied by a demand function with a single price aggregator, as in Matsuyama and Ushchev (2017) and Arkolakis et al. (2019a). Our setting covers the symmetric separable utility in Krugman (1979) and Zhelobodko et al. (2011). We derive semiparametric gravity equations of firm exports and show how to use them in counterfactual analysis.

2 Model

We consider an economy in which firms are heterogeneous in terms of productivity, demand, and trade costs. The equilibrium of this economy entails two semiparametric gravity equations for the extensive and intensive margins of firm exports. In general equilibrium, the functions in these two gravity equations along with country-level fundamentals determine trade flows, firm entry, price indices, and wages.

2.1 Environment

Demand. Each country j has a representative household that inelastically supplies $\bar{L}_j$ units of labor. In each country, the representative household has Constant Elasticity of Substitution (CES) preferences over the continuum of available varieties, ω:

$$U_j = \left(\sum_i \int_{\omega \in \Omega_{ij}} \left(\bar{b}_{ij} b_{ij}(\omega)\right)^{\frac{1}{\sigma}} \left(q_{ij}\left(\omega\right)\right)^{\frac{\sigma-1}{\sigma}} \; d\omega \right)^{\frac{\sigma}{\sigma-1}}, \qquad \sigma > 1,$$

where Ω_{ij} is the set of varieties from i available in j.

The utility maximization problem in country j determines the demand for variety ω from country i:

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

where, 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,

$$P_j^{1-\sigma} = \sum_i \int_{\Omega_{ij}} \left(\bar{b}_{ij} b_{ij}(\omega)\right) \left(p_{ij}\left(\omega\right)\right)^{1-\sigma} \; d\omega. \tag{2}$$

The ω-specific demand shifters, $b_{ij}(\omega)$, allows the model to match sales across varieties conditional on prices. The term $\bar{b}_{ij}$ is the component of bilateral taste shifters that is common to all varieties.

Production. Each variety is produced by a single firm, so we refer to a variety as a firm-specific good. The production function implies that, in order to sell q units in country j, firm ω from country i incurs in a labor cost of

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

The first term is the variable cost of selling q units in country j, including both a firm-specific iceberg shipping cost, $\bar{\tau}_{ij}\tau_{ij}(\omega)$, and productivity, $\bar{a}_i a_i(\omega)$. The second term is the fixed cost of labor necessary to enter j. As in Melitz (2003), we specify the fixed entry cost in terms of labor in the origin country. However, we depart from Melitz (2003) by allowing firms to be different not only in their productivity, but also in their variable and fixed costs of exporting. Eaton et al. (2011) show that these additional sources of firm heterogeneity are important for the model to match observed patterns of firm-level exports to different countries.

We consider a monopolistic competitive environment in which firms maximize profits given the demand in (1). For firm ω of country i, the optimal price in market j is $p_{ij}(\omega) = \frac{\sigma}{\sigma-1}\frac{\bar{\tau}_{ij}w_i}{\bar{a}_i}\frac{\tau_{ij}(\omega)}{a_i(\omega)}$ with an associated revenue of

$$R_{ij}(\omega) = \bar{r}_{ij} r_{ij}(\omega) \left[\left(\frac{w_i}{P_j} \right)^{1-\sigma} E_j \right], \tag{3}$$

where

$$r_{ij}(\omega) \equiv b_{ij}(\omega) \left(\frac{\tau_{ij}(\omega)}{a_i(\omega)} \right)^{1-\sigma} \quad \text{and} \quad \bar{r}_{ij} \equiv \bar{b}_{ij} \left(\frac{\sigma}{\sigma-1} \frac{\bar{\tau}_{ij}}{\bar{a}_i} \right)^{1-\sigma}. \tag{4}$$

We refer to $r_{ij}(\omega)$ as the *revenue potential* in j of firm ω from i. Conditional on entering market j, it is the firm-specific sales shifter in j that combines different sources of firm heterogeneity into a single term.

The firm's entry decision depends on the profit generated by the revenue in (3), $(1/\sigma)R_{ij}(\omega)$, and the fixed-cost of entry, $w_i \bar{f}_{ij} f_{ij}(\omega)$. Specifically, firm ω of i enters j if, and only if, $\pi_{ij}(\omega) = \frac{1}{\sigma}R_{ij}(\omega) - w_i \bar{f}_{ij} f_{ij}(\omega) \geq 0$. This yields the set of firms from i selling in j:

$$\omega \in \Omega_{ij} \quad \Leftrightarrow e_{ij}(\omega) \geq \sigma \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right], \tag{5}$$

where

$$e_{ij}(\omega) \equiv \frac{r_{ij}(\omega)}{f_{ij}(\omega)}. \tag{6}$$

We refer to $e_{ij}(\omega)$ as the *entry potential* of firm ω of i in j. Among firms with identical revenue potential, heterogeneity in the fixed-cost of entry generates heterogeneity in entry potentials and, therefore, in decisions to enter different destination markets. The difference between revenue and entry potentials of firms allows for imperfect cross-firm correlation between entry and sales across markets.

Entry. Firms in country i create a new variety by hiring $\bar{F}_i$ units of domestic labor. In this case, they take a draw of their variety characteristics from an arbitrary distribution:

$$v_i(\omega) \equiv \{a_i(\omega), b_{ij}(\omega), \tau_{ij}(\omega), f_{ij}(\omega)\}_j \sim G_i(v). \tag{7}$$

In equilibrium, free entry implies that N_i firms pay the fixed cost of entry in exchange for an ex-ante expected profit of zero,

$$\sum_j E\left[\max\{\pi_{ij}(\omega);\ 0\}\right] = w_i \bar{F}_i. \tag{8}$$

Market clearing. We follow Dekle et al. (2008) by introducing exogenous international transfers, so that spending is

$$E_i = w_i \bar{L}_i + \bar{T}_i, \quad \sum_i \bar{T}_i = 0. \tag{9}$$

Since labor is the only factor of production, labor income in i equals the total revenue of firms from i: $w_i L_i = \int_{\omega \in \Omega_{ij}} R_{ij}(\omega)\, d\omega$. Given the expression in (3),

$$w_i \bar{L}_i = \bar{r}_{ij} \left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j \left[\int_{\omega \in \Omega_{ij}} r_{ij}(\omega)\, d\omega\right]. \tag{10}$$

Equilibrium. Given the arbitrary distribution in (7), the equilibrium is defined as the vector $\{P_i, \{\Omega_{ij}\}_j, N_i, E_i, w_i\}_i$ satisfying equations (2), (5), (8), (9), (10) for all i.

2.2 Extensive and Intensive Margin of Firm Exports

We now use the definitions of entry and revenue potentials to characterize firm-level entry and sales in different markets in general equilibrium. We consider the CDF of $(r_{ij}(\omega), e_{ij}(\omega))$ implied by $G_i(v)$. Without loss of generality, this CDF can be decomposed as

$$r_{ij}(\omega) \sim H^r_{ij}(r|e), \qquad \text{and} \qquad e_{ij}(\omega) \sim H^e_{ij}(e). \tag{11}$$

Intuitively, firms draw their entry potential e from $H^e_{ij}(e)$. Conditional on having an entry potential of e, firms draw their revenue potential r from $H^r_{ij}(r|e)$. We impose the following regularity condition on the distribution of entry potentials.

Assumption 1. *Assume that $H^e_{ij}(e)$ is continuous and strictly increasing in $\mathbb{R}_+$ with $\lim_{e\to\infty} H^e_{ij}(e) = 1$.*

This assumption implies that that H_{ij}^e has full support in $\mathbb{R}_+$ and no mass points. These restrictions guarantee that any change in trade costs induces a positive mass of firms to switch their entry decisions. As described below, this is central for the change of variables necessary for our characterization of the equilibrium.[5]

Our specification allows for any pattern of heterogeneity in $(r_{ij}(\omega), e_{ij}(\omega))$, permitting any correlation between entry and revenue potentials. It encompasses several distributional assumptions in the literature. 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 case, we can specify the distribution of e_{ij} 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). Independent of the distributional assumption, in this case, the single source of heterogeneity implies a strict hierarchy of entry across destinations and a perfect cross-firm correlation between intensive and extensive margins of exports.

In addition, multiple papers incorporate additional sources of heterogeneity across firms that yields dispersion in both $r_{ij}(\omega)$ and $e_{ij}(\omega)$. For example, the demand and entry cost heterogeneity in Eaton et al. (2011) are modeled so that $a_i(\omega)$ is Pareto distributed while $b_{ij}(\omega)$ and $f_{ij}(\omega)$ are joint log-normally distributed. Arkolakis et al. (2019b) consider a further layer of product-firm heterogeneity that combines the assumption of Eaton et al. (2011) with functional forms governing the sales from each additional product. Fernandes et al. (2017) allow for a multivariate log-normal distribution of productivity shifters of sales across destinations.

We now focus on the share of firms from i selling in j, $n_{ij} = Pr\left[\omega \in \Omega_{ij}\right]$, and their average sales, $\bar{x}_{ij} \equiv E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right]$.

Extensive margin of firm-level exports. Given the assumption that H_{ij}^e has full support, it is possible to define the inverse distribution of entry potential: $\bar{\epsilon}_{ij}(n) \equiv \left(H_{ij}^e\right)^{-1}(1-n)$ where $\bar{\epsilon}_{ij}(n)$ is strictly decreasing, $\bar{\epsilon}_{ij}(1) = 0$, and $\lim_{n\to 0} \bar{\epsilon}_{ij}(n) = \infty$. In any equilibrium, expression (5) yields

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

This expression is a semiparametric gravity equation relating a function of the share of firms from i selling in j to a log-linear combination of exogenous bilateral trade shifters and endogenous outcomes in the origin and destination markets. We refer to $\bar{\epsilon}_{ij}(n)$ as the

[5] This assumption also implies positive trade flows between all origin-destination pairs. However, this is not essential. In Section 3.4, we allow for the possibility of zero bilateral trade by imposing that there exists $\bar{e}_{ij} < \infty$ such that $H_{ij}^e(\bar{e}_{ij}) = 1$.

extensive margin elasticity function of firm-level exports as it controls the sensitivity of the share of exporters in a market to changes in bilateral trade costs (holding constant other endogenous variables):

$$\frac{\partial \ln n_{ij}}{\partial \ln \bar{\tau}_{ij}} = \frac{\sigma - 1}{\varepsilon_{ij}(n_{ij})} \quad \text{such that} \quad \varepsilon_{ij}(n_{ij}) \equiv \left. \frac{\partial \ln \bar{\epsilon}_{ij}(n)}{\partial \ln n_{ij}} \right|_{n=n_{ij}}. \tag{13}$$

The extensive margin elasticity to trade costs is negative since $\sigma > 1$ and $\bar{\epsilon}_{ij}(n)$ is decreasing in n. Importantly, the model does not impose any restriction on the shape of $\bar{\epsilon}_{ij}(n)$, implying that the extensive margin elasticity may be locally increasing or decreasing on n_{ij}. We show below that parametric distributional assumptions restrict the relationship between the extensive margin elasticity and the share of firms of i exporting to j, n_{ij}. In the rest of the paper, we refer to n_{ij} simply as the exporter firm share.

Intensive margin of firm-level exports. We define the average revenue potential when a share n_{ij} of $i's$ firms sell in j as

$$\bar{\rho}_{ij}\left(n_{ij}\right) \equiv \frac{1}{n_{ij}} \int_0^{n_{ij}} \rho_{ij}(n)\ dn \tag{14}$$

where $\rho_{ij}(n) \equiv E[r|e = \bar{\epsilon}_{ij}(n)]$ is the average revenue potential in quantile n of the entry potential distribution.

The definition of sales per firm, $\bar{x}_{ij} \equiv E\left[R_{ij}\left(\omega\right)|\omega \in \Omega_{ij}\right]$, implies that

$$\bar{x}_{ij} = \bar{r}_{ij} \left[\left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j \right] \int_{e^*_{ij}}^{\infty} E\left[r|e\right] \frac{dH^e(e)}{1 - H^e(e^*_{ij})}, \quad e^*_{ij} \equiv \sigma \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j}\right)^{\sigma} \frac{P_j}{E_j} \right].$$

We consider the transformation $n = 1 - H^e_{ij}(e)$ such that $e = \bar{\epsilon}_{ij}(n)$ and $dH^e_{ij}(e) = -dn$. Since $1 - H^e_{ij}(e^*_{ij}) = n_{ij}$ and $\lim_{e\to\infty} H^e_{ij}(e) = \infty$,

$$\ln \bar{x}_{ij} - \ln \bar{\rho}_{ij}(n_{ij}) = \ln\left(\bar{r}_{ij}\right) + \ln\left(w_i^{1-\sigma}\right) + \ln\left(E_j P_j^{\sigma-1}\right). \tag{15}$$

This expression provides our second semiparametric gravity equation. It relates a measure of the composition-adjusted per-firm sales to a linear combination of exogenous bilateral trade shifters and endogenous outcomes in the origin and destination markets. We refer to $\bar{\rho}_{ij}(n)$ as the intensive margin elasticity function of firm-level exports as it controls the sensitivity of average per-firm sales to changes in bilateral trade costs (holding constant other endogenous variables):

$$\frac{\partial \ln \bar{x}_{ij}}{\partial \ln \bar{\tau}_{ij}} = -(\sigma-1) + \varrho_{ij}(n_{ij}) \frac{\sigma-1}{\varepsilon_{ij}(n_{ij})} \quad \text{such that} \quad \varrho_{ij}(n_{ij}) \equiv \left. \frac{\partial \ln \bar{\rho}_{ij}(n)}{\partial \ln n_{ij}} \right|_{n=n_{ij}}. \tag{16}$$

This elasticity combines two well-known forces. The first term is the reduction in the sales of the initial set of exporters in j arising from the constant elasticity of substitution across varieties. The second term measures how the change in the number of exporters affects the average revenue potential of firms selling in j. The sign of this term depends on how different marginal and infra-marginal exporters are in terms of revenue potential. Specifically,

$$\varrho_{ij}(n_{ij}) = \frac{\rho_{ij}(n_{ij})}{\bar{\rho}_{ij}(n_{ij})} - 1. \tag{17}$$

Notice that $\varrho_{ij}(n_{ij}) < 0$ if, and only if, $\rho_{ij}(n_{ij}) < \bar{\rho}_{ij}(n_{ij})$. In other words, $\bar{\rho}_{ij}(n)$ is decreasing in n whenever the average revenue potential of marginal exporters, $\rho_{ij}(n_{ij})$, is worse than that of infra-marginal exporters, $\bar{\rho}_{ij}(n_{ij})$.[6] Similarly, if marginal firms are better than infra-marginal firms, then $\rho_{ij}(n_{ij}) > \bar{\rho}_{ij}(n_{ij})$ and $\bar{\rho}_{ij}(n)$ is increasing in n. Since revenue potential is positive, $\rho_{ij}(n) > 0$ and $\varrho_{ij}(n_{ij}) > -1$.

One important feature of the model is that the intensive margin elasticity in (16) is a function of the exporter firm share, n_{ij}. Parametric assumptions on the distribution of firm fundamentals yield specific patterns of dependence between the intensive margin elasticity and the exporter firm share. We return to this point below.

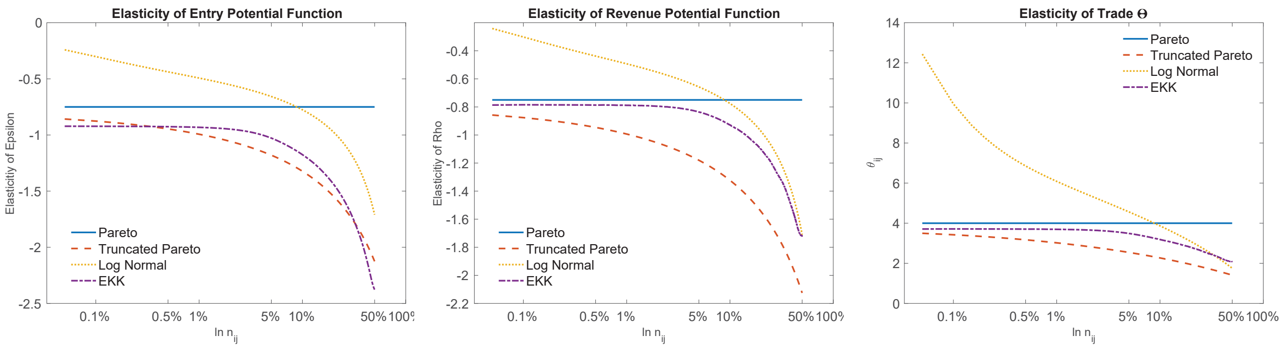
Bilateral trade flows. Bilateral trade flows combine the extensive and intensive margins of firm exports: $X_{ij} \equiv N_i n_{ij} \bar{x}_{ij}$. Thus, $\bar{\epsilon}_{ij}(n_{ij})$ and $\bar{\rho}_{ij}(n_{ij})$ determine the elasticity of bilateral trade flows to changes in bilateral trade costs (holding constant other endogenous variables):

$$\theta_{ij}(n_{ij}) \equiv -\frac{\partial \ln X_{ij}}{\partial \ln \bar{\tau}_{ij}} = -\left(\frac{\partial \ln n_{ij}}{\partial \ln \bar{\tau}_{ij}} + \frac{\partial \ln \bar{x}_{ij}}{\partial \ln \bar{\tau}_{ij}} \right) = (\sigma-1)\left(1 - \frac{1+\varrho_{ij}(n_{ij})}{\varepsilon_{ij}(n_{ij})}\right). \tag{18}$$

This expression indicates that n_{ij} acts like a state variable that determines the elasticity of bilateral trade flows to changes in trade costs – the so-called trade elasticity. This occurs because n_{ij} controls the elasticity of the extensive and intensive margins of firm exports. Notice that the trade elasticity is positive for all n_{ij} since $\varepsilon_{ij}(n_{ij}) < 0$ and $\varrho_{ij}(n_{ij}) > -1$.

[6] In Melitz (2003), the single source of firm heterogeneity ($r_{ij}(\omega) = e_{ij}(\omega)$) implies that $\rho_{ij}(n) = \bar{\epsilon}_{ij}(n)$. In this case, marginal exporters are worse than existing infra-marginal exporters since $\frac{\partial \bar{\rho}_{ij}(n)}{\partial n} = \frac{1}{n^2} \int_0^n \left(\bar{\epsilon}_{ij}(n) - \bar{\epsilon}_{ij}(x) \right) dx < 0$ and $\bar{\epsilon}_{ij}(n) < \bar{\epsilon}_{ij}(x)$ for all $x < n$.

Figure 1: Distributional assumptions and Elasticity of different margins of trade flows



Note. Left panel reports the elasticity of $\bar{\epsilon}_{ij}(n)$. Center panel reports the elasticity of $\bar{\rho}_{ij}(n)$. Right panel reports the trade elasticity $\theta_{ij}(n)$ in (18). Each line corresponds to the elasticity as a function of n implied by different parametric restrictions on the distribution of firm fundamentals. See main text for a description of each parametrization.

Distributional assumptions and elasticity of trade flows. We now show that distributional assumptions on firm fundamentals restrict how the different margins of bilateral trade responses vary with the exporter firm share. In Figure 1, we depict the elasticity functions implied by productivity distributions from the Pareto family (Chaney, 2008), the truncated Pareto family (Melitz and Redding, 2015), and the log-normal family (Head et al., 2014). We also consider the specification in Eaton et al. (2011) where productivity has a Pareto distribution and shifters of demand and entry costs have a joint log-normal distribution. In all cases, we use the baseline parameters reported in each paper.

The first plot indicates that the Pareto assumption yields constant elasticities of all margins. The other parameterizations yield a declining elasticity of $\bar{\epsilon}(n)$, which, by equation (13), implies that the extensive margin elasticity is more sensitive when the exporter firm share is low. Similarly, all other parameterizations yield a declining elasticity of $\bar{\rho}(n)$, indicating that new entrants and incumbents are more similar to each other when n_{ij} is small. This implies that composition effects are weaker when few firms export to a particular destination. The third panel combines these two margins to show that the trade elasticity is higher when n_{ij} is low. In all parametrizations, the trade elasticity falls below two when n_{ij} is above 50%. We show below that this implies a low elasticity in the domestic market where n_{ii} is high.

2.3 General Equilibrium

We use the functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ to characterize all aggregate variables in equilibrium.

Lemma 1. *Suppose Assumption 1 holds. Given* $\left\{\bar{T}_i, \bar{L}_i, \bar{F}_i, \{\bar{r}_{ij}, \bar{f}_{ij}\}_j\right\}_i$, *an equilibrium vector* $\{\{n_{ij}, \bar{x}_{ij}\}_j, P_i, N_i, E_i, w_i\}_i$ *satisfies the following conditions.*

1. The extensive and intensive margins of firm-level sales, n_{ij} *and* $\bar{x}_{ij}$*, satisfy (12) and (15) for all* i *and* j*. Together with* N_i*, they determine bilateral trade flows,* $X_{ij} = N_i n_{ij} \bar{x}_{ij}$*.*

2. For all i, total spending, E_i, satisfies (9).

3. For all i, the labor market clears,

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

4. For all j, the price index is given by

$$P_j^{1-\sigma} = \sum_i \bar{r}_{ij} w_i^{1-\sigma} \bar{\rho}_{ij}(n_{ij}) n_{ij} N_i. \tag{20}$$

5. For all i, the number of entrants is

$$N_i = \left[\sigma \frac{\bar{F}_i}{\bar{L}_i} + \sum_j \frac{n_{ij}\bar{x}_{ij}}{w_i \bar{L}_i} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})} \, dn} \right]^{-1}. \tag{21}$$

Proof. See Appendix A.1.

The lemma shows that all aggregate variables in equilibrium only depend on firm heterogeneity through the functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ – by (14), $\bar{\rho}_{ij}(n)$ yields $\rho_{ij}(n)$. Intuitively, as in Melitz (2003), all aggregate variables can be written in terms of the set of firms operating in each country pair, which is fully determined by the entry cutoff in equilibrium. Through the inversion argument in Section 2.2, we establish a one-to-one mapping between the entry cutoff and the firm exporter share. This allows us to re-write all aggregate variables in equilibrium in terms of the elasticity functions $(\bar{\epsilon}_{ij}(n), \bar{\rho}_{ij}(n))$ and, consequently, in terms of the exporter firm share n_{ij}

This lemma indicates that the functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ summarize the aggregate implications of the different dimensions of firm heterogeneity in the model. Thus, any parametric restriction on the distribution of firm fundamentals affects the economy's equilibrium insofar it determines the shape of these elasticity functions. We summarize this discussion in the following remark.

Remark 1. *All dimensions of heterogeneity can be folded into the two elasticity functions, $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$, that govern the semiparametric gravity equations for the extensive and intensive margins of firm exports, (12) and (15).*

The rest of the paper exploits this insight in two ways. In the next section, we build directly on Lemma 1 by using $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ to evaluate how firm heterogeneity affects the counterfactual impact of trade shocks on welfare. In Section 4, we will exploit the semiparametric gravity equations in (12) and (15) to estimate $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$.

3 Nonparametric Counterfactual Analysis

We now investigate the response of aggregate outcomes to trade cost shocks. Our main result establishes that, given trade outcomes in the initial equilibrium, $(\sigma, \bar{\epsilon}_{ij}(n), \bar{\rho}_{ij}(n))$ are sufficient to compute ex-ante counterfactual changes of aggregate outcomes and welfare. Thus, firm heterogeneity only matters insofar it affects the functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$. We then provide ex-post sufficient statistics for welfare changes in terms of changes in endogenous trade outcomes and the elasticity functions $(\sigma, \bar{\epsilon}_{ij}(n), \bar{\rho}_{ij}(n))$.

Our results indicate that the aggregate consequences of firm heterogeneity arise only from the fact that the trade elasticity varies with the exporter firm share. Whenever the elasticities of $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ are constant, we recover the result in Arkolakis et al. (2012) that the trade elasticity is constant and sufficient for counterfactual analysis. In contrast, if $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ vary with the exporter firm share, the trade elasticity also varies with the exporter firm share. This implies that changes in the exporter firm share following the shock generate changes in the bilateral trade elasticities along the path to the new counterfactual equilibrium, affecting the adjustment of all aggregate outcomes.

3.1 Ex-ante Nonparametric Sufficient Statistics for Counterfactual Analysis

Our first result establishes what is necessary to compute counterfactual changes in aggregate outcomes given exogenous changes in variable trade costs $\bar{\tau}_{ij}$.[7] We use $\hat{y}_j \equiv y'_j/y_j$ to express changes in any variable between its level in the observed equilibrium, y_i, and the counterfactual equilibrium y'_i. We also use bold letters to denote vectors, $\boldsymbol{y} = [y_i]_i$ and bold bar variables to denote matrices, $\bar{\boldsymbol{y}} = [y_{ij}]_{i,j}$.

Proposition 1. *Consider any change in $\hat{\bar{\tau}}_{ij}$ for $i \neq j$. Given the matrices of exporter firm shares and bilateral trade flows in the initial equilibrium $(\bar{\boldsymbol{n}}, \bar{\boldsymbol{X}})$, the substitution elasticity σ and the elasticity functions $(\bar{\boldsymbol{\epsilon}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{\rho}}(\bar{\boldsymbol{n}}))$ are sufficient to characterize counterfactual changes in aggregate outcomes, $\{\hat{\bar{\boldsymbol{n}}}, \hat{\bar{\boldsymbol{X}}}, \hat{\boldsymbol{P}}, \hat{\boldsymbol{N}}, \hat{\boldsymbol{E}}, \hat{\boldsymbol{w}}\}$.*

Proof. See Appendix A.2.

The proposition follows directly from the characterization of the equilibrium in Lemma 1. It outlines two sufficient requirements to compute counterfactual changes in aggregate outcomes. First, it is not necessary to know the entire distribution of firm fundamentals in

[7] In the proof in Appendix A.2, we show that Proposition 1 also holds for changes in fixed costs $(\bar{f}_{ij})$, transfers $(\bar{T}_i)$, productivity levels $(\bar{a}_i)$, entry costs $(\bar{F}_i)$, and country sizes $(\bar{L}_i)$.

the initial equilibrium. Instead, one only needs to obtain the fraction of the firms of country i selling in each other country j (i.e., the exporter firm share n_{ij}), and the two elasticity functions of the extensive and intensive margins of firm exports, $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$. Second, it is also necessary to obtain bilateral trade flows between countries, as in the original "hat algebra" methodology in Dekle et al. (2008), and the elasticity of substitution σ, as in the "hat algebra" for heterogeneous firm models (see Costinot and Rodriguez-Clare (2013)).[8]

This result implies that the different dimensions of firm heterogeneity in the model only matter through the responses of the extensive and intensive margins of firm exports, as summarized by the functions $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$. Thus, conditional on knowing $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$, Proposition 1 implies that we can compute counterfactual changes in aggregate outcomes without imposing parametric restrictions on the distribution of firm fundamentals.

The trade elasticity, $\theta_{ij}(n_{ij})$, does not play a direct role in Proposition 1. However, as the next proposition shows, the trade elasticity is the channel through which the extensive and intensive margin elasticities affect counterfactual changes in aggregate outcomes

Proposition 2. *Consider a small trade cost shock between origin o and destination d, $d\ln\bar{\tau}_{od}$.*

1. Let $Y_i \equiv \{\{X_{ij}\}_j, P_i, N_i, E_i, w_i\}$. The elasticity of any element of Y_i to $\bar{\tau}_{od}$ is a function of $\left(\sigma, \bar{\boldsymbol{\theta}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{X}}\right)$:

$$\frac{d\ln Y_i}{d\ln\bar{\tau}_{od}} = \Psi_{i,od}\left(\sigma, \bar{\boldsymbol{\theta}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{X}}\right), \tag{22}$$

where $\theta_{ij}(n)$ is the trade elasticity function defined in (18).

2. The elasticity of the exporter firm share n_{ij} to $\bar{\tau}_{od}$ is a function of $\left(\sigma, \bar{\boldsymbol{\theta}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{X}}\right)$ and $\varepsilon_{ij}(n_{ij})$:

$$\frac{d\ln n_{ij}}{d\ln\bar{\tau}_{od}} = \tilde{\Psi}_{ij,od}\left(\sigma, \bar{\boldsymbol{\theta}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{X}}, \varepsilon_{ij}(n_{ij})\right). \tag{23}$$

Proof. See Appendix A.3.

The first part of the proposition establishes that the elasticity of aggregate outcomes, $\{\{X_{ij}\}_j, P_i, N_i, E_i, w_i\}$, to bilateral trade costs is a function of the initial aggregate trade matrix, $\bar{\boldsymbol{X}}$, the elasticity of substitution σ, and the bilateral trade elasticity matrix, $\bar{\boldsymbol{\theta}}(\bar{\boldsymbol{n}})$,. Conditional on the trade elasticity matrix, firm heterogeneity does not affect counterfactual responses. However, firm heterogeneity determines how the trade elasticity varies with the exporter firm share. Such a dependence arises because n_{ij} pins down the magnitude of the response of the extensive and intensive margins of firm exports (see equations (13) and (16)), which jointly determine the response of bilateral trade flows (see the definition of $\theta_{ij}(n_{ij})$

[8] The elasticity of substitution is necessary when the entry cost is set in terms of the origin country wage. In this case, the origin fixed-effects in the gravity equations contain the origin wage with an elasticity determined by σ.

in (18)). Thus, to compute the bilateral trade elasticity matrix, one must know the trade elasticity function, $\theta_{ij}(n)$, and the exporter firm shares in the initial equilibrium, n_{ij}.

In contrast to Proposition 1, the first part of Proposition 2 indicates that, at least for local responses, separate knowledge of the extensive and intensive margin elasticities is not required. The second part of Proposition 2 explains exactly where this knowledge is necessary. In particular, we need to know the elasticity of $\bar{\epsilon}_{ij}(n)$ to determine the elasticity of the exporter firm share, n_{ij}, to bilateral trade costs, $\bar{\tau}_{od}$. This determines the change in the trade elasticity along the path to the new counterfactual equilibrium. Notice that, given the definition of $\theta_{ij}(n_{ij})$ in (18), knowing $\varepsilon_{ij}(n)$ and $\theta_{ij}(n)$ is equivalent to knowing $\varepsilon_{ij}(n)$ and $\varrho_{ij}(n)$. This is the reason why counterfactual changes in Proposition 1 can be written in terms of $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$.

Our second remark summarizes this discussion.

Remark 2. *Firm heterogeneity only matters for aggregate counterfactual outcomes through* σ *and the shape of the elasticity functions* $(\bar{\boldsymbol{\epsilon}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{\rho}}(\bar{\boldsymbol{n}}))$. *For small trade shocks,* $(\bar{\boldsymbol{\epsilon}}(\bar{\boldsymbol{n}}), \bar{\boldsymbol{\rho}}(\bar{\boldsymbol{n}}))$ *matter only though their combine effect on the trade elasticity* $\bar{\boldsymbol{\theta}}(\bar{\boldsymbol{n}})$.

It is also important to discuss this result in the context of previous results in the literature. In particular, the result is a generalization of Proposition 2 in Arkolakis et al. (2012) that characterizes counterfactual predictions in gravity models with a constant trade elasticity – i.e. $\theta_{ij}(n_{ij}) = \theta > 0$ for all n_{ij}, i and j. We return to this point in Section 3.3. Furthermore, this result provides a characterization of nonparametric counterfactual for large changes. Arkolakis et al. (2019a) show that knowledge of the trade elasticity is sufficient, locally, for nonparametric counterfactuals in the case of two symmetric countries. We show that their results hold locally, for many asymmetric countries, but it not true when wee consider large changes as the second part of Proposition 2 indicates.

3.2 Ex-post Nonparametric Sufficient Statistics for Welfare Gains

We next link welfare gains to observable variables and measurable elasticity functions. That is, we derive ex-post nonparametric sufficient statistics for welfare changes triggered by trade shocks. Our formulas hold under the assumption of trade balance (i.e., $T_i = 0$ for all i). We discuss here the main formulas and present their derivations in Appendix A.4.

We first express changes in the real wage in terms of the change in the share of domestic active firms, $\hat{n}_{ii}$, and the extensive margin elasticity function in the domestic market, $\bar{\epsilon}_{ii}(n)$. From (12),

$$\ln\left(\frac{\hat{w}_i}{\hat{P}_i}\right) = \frac{1}{\sigma - 1}\ln\left(\frac{\bar{\epsilon}_{ii}(n_{ii}\hat{n}_{ii})}{\bar{\epsilon}_{ii}(n_{ii})}\right). \tag{24}$$

Since $\bar{\epsilon}_{ii}(n)$ is decreasing and $\sigma > 1$, the real wage increases if, and only if, n_{ii} falls (i.e., $\hat{n}_{ii} < 1$). This expression illustrates the main new source of gains from trade in Melitz (2003): the consumption-equivalent gain of reallocating resources from domestic firms with a lower entry potential to firms with a higher entry potential. Given the change in the share of domestic active firms $\hat{n}_{ii}$, welfare gains are higher if $\bar{\epsilon}_{ii}(n)$ is more elastic. Intuitively, as $\bar{\epsilon}_{ii}(n)$ becomes steeper, the difference in entry potential between incumbent and marginal firms becomes stronger, implying larger reallocation gains.[9]

Our second formula provides a decomposition of the gains from trade into three components. From (15),

$$\ln\left(\frac{\hat{w}_i}{\hat{P}_i}\right) = -\frac{1}{\sigma-1}\ln\left(\hat{x}_{ii}\right) + \frac{1}{\sigma-1}\ln\left(\hat{N}_i\hat{n}_{ii}\right) + \frac{1}{\sigma-1}\ln\left(\frac{\bar{\rho}_{ii}(n_{ii}\hat{n}_{ii})}{\bar{\rho}_{ii}(n_{ii})}\right). \tag{25}$$

The first component captures the substitution between domestic and foreign varieties implied by CES preferences. It depends on the change in the domestic spending share adjusted by the elasticity of substitution across varieties. The second component is the welfare impact of the change in the number of firms selling in the domestic market, adjusted by the variety substitution elasticity $1/(\sigma-1)$. This arises from the impact of trade shocks on competition in the domestic market. The last component arises from firm heterogeneity in terms of domestic revenue potential. It measures the difference in the revenue potential of marginal entrants compared to incumbents in the domestic market.[10]

Finally, we obtain our third formula by combining (24) and (25) to solve for the change in n_{ii} as a function of the change in $\ln\left(x_{ii}/N_i\right)$. Locally,

$$d\ln\left(\frac{w_i}{P_i}\right) = -\frac{1}{\theta_{ii}(n_{ii})}\left(d\ln x_{ii} - d\ln N_i\right). \tag{26}$$

Notice that, using the definition in (18), $\theta_{ii}(n_{ii})$ can alternatively be written in terms of σ, $\bar{\rho}_{ii}(n)$ and $\bar{\epsilon}_{ii}(n)$.

[9]Expression (24) is related to the characterization of the gains from trade in terms of the domestic productivity cutoff in Melitz (2003). Notice however that such characterization lacks empirical analogs due to the lack of measures of the productivity cutoff. Instead, our formula expresses the gains from trade in terms of the change in the share of active domestic firms $\hat{n}_{ii}$ and the extensive margin elasticity function $\bar{\epsilon}_{ii}(n)$. The next section shows that it is possible to obtain measures of both of these elements.

[10]The ratio in the third component can be alternatively written as $\frac{\bar{\rho}_{ii}(n_{ii}\hat{n}_{ii})}{\bar{\rho}_{ii}(n_{ii})} = \frac{\hat{\bar{x}}_{ii}}{\hat{\bar{x}}_{ii}(\Omega^c_{ii})}$ where Ω^c_{ii} is the set of firms of i that operate in i in any two equilibria and $\hat{\bar{x}}_{ii}(\Omega^c_{ii})$ is the change in the average domestic sales of this set of incumbent domestic firms. This alternative way of expressing the welfare impact of changing firm composition resembles the variety correction term for CES price indices introduced by Feenstra (1994). This version of expression (25) is also related to the decomposition in Hsieh et al. (2016). Our derivation provides an alternative way of measuring the importance of firm heterogeneity using the intensive margin elasticity function $\bar{\rho}_{ij}(n)$.

This expression shows that, for any given $d\ln(x_{ii}/N_i)$, the real wage change is stronger whenever the domestic trade elasticity $\theta_{ii}(n_{ii})$ is lower. Intuitively, the lower trade elasticity implies that it is harder for the economy to substitute consumption from foreign varieties to domestic varieties (through both the extensive and the intensive margins). This amplifies the cost of reducing the spending share on foreign varieties.

This expression is closely related to the welfare formulas derived in Arkolakis et al. (2009) (footnote 17) and Melitz and Redding (2013) (equation 33). The critical departure in our case is that the trade elasticity is a function of the observable share of active firms n_{ii}. This implies that, for large shocks, the computation of welfare gains must account for the correlation between changes in the domestic trade elasticity and the domestic spending share. Such a correlation arises from endogenous changes in the share of domestic active firms and its implied effect on the domestic trade elasticity.

We summarize the discussion above in the following remark.

Remark 3. *The elasticity of substitution, σ, and the domestic elasticity functions, $\bar{\epsilon}_{ii}(n)$ and $\bar{\rho}_{ii}(n)$, can be used to compute nonparametric sufficient statistics for the welfare gains implied by trade shocks.*

Gains from Trade. It is possible to use expressions (24)–(25) to compute the gains from trade. This requires solving for the changes in n_{ii} and N_i when country i moves from the initial equilibrium to the autarky equilibrium. In Appendix A.5, we show that the changes in n_{ii} and N_i are the solution of a nonlinear system of two equations and two unknowns. This system is a special case of the general system used to compute the nonparametric counterfactual changes in Proposition 1. It depends on three ingredients: (i) data on exporter firm shares and export flows of country i in the initial equilibrium $\{n_{ij}, X_{ij}\}_j$, (ii) the elasticity of substitution σ, and (iii) the two elasticity functions of firm exports for country i, $\{\bar{\epsilon}_{ij}(n), \bar{\rho}_{ij}(n)\}_j$. Accordingly, this characterization implies that firm heterogeneity only affects the welfare gains from trade of i through the shape of $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$.

3.3 Constant Elasticity Benchmark

The importance of firm heterogeneity for counterfactual analysis depends on how the two elasticity functions of firm exports vary with the exporter firm shares. To illustrate this, we contrast the results above for the general model to those obtained for a benchmark special case where the extensive and intensive margin elasticities are constant. Specifically, we impose the following functional form assumptions on $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$.

Assumption 2. ***Constant Elasticity Model.*** *For $\varrho_{ij} > -1$ and $\varepsilon_{ij} < 0$,*

$$\bar{\rho}_{ij}(n) = n^{\varrho_{ij}} \quad \textit{and} \quad \bar{\epsilon}_{ij}(n) = n^{\varepsilon_{ij}}. \tag{27}$$

Assumption 2 imposes that the extensive and intensive margin functions are isoelastic between each bilateral trading pair. This implies that the elasticities of all trade margins to changes in trade costs do not vary with the exporter firm share. With the functional forms in (27), expressions (12) and (15) imply log-linear gravity equations for the intensive and extensive margins of firm exports,

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

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

The Melitz-Pareto model in Chaney (2008) yields a special case of Assumption 2. It implies that the two functions are identical and common across countries ($\varrho_{ij} = \varepsilon_{ij} = \tilde{\theta}$).[11] In this case, the intensive margin does not respond to changes in variable trade costs because $\ln \bar{x}_{ij}$ does not depend on $\ln \bar{r}_{ij}$. The additional degrees of freedom in (27) allow the model to independently capture the impact of trade cost shocks on the extensive and intensive margins of firm-level exports.[12]

The next proposition shows that, even when both trade margins are active, the response of aggregate outcomes to trade shocks only depends on the constant trade elasticity and the elasticity of substitution.

Corollary 1. *Suppose Assumption 2 holds.*

1. Let $Y_i \equiv \{\{X_{ij}\}_j, P_i, N_i, E_i, w_i\}$. The elasticity of any element of Y_i to $\bar{\tau}_{od}$ is a function of $\left(\sigma, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{X}}\right)$:

$$\frac{d \ln Y_i}{d \ln \bar{\tau}_{od}} = \Psi_{i,od}\left(\sigma, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{X}}\right) \quad \textit{such that} \quad \theta_{ij} = (\sigma - 1)\left(1 - \frac{1+\varrho_{ij}}{\varepsilon_{ij}}\right). \tag{28}$$

[11] Assume that firms only differ with respect to their productivity such that $r_{ij}(\omega) = \epsilon_{ij}(\omega) = (a_i(\omega))^{\sigma-1}$, and $a_i(\omega) \sim 1 - a^{-\theta}$ with $\theta > \sigma - 1$. So, $e_{ij}(\omega) \sim H^e(e) = 1 - e^{-\frac{\theta}{\sigma-1}}$. This immediately implies that the extensive margin function is $\bar{\epsilon}_{ij}(n) = (H^e)^{-1}(1-n) = n^{-\frac{\sigma-1}{\theta}}$. Also, conditional on $e = \bar{\epsilon}_{ij}(n)$, $H^r_{ij}\left(r | e = \epsilon_{ij}(n)\right)$ is degenerate at $r = \bar{\epsilon}_{ij}(n)$, which implies that $\rho_{ij}(n) = n^{-\frac{\sigma-1}{\theta}}$ and $\bar{\rho}_{ij}(n) = \left(1 - \frac{\sigma-1}{\theta}\right)^{-1} n^{-\frac{\sigma-1}{\theta}}$.

[12] Evidence in Fernandes et al. (2017) shows that trade costs affect both the extensive and the intensive margins of firm exports.

2. Assume that trade is balanced (i.e., $T_i = 0$) and $o \neq d$. The real wage change is

$$\Delta \ln \left(\frac{w_i}{P_i} \right) = -\frac{1}{\theta_{ii}} \Delta \left(\ln x_{ii} / N_i \right). \tag{29}$$

3. If we assume further that $\varrho_{ij} = \varrho_i$ and $\varepsilon_{ij} = \varepsilon_i$, then $\theta_{ij} = \theta_i$ and $\Delta \ln N_i = 0$.

Proof. Expression (28) follows directly from Proposition 2 when $\theta_{ij}(n) = \theta_{ij}$ for all i and j. Expression (29) follows directly from directly from (26) when $\theta_{ii}(n) = \theta_{ii}$. Appendix A.6 establishes the last part.

The first part of Corollary 1 shows that, for any trade shock, computing changes in aggregate outcomes only requires the matrix of constant bilateral trade elasticities, $\bar{\boldsymbol{\theta}} = [\theta_{ij}]_{i,j}$. For larger shocks, we can obtain changes in all aggregate outcome by integrating the local responses in (28) without tracking the initial exporter share n_{ij}. Thus, in this case, it is not necessary to separately know the elasticities of the intensive and extensive margins, ε_{ij} and ϱ_{ij}, nor exporter firm shares in the initial equilibrium, n_{ij}. Thus, firm heterogeneity does not matter for aggregate outcomes in this benchmark constant elasticity economy.

The second part of the proposition establishes the gains from trade in this constant elasticity case. It is just the integral of (26), with a constant domestic trade elasticity. The last part indicates that, when the trade elasticity is further restricted to be identical across all destinations, we recover a generalized version of the sufficient statistic for the gains from trade in gravity trade models of Arkolakis et al. (2012) where the trade elasticity is origin-specific. This follows from the fact that, in this case, the number of entrants, N_i, does not change with bilateral trade costs. The following remark summarizes this.

Remark 4. The aggregate implications of firm heterogeneity depend on how the elasticities functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ vary with the exporter firm share. If these functions are constant elasticity then the bilateral trade elasticities θ_{ij} are constant and sufficient to compute counterfactual predictions.

3.4 Extensions

We derive five extensions of our baseline framework in Appendix B. For each extension, we derive semiparametric gravity equations for different margins of firm exports. Furthermore, for the first three extensions, we show how to conduct counterfactual analysis without parametric restriction on the distribution of firm fundamentals using the functions in the semiparametric gravity equations.

Multiple sectors, multiple factors, input-output links. We extend our baseline model to include multiple factors of production and input-output links between multiple sectors. Specifically, we extend the multi-sector multi-factor gravity model of Costinot and Rodriguez-Clare (2013) in which, as in our baseline, firms in each sector are heterogeneous with respect to productivity, preferences, and variable and fixed trade costs. We restrict all firms in a sector to have the same nested constant elasticity of substitution (CES) production technology that uses multiple factors and multiple sectoral composite goods. In this setting, we derive sector-specific analogs of (12) and (15) that can be used to perform nonparametric counterfactual analysis with respect to trade cost shocks.

Allowing for zero bilateral trade flows. We extend our baseline framework to allow for zero trade flows between two countries. As in Helpman et al. (2008), we allow the support of the entry potential distribution to be bounded: $H^e_{ij}(e)$ has full support over $[0, \bar{e}_{ij}]$. The bounded support does not affect the intensive margin gravity equation (15), but it introduces a censoring structure in the extensive margin equation (12). Under the assumption that zero trade flows remain equal to zero, we use these extended gravity equations to compute nonparametric counterfactual changes in aggregate outcomes following trade cost shocks.

Allowing for import tariffs. Third, we follow Costinot and Rodriguez-Clare (2013) and introduce bilateral import tariffs in the model. In this setting, market clearing and spending must account for the fact that tariff revenue remains in the destination country. We show that the semiparametric gravity equations above still hold, but now bilateral trade costs also include ad-valorem import tariffs. We then characterize the system of equations that determines the model's counterfactual predictions without parametric distributional assumptions. It depends on the same elements as before, with one addition, the tariff levels in the initial equilibrium.

Multi-product firms. We follow Bernard et al. (2011) and Arkolakis et al. (2019b) to formulate a model of multi-product firms without parametric assumptions on firm fundamentals. We allow each potential firm to produce an exogenous number of varieties, and receive product-specific demand, trade cost, fixed cost, and productivity draws. We derive three types of gravity equations: (i) the extensive margin of firm entry, (ii) the extensive margin of average number of products per exporter firm, and (iii) the intensive margin of exports per product (averaged across all exporters).

Non-CES Preferences. We adapt our framework to general Marshallian demand functions that can be written as a function of the destination's price aggregator and income level. Our demand system subsumes the settings in Arkolakis et al. (2019a) and Matsuyama and Ushchev (2017). In our case, we abstract from fixed entry costs and incorporate endogenous firm entry through a choke price in demand. We show how to extend our inversion argument to derive an extensive margin gravity equation analogous to the one in (12). Because revenue and entry potentials are identical, the same function in the extensive margin gravity equation determines the intensive margin of average firm exports.

4 Estimation Strategy: Semiparametric Gravity Equations of Firm Exports

Firm heterogeneity affects the impact of trade shocks on aggregate outcomes and welfare through the shape of $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$. In the model, these two functions control the semiparametric gravity equations for the extensive and intensive margins of firm exports. This section outlines a strategy to estimate $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$ using these semiparametric gravity equations.

4.1 Semiparametric Gravity Equations of Firm Exports

The gravity equations in (12) and (15) imply the following semiparametric specifications:

$$\ln \bar{\epsilon}_{ij}(n_{ij}) = \ln\left(\bar{f}_{ij}\bar{\tau}_{ij}^{\sigma-1}\right) + \tilde{\delta}_i^{\epsilon} + \tilde{\zeta}_j^{\epsilon}, \tag{30}$$

$$\ln \bar{x}_{ij} - \ln \bar{\rho}_{ij}(n_{ij}) = \ln\left(\bar{\tau}_{ij}^{\sigma-1}\right) + \tilde{\delta}_i^{\rho} + \tilde{\zeta}_j^{\rho}. \tag{31}$$

where $\tilde{\delta}_i^{\epsilon} \equiv \ln(\sigma^{\sigma}(\sigma-1)^{1-\sigma}\bar{a}_i^{1-\sigma}w_i^{\sigma})$, $\tilde{\zeta}_j^{\epsilon} \equiv -\ln(E_j P_j^{\sigma-1})$, $\tilde{\delta}_i^{\rho} \equiv \ln(\sigma w_i) - \tilde{\delta}_i^{\epsilon}$, and $\tilde{\zeta}_j^{\rho} \equiv -\tilde{\zeta}_j^{\epsilon}$. Without loss of generality, we normalize $\bar{b}_{ij} \equiv 1$ since bilateral shifters of demand and trade costs are isomorphic in the model – i.e., the equilibrium only depend on $\bar{\tau}_{ij}^{1-\sigma}\bar{b}_{ij}$.

These two equations form the basis of our empirical strategy. They link average firm revenue and the two elasticity functions of the exporter firm share to bilateral shifters of variable and fixed costs of exporting, as well as to exporter and importer fixed-effects. We can then estimate $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$ using these equations along with bilateral data on average firm exports, exporter firm shares, and trade cost shifters.

Remark 5. *$\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$ can be estimated with the semiparametric specifications (30)–(31).*

In the rest of this section, we first describe sufficient assumptions on the data generating process that allow us to estimate $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$ using the semiparametric equations in (30) and (31). We then outline an estimator of $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$ based on cross-country variation in bilateral trade cost shifters.

4.2 Data Generating Process

Our goal is to estimate the functions $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$. Throughout our analysis, we use estimates in the literature to calibrate the elasticity of substitution, σ. In particular, we set $\sigma = 3.9$ to match the median estimate in Hottman et al. (2016).

We assume that the equilibrium of the model of Section 2 determines all outcomes in the world economy. We start by describing the observed and unobserved variables in the economy. For a set of origin-destination pairs (i, j), we observe the share of firms of i selling in j, n_{ij}, and their average sales, $\bar{x}_{ij}$. We assume that we observe a component of variable trade costs (denoted by τ_{ij}), and an exogenous shifter of bilateral trade costs (denoted by z_{ij}). Previewing our empirical application, we use data on bilateral freight costs to measure τ_{ij}, and data on bilateral distance to measure z_{ij}.

Assumption 3. *Assume that we observe a component of variable trade costs, τ_{ij}, such that*

$$\ln \bar{\tau}_{ij} = \ln \tau_{ij} + \eta^u_{ij}. \tag{32}$$

Assume also that there exists an observed bilateral trade shifter, z_{ij}, such that

$$\begin{aligned} \ln \bar{\tau}_{ij} &= z_{ij}\kappa^\tau + \delta^\tau_i + \zeta^\tau_j + \eta^\tau_{ij}, \\ \ln \bar{f}_{ij} &= z_{ij}\kappa^f + \delta^f_i + \zeta^f_j + \eta^f_{ij}. \end{aligned} \tag{33}$$

These equations are equivalent to the first-stage equations in the estimation of the semiparametric gravity equations of firm exports. They link variable and fixed trade costs to an observed shifter, accounting for the fact that we may not observe all components of trade costs.[13]

To estimate $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$, we impose assumptions on the data generating process of trade costs in the economy.

Assumption 4. *Assume that $E[\eta^\tau_{ij}|z_{ij}, D_{ij}] = E[\eta^f_{ij}|z_{ij}, D_{ij}] = E[\eta^u_{ij}|z_{ij}, D_{ij}] = 0$, where D_{ij} is a vector of origin and destination fixed-effects.*

[13] This specification allows z_{ij} to affect the fixed cost of entering foreign markets and, therefore, it is weaker than the requirement in Helpman et al. (2008) that the instrument cannot affect entry costs.

This orthogonality assumption is the basis of the estimation of constant elasticity gravity equations of international trade flows – for a review, see Head and Mayer (2013). Conditional on origin and destination fixed-effects, the observed shifter must be mean independent from unobserved shifters of trade costs.

Finally, we impose the following restrictions on $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$.

Assumption 5. *Assume that origin-destination pairs are divided into groups* $(g = 1, ..., G)$ *such that, for all* $(i, j) \in g$,

$$\begin{bmatrix} \ln \bar{\rho}_{ij}(n) \\ \ln \bar{\epsilon}_{ij}(n) \end{bmatrix} = \begin{bmatrix} \ln \bar{\rho}_{g}(n) \\ \ln \bar{\epsilon}_{g}(n) \end{bmatrix} = \sum_{k=1}^{K} \begin{bmatrix} \gamma^{\rho}_{g,k} f_k(\ln n) \\ \gamma^{\epsilon}_{g,k} f_k(\ln n) \end{bmatrix} \tag{34}$$

where $f_k(\ln n)$ *denotes restricted cubic splines over knots* $k = 1, ..., K$.

This assumption imposes two types of restrictions on $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$. First, these functions are identical among origin-destination pairs in the same group g. This allows us to estimate $\bar{\rho}_g(n)$ and $\bar{\epsilon}_g(n)$ using variation in the observed shifters of trade costs across origin-destination pairs at a point in time.[14] In our empirical application, we specify that all countries belong to a single group, so that $\bar{\epsilon}_{ij}(n) = \bar{\epsilon}(n)$ and $\bar{\rho}_{ij}(n) = \bar{\rho}(n)$ for all i and j. In Appendix C.5, we provide estimates with multiple country groups defined in terms of characteristics of origin and destination countries.

Second, Assumption 5 specifies a flexible function basis for $\bar{\rho}_g(n)$ and $\bar{\epsilon}_g(n)$. We approximate the shape of these functions with a series of restricted cubic polynomials. Specifically, we specify K knots that form a series of intervals, $\mathcal{U}_k \equiv [u_k, u_{k+1}]$. In each interval, a restricted cubic spline governs the behavior of the elasticity function. See Appendix C.2 for details.

4.3 Estimating Moment Conditions

We use the above assumptions to construct moment conditions for the estimation of $\bar{\rho}_g(n)$ and $\bar{\epsilon}_g(n)$.

Pass-through from observed shifter to variable trade cost. We first specify an equation for the estimation of the pass-through from the observed cost shifter z_{ij} to the observed component of variable trade costs τ_{ij}. Assumption 3 implies that

$$v^{\tau}_{ij} = \ln \tau_{ij} - z_{ij}\kappa^{\tau} - \delta^{\tau}_{i} - \zeta^{\tau}_{j}. \tag{35}$$

[14] Our notation allows groups to be defined as destination-origin country pairs over different years. In this case, one can easily extend our strategy to exploit variation over time in the observed trade cost shifter to obtain bilateral-specific estimates of the elasticity functions.

Here, $v_{ij} = \eta_{ij}^{\tau} - \eta_{ij}^{u}$, implying that $E[v_{ij}^{\tau}|z_{ij}, D_{ij}] = 0$ by Assumption 4. We exploit this condition to estimate κ^{τ} using the linear equation in (35).

Semiparametric gravity equations of firm exports. To estimate $\bar{\rho}_g(n)$ and $\bar{\epsilon}_g(n)$, we show in Appendix A.7 that, under Assumptions 3 and 5, equations (30)–(31) are equivalent to

$$\begin{bmatrix} v_{ij}^{\epsilon} \\ v_{ij}^{\rho} \end{bmatrix} = \begin{bmatrix} z_{ij} \\ \ln \bar{x}_{ij} + \tilde{\sigma}\kappa^{\tau} z_{ij} \end{bmatrix} - \sum_{k=1}^{K} \begin{bmatrix} \kappa^{\epsilon}\gamma_{g,k}^{\epsilon} f_k(\ln n) \\ \gamma_{g,k}^{\rho} f_k(\ln n) \end{bmatrix} - \begin{bmatrix} \delta_i^{\epsilon} + \zeta_j^{\epsilon} \\ \delta_i^{\rho} + \zeta_j^{\rho} \end{bmatrix}, \tag{36}$$

where $\tilde{\sigma} \equiv \sigma - 1$ and $\kappa^{\epsilon} \equiv 1/\left(\tilde{\sigma}\kappa^{\tau} + \kappa^{f}\right)$.

In terms of the structural unobserved shifters introduced above, $v_{ij}^{\epsilon} \equiv \kappa^{\epsilon}(v_{ij}^{\rho} - \eta_{ij}^{f})$ and $v_{ij}^{\rho} \equiv -\tilde{\sigma}\eta_{ij}^{\tau}$. Thus, Assumption 4 implies that $E[v_{ij}^{\epsilon}|z_{ij}, D_{ij}] = E[v_{ij}^{\rho}|z_{ij}, D_{ij}] = 0$. Combined with (36), these moment conditions can be used to estimate the parameters $\gamma_{g,k}^{\rho}$ and $\gamma_{g,k}^{\epsilon}$ for each knot k.[15]

Pass-through from observed shifter to fixed entry cost. To estimate the scale parameter κ^{ϵ}, we exploit the restriction imposed by the specification of entry costs in terms of labor in the origin country. Under this assumption,

$$v_j^{f} = \kappa^{\epsilon}\zeta_j^{\rho} - \zeta_j^{\epsilon}. \tag{37}$$

Here, $v_j^{f} \equiv \kappa^{\epsilon}\zeta_j^{f}$ is the destination fixed-effect in the first-stage specification for the entry cost in (33). Since there is a constant in (33), $E[v_j^{f}] = 0$. We use this moment condition to estimate κ^{ϵ}.

Estimator. Expressions (35)–(37) can be used to compute $(v_{ij}^{\tau}, v_{ij}^{\epsilon}, v_{ij}^{\rho}, v_j^{f})$ conditional on our main parameters of interest, $\boldsymbol{\Theta} \equiv \left(\kappa^{\epsilon}, \kappa^{\tau}, \left\{\gamma_{g,k}^{\rho}, \gamma_{g,k}^{\epsilon}\right\}_{g,k=1}^{G,K}\right)$, as well as the set of origin fixed-effects, $\boldsymbol{\delta} \equiv \{\delta_i^{\tau}, \delta_i^{\epsilon}, \delta_i^{\rho}\}_{i=1}^{N}$ and destination fixed-effects, $\boldsymbol{\zeta} \equiv \left\{\zeta_i^{\tau}, \zeta_j^{\epsilon}, \zeta_j^{\rho}\right\}_{j=1}^{N}$. We use the recovered structural residuals to construct the following GMM estimator for $(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\zeta})$:

$$\min_{(\Theta,\delta,\zeta)} h\left(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\zeta}\right)' \hat{\Omega} h\left(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\zeta}\right), \quad \text{where} \quad h\left(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\zeta}\right) \equiv \begin{bmatrix} \sum_{ij}\left(v_{ij}^{\tau} z_{ij},\ v_{ij}^{\tau} D_{ij}\right)' \\ \sum_{ij}\left(v_{ij}^{\epsilon} F(z_{ij}), v_{ij}^{\epsilon} D_{ij}\right)' \\ \sum_{ij}\left(v_{ij}^{\rho} F(z_{ij}), v_{ij}^{\rho} D_{ij}\right)' \\ \sum_{j} v_j^{f} \end{bmatrix}, \tag{38}$$

[15]Conditional on observing κ^{τ}, the assumption that $E[\eta_{ij}^{u}|z_{ij}, D_{ij}] = 0$ is not necessary for the estimation of $\bar{\epsilon}_g(n)$ and $\bar{\rho}_g(n)$ using (36). The assumption of $E[\eta_{ij}^{u}|z_{ij}, D_{ij}] = 0$ only affects the estimation of κ^{τ} based on (35). Accordingly, as in Adao et al. (2017), it is possible to estimate $\bar{\epsilon}_g(n)$ and $\bar{\rho}_g(n)$ with the alternative assumption of perfect pass-through from z_{ij} to $\bar{\tau}_{ij}$ (i.e., $\kappa^{\tau} \equiv 1$).

and $\hat{\Omega}$ is the two-step optimal matrix of moment weights.[16]

In our estimator, $F(z_{ij})$ is a vector function of the bilateral cost shifter. We specify the instrument vector to match the functional form in Assumption 5:

$$F(z_{ij}) \equiv \left\{ \mathbb{I}_{(ij \in g)} \mathbb{I}_{(n \in \mathcal{U}_k)} (z_{ij})^d \right\}_{g=1,k=1,d=1}^{G,K,3} .$$

5 Empirical Estimation

We use the strategy above to estimate $\bar{\rho}_{ij}(n)$ and $\bar{\epsilon}_{ij}(n)$ using the semiparametric gravity equations for the extensive and intensive margins of firm exports. Our results show how the two elasticity functions of firm exports vary with the exporter firm share. In the next section, we use our estimates to evaluate how much firm heterogeneity matters for the measurement of the gains from trade.

5.1 Data

Our baseline data source for bilateral trade flows is the 2016 release of the World Input-Output Database (WIOD). It contains domestic sales, X_{ii}, as well as bilateral trade flows, X_{ij}, for 43 countries between 2000 and 2014. The first columns in Table 4 in Appendix C.1 presents the list of countries with trade flows in the WIOD. Our sample of countries accounts for 90% of world trade and entails positive bilateral flows for almost all exporter-importer pairs.[17]

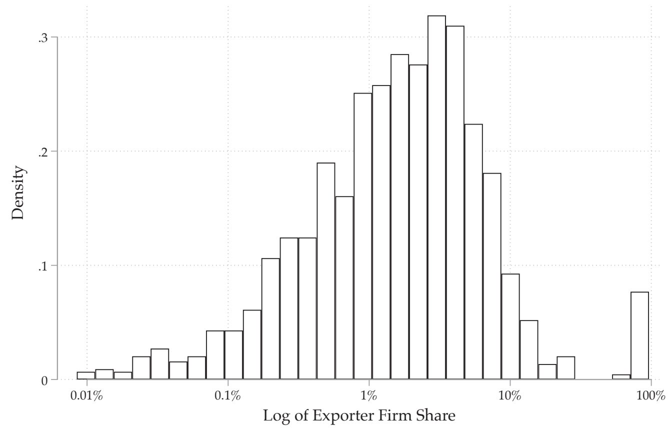
The estimator in equation (38) requires four bilateral variables: (i) the exporter firm share, n_{ij}; (ii) the average firm revenue, $\bar{x}_{ij}$; (iii) the trade cost shifter, z_{ij}; and (iv) the observed component of trade costs, τ_{ij}.

We use various sources to construct n_{ij} and $\bar{x}_{ij}$ for a subset of 35 origin countries in the WIOD – for the full list of countries, see columns (2) and (3) of Table 4 in Appendix C.1. We construct the data in two steps. We first use the OECD Trade by Enterprise Characteristics (TEC) database to obtain the number of manufacturing firms from i selling in j, N_{ij} for $i \neq j$. For origin-destination pairs not in the OECD TEC database, we obtain N_{ij} from the World Bank Exporter Dynamics Database (EDD). These datasets also contain the total exports of the same set of firms from i exporting to j. We use this information to compute the average revenue of firms from i selling to j, $\bar{x}_{ij}$.

[16] As the fixed effects in equations (35)–(36) are linear, we can follow a similar procedure to Berry et al. (1995) and partial-out the fixed effects without having to iterate over these terms in the GMM estimator in (38). This reduces the dimensionality of the minimization problem and, therefore, the computational burden of estimation.

[17] This attenuates concerns related to the estimation of gravity equations with zero trade flows, as in Helpman et al. (2008) and Silva and Tenreyro (2006).

Figure 2: Empirical distribution of $\ln n_{ij}$, 2012



Note. Empirical distribution of $\ln(n_{ij})$ in the cross-section of origin-destination pairs in 2012. For the list of countries, see Table 4 in Appendix C.1.

The second step is the construction of the number of entrants N_i, which is not readily available in national statistics. Together with the number of exporters N_{ij}, we use N_i to construct the exporter firm shares, $n_{ij} = N_{ij}/N_i$.[18] We compute the number of entrants as $N_i = N_{ii}/n_{ii}$ where, in country i, N_{ii} is the number of active manufacturing firms and n_{ii} is the survival probability of new manufacturing firms. Our approach assumes 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. To maximize country coverage, we obtain N_{ii} from several datasets: the OECD Demographic Business Statistics (SDBS), the OECD Structural Statistics for Industry and Services (SSIS), and the World Bank Enterprise Surveys. In addition, we obtain n_{ii} from the one-year survival rate of manufacturing firms in the OECD SDBS.[19]

Our measure of the bilateral trade cost shifter, z_{ij}, is the bilateral distance (population-weighted) in the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII). This dataset includes not only distance between countries, but also distances within a country z_{ii} due to the nature of population weighting. We use this information to include observations associated with domestic trade in our baseline sample.

[18]Prior research circumvents this data requirement by assuming that $N_i = N_{ii}$ and $n_{ii} = 1$ – e.g., see Fernandes et al. (2017). However, this limits the potential sources of gains from trade in our model by shutting down welfare gains implied by changes in domestic firm composition and firm performance – see equation (25).

[19]This data is only available for 80% of the origin countries 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 that our results are robust to excluding from the sample countries without survival rate data.

Finally, we use the bilateral freight cost to measure the observed component of variable trade costs τ_{ij}. This is only necessary for the estimation of κ^τ using the linear specification in (35). We consider a subset of countries for which we observe CIF/FOB import margins in the OECD freight cost database. Columns (4) and (5) of Table 4 in Appendix C.1 report the list of countries with available data on bilateral freight costs.

The availability of data on $\bar{x}_{ij}$, n_{ij}, and z_{ij} determines our sample for the estimation of the last three moment conditions in (38). Table 4 in Appendix C.1 reports the list of countries in our baseline sample. Figure 2 summarizes the distribution of $\ln(n_{ij})$ for all bilateral pairs in 2012.[20] The empirical distribution of n_{ij} is central for our analysis: because n_{ij} is the only input of the elasticity functions $\bar{\rho}_g(n)$ and $\bar{\epsilon}_g(n)$, we are only able to precisely estimate these functions in the part of the support in which we observe values of n_{ij}.

5.2 Pass-Through of Distance to Freight Costs

We start by estimating the pass-through parameter κ^τ from the linear specification in (35). We pool data from 2008-2014 to estimate the following regression:

$$\log \tau_{ij,t} = \kappa^\tau \log z_{ij} + \delta^\tau_{i,t} - \zeta^\tau_{j,t} + \epsilon_{ij,t},$$

where $\delta^\tau_{i,t}$ and $\zeta^\tau_{j,t}$ are respectively origin-year and destination-year fixed-effects.

Table 1 reports the pass-though estimates along with standard errors clustered at the destination-origin level. We estimate an elasticity of trade costs to distance of roughly 0.35. We obtain similar pass-through estimates in the presence of different sets of fixed-effects. This is reassuring given that the fixed-effects absorb a great deal of variation in freight costs in our sample – the R^2 increases from 0.47 in columns (1) to 0.82 in column (3).

5.3 Constant Elasticity Gravity Estimation

As a benchmark, we estimate $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ under the constant elasticity functional form in (27). In this special case, these functions are fully characterized by the parameters ε and ϱ that respectively capture the inverse elasticity of firm entry to trade costs and the elasticity of average revenue potential to firm entry. These parameters are intrinsically tied to the elasticities of firm-level entry and sales, n_{ij} and $\bar{x}_{ij}$, to the bilateral trade shifter, z_{ij}. Under

[20] We obtain a similar country coverage for ($\bar{x}_{ij}$, n_{ij}, z_{ij}) in every year between 2010 and 2014. In addition, 2012 is the year with the most observations of the freight cost τ_{ij} used in estimation. In the appendix, we show that our results are similar when we use data for 2010 and 2014.

Table 1: Estimation of κ^τ

	Dep. Var.: Log of Freight Cost		
	(1)	(2)	(3)
Log of Distance	0.351***	0.349***	0.359***
	(0.059)	(0.085)	(0.100)
R^2	0.479	0.722	0.871
Fixed-Effects:			
Year	Yes	Yes	No
Origin, Destination	No	Yes	No
Origin-Year, Destination-Year	No	No	Yes

Note. Sample of 522 origin-destination-year triples described in Table 4 of Appendix C.1. Standard errors clustered by origin-destination pair. *** p < 0.01

(27), the gravity equations in (36) imply the following log-linear specifications:

$$\begin{aligned} \ln n_{ij} &= \beta^\epsilon z_{ij} + \tilde{\delta}_i^\epsilon + \tilde{\zeta}_j^\epsilon + \eta_{ij}^\epsilon \\ \ln \bar{x}_{ij} &= \beta^\rho z_{ij} + \tilde{\delta}_i^\rho + \tilde{\zeta}_j^\rho + \eta_{ij}^\rho, \end{aligned} \tag{39}$$

where

$$\beta^\epsilon \equiv (\kappa^\epsilon \varepsilon)^{-1} \quad \text{and} \quad \beta^\rho \equiv -\tilde{\sigma}\kappa^\tau + \varrho\beta^\epsilon. \tag{40}$$

Table 2 presents the estimates of (39). Column (1) indicates that the exporter firm share falls sharply with distance: a 1% higher bilateral distance leads to a 1.2% decline in exporter firm share. Column (2) indicates that average sales also decline with distance. As pointed out by Fernandes et al. (2015), this evidence is inconsistent with the lack of average revenue responses in the Melitz-Pareto model (Chaney, 2008). Finally, column (3) reports an elasticity of bilateral trade flows to distance of -2, which is slightly lower than the typical estimates in the literature reviewed by Head and Mayer (2013).

In Panel B of Table 2, we use the expressions in (40) to recover $\kappa^\epsilon\varepsilon$ and ϱ. We use our baseline calibration of $\tilde{\sigma}\kappa^\tau = 1.04$ that sets $\tilde{\sigma} = \sigma - 1 = 2.9$ from Hottman et al. (2016) and $\kappa^\tau = 0.36$ from column (3) of Table 1. The negative extensive margin elasticity implies that $\kappa^\epsilon\varepsilon < 0$. Thus, in line with our model, $\varepsilon < 0$ whenever distance increases trade costs, $\kappa^\epsilon > 0$. In addition, the implied value of ϱ indicates that the average revenue potential of all exporters falls by 0.2% when the exporter firm share increases by 1%. Hence, marginal exporters have a lower revenue potential than incumbent exporters in each market.

Finally, Table 3 presents the estimates of ε and ϱ obtained with the GMM estimator in (38) under the constant elasticity assumption in (27). Relative to the discussion above, the full structural estimation yields separate estimates of κ^ϵ and ε. Specifically, our estimate of ε indicates that a 1% increase in bilateral trade costs triggers a reduction of 1.1% in the

Table 2: Constant Elasticity Gravity of Firm Exports

Dep. Var.:	$\ln n_{ij}$	$\ln \bar{x}_{ij}$	$\ln X_{ij}$
	(1)	(2)	(3)
Panel A: Constant elasticity gravity estimation			
Log of Distance	-1.187***	-0.795***	-1.982***
	(0.0487)	(0.0434)	(0.0724)
R^2	0.906	0.693	0.867
Panel B: Implied structural parameters			
	$\kappa^{\epsilon} \times \varepsilon$	ϱ	
	-0.842	-0.208	

Note. Sample of 1,479 origin-destination pairs in 2012 – see Table 4 of Appendix C.1. All specifications include origin and destination fixed-effects. Implied structural parameters computed with $\tilde{\sigma} = 2.9$ from Hottman et al. (2016) and $\kappa^{\tau} = 0.36$ from column (3) of Table 1. Standard errors clustered by origin-destination pair. *** $p < 0.01$

exporter firm share. This is consistent with a skewed distribution of firm entry potentials. Due to the average sales responses in Panel A of Table 2, we reject the hypothesis that $\varrho = \varepsilon$ which holds in the Melitz-Pareto model of Chaney (2008).

We can use our estimates of ε and ϱ to compute an estimate of the elasticity of bilateral trade to bilateral trade cost using the definition in (28). The last column of Table 3 indicates an implied trade elasticity of 5. This is within the range of estimates in the literature reviewed by Costinot and Rodriguez-Clare (2013).

Table 3: Constant Elasticity Gravity of Firm Exports with $\bar{\epsilon}_{ij}(n) = n^{\varepsilon}$ and $\bar{\rho}_{ij}(n) = n^{\varrho}$

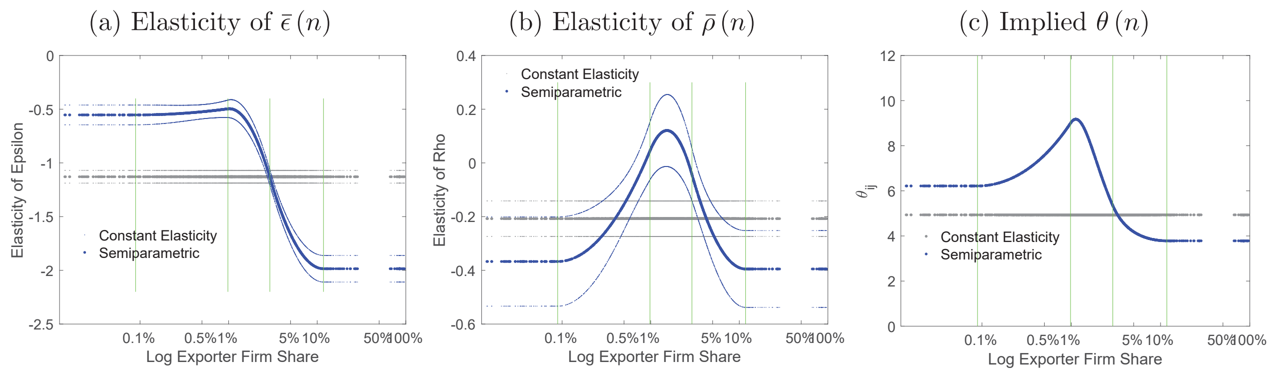
ε	ϱ	θ
-1.13	-0.21	4.94
(0.03)	(0.03)	

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 of Appendix C.1. Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

5.4 Semiparametric Gravity Estimation

We now turn to our semiparametric estimates of $\bar{\epsilon}_g(n)$ and $\bar{\rho}_g(n)$ for a single group pooling all countries. Thus, to simplify notation, we drop the subscript g. Figure 3 presents estimates of the elasticities of $\bar{\epsilon}(n)$ and $\bar{\rho}(n)$ with respect to the exporter firm share. We use green bars to denote the estimation knots. We overlay our baseline estimates with the estimates of the constant elasticity specification presented in Table 3. We report the elasticity of the extensive margin function $\bar{\epsilon}(n)$ in Panel (a), the elasticity of the intensive margin function $\bar{\rho}(n)$ in Panel (b), and the implied trade elasticity $\theta(n)$ in Panel (c) (obtained from (18)).

Figure 3: Semiparametric Gravity of Firm Exports with $\bar{\epsilon}_{ij}(n) = \bar{\epsilon}(n)$ and $\bar{\rho}_{ij}(n) = \bar{\rho}(n)$



Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 of Appendix C.1. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

Our estimates show that the elasticity functions vary with the share of firms exporting to a market. Since the extensive margin elasticity $\frac{\partial \ln n_{ij}}{\partial \ln \bar{\tau}_{ij}}$ is inversely proportional to $\varepsilon(n)$ (see equation (13)), Panel (a) shows that the extensive margin becomes less responsive as more firms serve a market. For low levels of entry, the expression in (13) implies that a 1 log-point increase in trade costs reduces the share of exporting firms by 4.8 log-points. This elasticity is lower for higher levels of firm entry. In the top knot, a 1 log-point increase in trade costs reduces the exporter firm share by 1.5 log-point. This implies that, at first, exporters are very sensitive to changes in trade frictions. However, high levels of entry potential are rare in the economy: when many firms export, small changes in trade frictions lead to smaller responses in the share of firms that decide to export.

Panel (b) indicates that selection patterns change with the share of firms exporting to a market. For low and high levels of the exporter firm share, the intensive margin elasticity is negative, being around -0.35. This indicates that marginal firms with high and low levels of entry potential have a lower revenue potential than infra-marginal firms already operating in each market. In other words, the marginal entrants have a lower revenue potential than incumbents in the market when the exporter firm share is either high or low. In contrast, selection forces are much weaker for mid-levels of firm entry potential. In the middle of the support, the elasticity of $\bar{\rho}(n)$ is not statistically different from zero. This indicates that revenue potential is similar among middle-ranked firms in terms of entry potential.

Panel (c) shows what the firm-level elasticity margins imply for the response of bilateral trade flows to changes in bilateral trade costs. Our estimates shows that the declining extensive margin leads to a lower trade elasticity when the exporter firm share is high. Thus, in line with the product-level evidence in Kehoe and Ruhl (2013b) and Kehoe et al. (2015), the trade elasticity tends to be lower when trade volumes are high. Notice that the trade elasticity varies between 3.5 and 8. Such values are consistent with the range of trade elasticity estimates in the literature – see Costinot and Rodriguez-Clare (2013).[21]

In Figure 7 of Appendix C.3, we compare our estimated trade elasticity function to that implied by parametric assumptions and their associated estimates about the distribution of firm fundamentals currently present in the literature. The log-normal assumption in Bas et al. (2017) and Head et al. (2014) implies a much steeper trade elasticity function. The trade elasticity for high levels of n_{ij} is below two, while it is above twelve for low levels of n_{ij}. The truncated Pareto assumption in Melitz and Redding (2015) yields a trade elasticity function that is uniformly low. It is always below four and falls below two for high levels of n_{ij}.

These comparisons highlight the difference between our approach based on semiparametric gravity equations and approaches based on parametrizations of cross-section variation in firm-level outcomes. While we directly estimate the elasticity functions driving the model's aggregate predictions, the parametric micro approach extrapolates from heterogeneity in firm-level outcomes to obtain these elasticity functions. Our results indicate that this extrapolation may lead to elasticity functions that are substantially different from those implied by estimates of the semiparametric gravity equations of firm exports. In the next section, we investigate the quantitative implications of such differences for the model's counterfactual predictions.

Robustness of baseline estimates. In Appendix C.4, we investigate the robustness of the baseline estimates in Figure 3. We obtain similar estimates using years that have a similar country coverage. We also show that the trade elasticity function is similar when we exclude observations associated with domestic sales. In addition, we investigate the sensitivity of our estimates to the procedure to measure N_i. The estimated elasticity functions are similar when n_{ii} is either the two-year or the three-year survival rate of manufacturing firms. Estimates are also similar when we exclude from the sample origin countries for which we impute the one-year survival rate. Finally, we re-estimate the elasticity functions under the assumption that all entrants sell in the domestic market (i.e., $n_{ii} = 1$ and $\bar{f}_{ii} = 0$). In this case, for low levels of the exporter firm share, extensive margin elasticity is higher, leading to a higher trade elasticity. Outside the bottom of the support, these elasticity function estimates become similar to our baseline estimates.

[21]It is possible that existing trade elasticity estimates are average treatment effects obtained from variation in particular parts of the support of exporter firm shares. Our approach then just captures how the trade elasticity varies across the support of values of n_{ij}.

Additional estimates for multiple country groups. Our baseline estimates impose identical elasticity functions across all exporter-destination pairs ($G = 1$). In Appendix C.5, we allow the elasticity functions to vary across country groups. Specifically, we investigate whether the trade elasticity functions vary with the country's per capita income, as in Adao et al. (2017). We find that, for developed origins, the trade elasticity varies between five and seven depending on the level of n_{ij}. However, for developing origins, the trade elasticity is nine for a low n_{ij}, but it is only four for a high n_{ij}. The trade elasticity does not vary with the destination's development level. We also find that the elasticity functions do not differ for country pairs inside and outside Free Trade Areas.

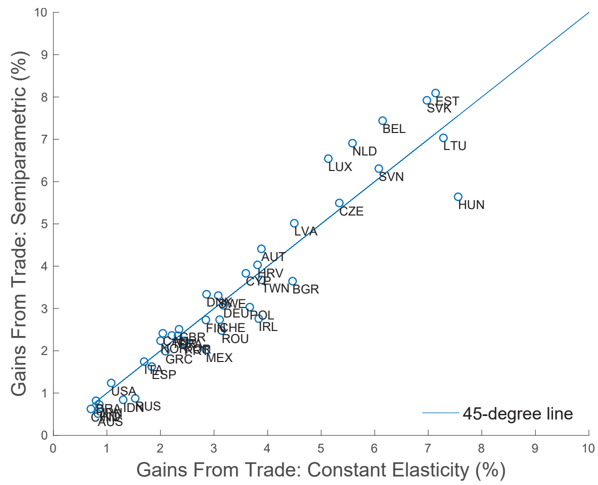
6 Quantifying the Gains From Trade

We conclude by revisiting the question: how large are the gains from trade? We first measure the importance of firm heterogeneity for the gains from trade. We compare the gains from trade implied by our semiparametric estimates of the gravity equations of firm exports and a benchmark constant elasticity gravity model of bilateral trade flows. We then evaluate the quantitative importance of incorporating firm heterogeneity with the sufficient elasticity functions governing the extensive and intensive margins of firm exports. In this case, we compare our baseline estimates of the gains from trade to those implied by alternative parametric methodologies that estimate the distribution of firm fundamentals by matching cross-sectional dispersion in firm-level outcomes.

6.1 The Gains from Trade: Measuring the Implications of Firm Heterogeneity

We now investigate how much firm heterogeneity matters for the gains from trade. We do so by comparing the gains from trade implied by our estimates under two assumptions: the constant elasticity specification in Table 3 and the semiparametric specification in Figure 3. For the constant elasticity specification, the results in Section 3.3 indicate that the gains from trade only depend on the domestic spending share and the constant aggregate trade elasticity – that is, the gains are given by the sufficient statistic in Arkolakis et al. (2012). For this reason, we take it to be the benchmark in which firm heterogeneity does not matter for the model's aggregate predictions. In contrast, the non-constant elasticity functions reported in Figure 3 yield gains from trade that differ from those implied by this benchmark specification. For our general specification, we compute the gains from trade with the nonparametric sufficient statistics of Section 3.2 where $\hat{n}_{ii}$ and $\hat{N}_i$ solve the system in Appendix A.5. We

Figure 4: Gains from Trade



Note. Gains from Trade is the percentage change in the real wage implied by moving from autarky to the observed equilibrium in 2012. Gains from trade for semiparametric specification computed with the formula in Section 3.2 for $\hat{n}_{ii}$ and $\hat{N}_i$ solving the system in Appendix A.5 and the baseline spline estimates in Figure 3. Gains from trade for the constant elasticity specification computed with the formula in Section 3.3 and the trade elasticity of five reported in Table 3.

consider our baseline sample for 2012.[22]

Figure 4 compares the gains from trade implied by the constant elasticity and semiparametric specifications. The two specifications yield highly correlated gains from trade. The cross-country correlation of the two measures is 0.96. As pointed out in Section 3.2, this correlation is a consequence of the fact that the domestic trade share remains an important driver of the gains from trade in our general specification – even though it is no longer a sufficient statistic. Notice that firm heterogeneity may still have a substantial impact on the gains from trade of some countries. It yields gains from trade that are more than 20% higher for Luxembourg, Belgium and Netherlands. However, the gains are more than 30% lower for Russia, India and Australia. Overall, when we account for firm heterogeneity, the absolute average change in the gains from trade is 15%.

In Figure 5, we investigate further the mis-measurement in the gains from trade introduced by ignoring the implications of firm heterogeneity. We compute, for each country, the ratio between the gains from trade implied by the semiparametric and the constant elasticity specifications. We plot this ratio against the log of domestic spending ratio in panel (a) and the average exporter firm share in panel (b).

Panel (a) shows that the domestic spending share, a familiar sufficient statistic for the gains from trade, cannot systematically explain the magnitude of the change in the gains

[22]We have data on n_{ij} for 80% of country pairs in our baseline sample, accounting for 71% trade flows in 2012. To compute gains from trade, we impute the exporter firm share for the subset of pairs with missing data using estimates of the constant elasticity gravity equations reported in column of (1) of Table 3.

from trade implied by firm heterogeneity. Notice that both versions of the model yield the same change in domestic trade share moving to autarky, $\Delta \ln x_{ii}$. Intuitively, equation (26) shows that for domestic trade share to play an important role for explaining this difference, the estimated trade elasticities, $1/\theta(n_{ii})$, in the two specifications need to be substantially different. But for the relevant range of n_{ii}, our aggregate trade elasticity estimates in Figure 3 are about 4 for the semiparametric and 5 for the constant elasticity specifications. Thus, there is little room for the domestic trade shares to explain the differences in the welfare gains from trade.

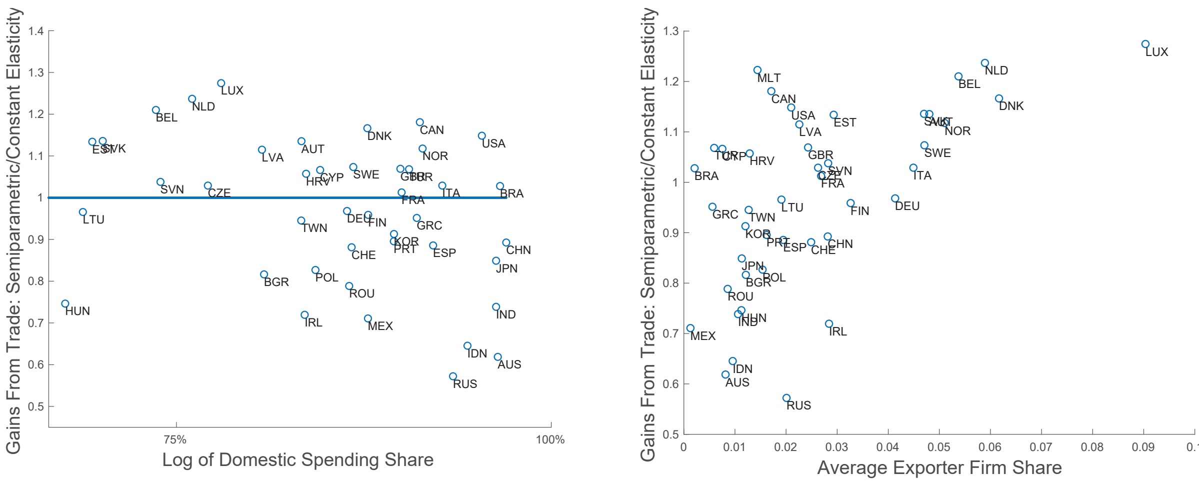
Panel (b) instead shows that the impact of firm heterogeneity on the gains from trade can be explained by the share of firms exporting in a country. There is a 0.6 correlation between the ratio of the gains implied by two specifications and the average exporter firm share (i.e., average n_{ij} across $j \neq i$). Firm heterogeneity amplifies the gains from trade in countries with a higher share of exporting firms. Scale economies is the force behind this role of firm heterogeneity. In particular, when a high fraction of firms exports, there are more resources allocated to covering the fixed cost of entering foreign markets. This implies that competition pressures in the domestic labor market are stronger, leading to a stronger decline in the domestic survival rate, n_{ii}, when the country moves from autarky to the trade equilibrium. This in turn creates higher gains from trade.[23]

6.2 The Gains from Trade: Measuring the Importance of Estimating Semiparametric Gravity Equations of Firm Exports

Lastly, we investigate the quantitative importance of measuring firm heterogeneity in the model by directly estimating the two sufficient elasticity functions for counterfactual analysis. To do this, we compare our baseline estimates of the gains from trade to those implied by parametric distributional assumptions and their associated estimates in the literature. Specifically, we compute the ratio between the gains from trade implied by our semiparametric gravity specification and the gains implied by specifications based on the assumption that firm productivity has either the Truncated Pareto distribution in Melitz and Redding (2015) or the Log-normal distribution in Bas et al. (2017). We use the parameter estimates reported on these papers. Figure 6 presents the cross-country relationship between these ratios and initial trade outcomes.

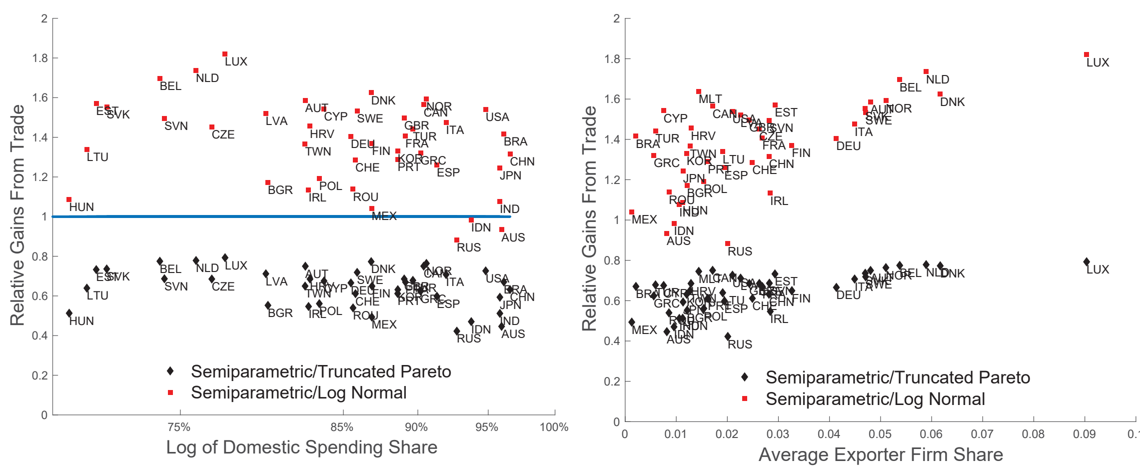
[23]Panel (a) of Figure 18 in Appendix D.1 shows that there is an almost perfect correlation between the ratio of welfare gains and the ratio of changes in n_{ii} implied by the two specifications. To understand why changes in n_{ii} explain the differences in welfare gains, consider the expression for welfare gains in (24). Since the elasticity estimates $\bar{\epsilon}(n)$ for the relevant range of n_{ii} is roughly constant in our semiparametric specification, we have that $\ln \hat{W}_i \approx \frac{\bar{\epsilon}}{\sigma-1} \ln \hat{n}_{ii}$. Thus, the gains from trade increase more in the countries for which n_{ii} falls by more with the semiparametric specification (relative to the constant elasticity specification).

Figure 5: Importance of Firm Heterogeneity for the Gains from Trade



Note. Gains from Trade is the percentage change in the real wage implied by moving from autarky to the observed equilibrium in 2012. Gains from trade for semiparametric specification computed with the formula in Section 3.2 for $\hat{n}_{ii}$ and $\hat{N}_i$ solving the system in Appendix A.5 and the baseline spline estimates in Figure 3. Gains from trade for constant elasticity specification computed with the formula in Section 3.3 and the trade elasticity of five reported in Table 3.

Figure 6: Importance of Functional Form Assumptions for the Gains from Trade



Note. Gains from Trade is the percentage change in the real wage implied by moving from autarky to the observed equilibrium in 2012. For each specification, gains from trade are computed with the formula in Section 3.2 for $\hat{n}_{ii}$ and $\hat{N}_i$ solving the system in Appendix A.5. Gains for semiparametric specification computed with the spline estimates in Figure 3. Gains for Truncated Pareto specification computed with elasticity functions implied by the productivity distribution in Melitz and Redding (2015). Gains for Log-normal specification computed with elasticity functions implied by the productivity distribution in Head et al. (2014).

The diamond-shaped dots in Figure 6 show that the Truncated Pareto specification leads to much higher gains from trade for all countries. On average, our baseline estimates yield gains from trade that are 35% lower. This is a direct consequence of the low trade elasticities implied by the parametrization in Melitz and Redding (2015) – see Figure 7 of Appendix C.3. In contrast, the square-shaped dots in Figure 6 show that the gains from trade are higher for the Log-normal specification. Again, this follows from the average trade elasticity implied by the productivity distribution in Bas et al. (2017). Figure 7 of Appendix C.3 shows that the implied trade elasticity in the log-normal case is higher than our baseline estimate for most values of the exporter firm share.

Both parametric assumptions have quantitatively large impacts on the gains from trade. However, they affect results in opposite directions. While the Truncated Pareto parametrization yields gains from trade that are too large, the Log-normal parametrization leads to gains from trade that are too small. This is a consequence of the opposite implications that these assumptions have for the trade elasticity function. These results indicate that one should be cautious when extrapolating elasticity functions from cross-sectional dispersion in firm-level outcomes.

7 Conclusion

We suggest a new way of modeling and estimating firm heterogeneity in the workhorse monopolistic competition framework. This new approach helps us revisit a number of open questions about the aggregate implications of firm heterogeneity through the lens of a new *not*-parametric point of view. Instead of focusing on parametrically specifying the distribution of various firm-specific wedges, we instead show that they can be folded into two semiparametric gravity equations that intuitively shape how trade costs affect firm-level entry and sales across country pairs. Given the initial equilibrium, the different sources of firm heterogeneity, and any associated parametric assumption imposed, only matter for counterfactual predictions through the shape of these two elasticity functions of firm exports. This characterization also allows us to obtain nonparametric ex-post sufficient statistics for the impact of trade shocks on welfare. Our results indicate that a key new statistic for aggregate gains from trade is the share of exporting firms. We evaluate its impact on the trade elasticity and gains from trade. Our estimates indicate that the impact of firm heterogeneity on the gains from trade crucially depends on the country's average exporter firm share.

References

Adão, R. (2015). Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil. *Unpublished paper, MIT*.

Adao, R., Costinot, A., and Donaldson, D. (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *The American Economic Review*, 107(3):633--689.

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy*, 118(6):1151--1199.

Arkolakis, C., Costinot, A., Donaldson, D., and Rodríguez-Clare, A. (2019a). The elusive pro-competitive effects of trade. *The Review of Economic Studies*, 86(1):46--80.

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2009). New trade models, same old gains? *NBER Working Paper*, (w15628).

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94--130.

Arkolakis, C., Ganapati, S., and Muendler, M.-A. (2019b). The extensive margin of exporting products: A firm-level analysis.

Arkolakis, K., Klenow, P., Demidova, S., and Rodriguez-Clare, A. (2008). The gains from trade with endogenous variety. In *American Economic Review Papers and Proceedings*, volume 98, pages 444--450.

Bartelme, D., Lan, T., and Levchenko, A. (2019). Specialization, market access and medium-term growth.

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

Bernard, A. B., Redding, S. J., and Schott, P. (2011). Multi-product firms and trade liberalization. *Quarterly Journal of Economics*, 126(3):1271--1318.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841--890.

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

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707--1721.

Costinot, A. and Rodriguez-Clare, A. (2013). Trade theory with numbers: Quantifying the consequences of globalization. Technical report, National Bureau of Economic Research.

Dekle, R., Eaton, J., and Kortum, S. (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF Staff Papers*, 55(3):511--540.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms. *Econometrica*, 79(5):1453--1498.

Fajgelbaum, P. D. and Khandelwal, A. K. (2016). Measuring the unequal gains from trade. *The Quarterly Journal of Economics*, 131(3):1113--1180.

Feenstra, R. (1994). New product varieties and the measurement of international prices. *American Economic Review*, 84(1):157--177.

Feenstra, R. C. (2018). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. *Journal of International Economics*, 110:16--27.

Feenstra, R. C. and Weinstein, D. E. (2017). Globalization, markups, and us welfare. *Journal of Political Economy*, 125(4):1040--1074.

Fernandes, A., Klenow, P., Denisse Periola, M., Meleshchuk, S., and Rodriguez-Clare, A. (2017). The intensive margin in trade: Moving beyond pareto. *Unpublished manuscript*.

Fernandes, A. M., Klenow, P. J., Meleshchuk, S., Pierola, M. D., and Rodrıguez-Clare, A. (2015). The intensive margin in trade: moving beyond pareto. *Manuscript, Stanford University*.

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

Head, K. and Mayer, T. (2013). *Gravity equations: Workhorse, toolkit, and cookbook*. Centre for Economic Policy Research.

Head, K., Mayer, T., and Thoenig, M. (2014). Welfare and trade without pareto. *American Economic Review*, 104(5):310--16.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, 2(5):441--487.

Hottman, C. J., Redding, S. J., and Weinstein, D. E. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291--1364.

Hsieh, C.-T., Li, N., Ossa, R., and Yang, M.-J. (2016). Accounting for the new gains from trade liberalization. Technical report, National Bureau of Economic Research.

Kehoe, T. J., Rossbach, J., and Ruhl, K. J. (2015). Using the new products margin to predict the industry-level impact of trade reform. *Journal of International Economics*, 96(2):289--297.

Kehoe, T. J. and Ruhl, K. J. (2013a). How important is the new goods margin in international trade? *Journal of Political Economy*, 121(2):358 -- 392.

Kehoe, T. J. and Ruhl, K. J. (2013b). How important is the new goods margin in international trade? *Journal of Political Economy*, 121(2):358 -- 392.

Krugman, P. (1979). Increasing returns monopolistic competition and international trade. *Journal of International Economics*, 9(4):469--479.

Lind, N. and Ramondo, N. (2018). Trade with correlation. Technical report, National Bureau of Economic Research.

Matsuyama, K. and Ushchev, P. (2017). Beyond ces: Three alternative cases of flexible homothetic demand systems. *Buffett Institute Global Poverty Research Lab Working Paper*, (17-109).

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695--1725.

Melitz, M. J. and Ottaviano, G. I. P. (2008). Market size, trade, and productivity. *The Review of Economic Studies*, 75(1):295--316.

Melitz, M. J. and Redding, S. J. (2013). New trade models, new welfare implications. Technical report, National Bureau of Economic Research.

Melitz, M. J. and Redding, S. J. (2014). Heterogeneous firms and trade. In *Handbook of international economics*, volume 4, pages 1--54. Elsevier.

Melitz, M. J. and Redding, S. J. (2015). New trade models, new welfare implications. *American Economic Review*, 105(3):1105--46.

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

Redding, S. J. (2011). Theories of heterogeneous FIrms and trade. *Annual Review of Economics*, 3:77--105.

Silva, J. M. C. S. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4):641--658.

Zhelobodko, E., Kokovin, S., Parenti, M., and Thisse, J.-F. (2011). Monopolistic competition in general equilibrium: Beyond the CES. *mimeo*.

A Theory Appendix

A.1 Proof of Lemma 1

Part 1. It follows immediately from the derivations in Section (2.2).

Part 3. To derive the labor market clearing condition notice that there are three sources of demand for labor: production of goods, fixed-cost of entering a market and fixed-cost of creating a variety. Thus,

$$w_i\bar{L}_i = \sum_j N_i Pr[\omega \in \Omega_{ij}] \left(1 - \frac{1}{\sigma}\right) E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] + \sum_j N_i Pr[\omega \in \Omega_{ij}] w_i \bar{f}_{ij} E\left[f_{ij}(\omega) | \omega \in \Omega_{ij}\right] + N_i w_i \bar{F}_i$$

From the free entry condition, we know that

$$w_i\bar{F}_i = \sum_j \mathbb{E}\left[\max\left\{\pi_{ij}(\omega);\ 0\right\}\right] = \sum_j Pr[\omega \in \Omega_{ij}] \left(\frac{1}{\sigma} E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] - w_i \bar{f}_{ij} E\left[f_{ij}(\omega) | \omega \in \Omega_{ij}\right]\right),$$

which implies that

$$w_i\bar{L}_i = \sum_j N_i Pr[\omega \in \Omega_{ij}] E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right].$$

Thus, since $\bar{x}_{ij} \equiv E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right]$ and $n_{ij} = Pr[\omega \in \Omega_{ij}]$, this immediately implies (19).

Part 4. Since $p_{ij}(\omega) = \frac{\sigma}{\sigma-1} \frac{\bar{\tau}_{ij} w_i}{\bar{a}_i} \frac{\tau_{ij}(\omega)}{a_i(\omega)}$, the expression for $P_j^{1-\sigma}$ in (2) implies that

$$P_j^{1-\sigma} = \sum_i \left[\bar{b}_{ij} \left(\frac{\sigma}{\sigma - 1} \frac{\bar{\tau}_{ij}}{\bar{a}_i}\right)^{1-\sigma}\right] \left(w_i^{1-\sigma}\right) \int_{\Omega_{ij}} \left(b_{ij}(\omega)\right) \left(\frac{\tau_{ij}(\omega)}{a_i(\omega)}\right)^{1-\sigma} d\omega$$

Using the definitions in (4), we can write this expression as

$$P_j^{1-\sigma} = \sum_i \bar{r}_{ij} \left(w_i^{1-\sigma}\right) \int_{\Omega_{ij}} r_{ij}(\omega)\ d\omega$$

Notice that $\int_{\Omega_{ij}} r_{ij}(\omega)\ d\omega = N_i Pr[\omega \in \Omega_{ij}] E[r | \omega \in \Omega_{ij}] = N_i n_{ij} \bar{\rho}_{ij}(n_{ij})$. This immediately yields expression (20).

Part 5. We start by writing

$$\begin{aligned}
\mathbb{E}\left[\max\left\{\pi_{ij}(\omega);\ 0\right\}\right] &= Pr[\omega \in \Omega_{ij}] E\left[\pi_{ij}(\omega) | \omega \in \Omega_{ij}\right] + Pr[\omega \notin \Omega_{ij}] 0 \\
&= Pr[\omega \in \Omega_{ij}] \left(\tfrac{1}{\sigma} E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] - w_i \bar{f}_{ij} E\left[f_{ij}(\omega) | \omega \in \Omega_{ij}\right]\right) \\
&= n_{ij} \left(\tfrac{1}{\sigma} \bar{x}_{ij} - w_i \bar{f}_{ij} E\left[r_{ij}(\omega)/e_{ij}(\omega) | \omega \in \Omega_{ij}\right]\right)
\end{aligned}$$

where the second equality follows from the expression for $\pi_{ij}(\omega) = (1/\sigma) R_{ij}(\omega) - w_i \bar{f}_{ij} f_{ij}(\omega)$, and the third equality follows from the definitions of $\bar{x}_{ij} \equiv E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right]$ and $e_{ij}(\omega) \equiv r_{ij}(\omega)/f_{ij}(\omega)$.

By defining $e_{ij}^* \equiv \sigma \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j}\right)^\sigma \frac{P_j}{E_j}\right]$, we can write

$$E\left[r_{ij}(\omega)/e_{ij}(\omega) | \omega \in \Omega_{ij}\right] = \int_{e_{ij}^*}^\infty \frac{1}{e} \left[\int_0^\infty r dH_{ij}^r(r|e)\right] \frac{dH^e(e)}{1 - H^e(e_{ij}^*)}$$

Consider the transformation $n = 1 - H_{ij}(e)$ such that $e = \bar{\epsilon}_{ij}(n)$. In this case, $dH_{ij}(e) = -dn$ and

$n_{ij} = 1 - H_{ij}(e^*_{ij})$, which implies that

$$E\left[r_{ij}(\omega)/e_{ij}(\omega)|\omega \in \Omega_{ij}\right] = \frac{1}{n_{ij}} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} dn.$$

Thus,

$$\mathbb{E}\left[\max\left\{\pi_{ij}(\omega);\ 0\right\}\right] = \frac{1}{\sigma} n_{ij}\bar{x}_{ij} - w_i \bar{f}_{ij} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn.$$

Thus, the free entry condition is

$$\sigma w_i \bar{F}_i = \sum_j n_{ij}\bar{x}_{ij} - \sum_j \left(\sigma w_i \bar{f}_{ij}\right) \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn. \tag{41}$$

Notice that the summation of (12) and (15) implies that

$$\ln\left(\sigma w_i \bar{f}_{ij}\right) = \ln \bar{x}_{ij} - \ln \bar{\rho}_{ij}(n_{ij}) + \ln \bar{\epsilon}_{ij}(n_{ij})$$

which yields

$$\sigma w_i \bar{F}_i = \sum_j n_{ij}\bar{x}_{ij} - \sum_j \bar{x}_{ij} \frac{\bar{\epsilon}_{ij}(n_{ij})}{\bar{\rho}_{ij}(n_{ij})} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn.$$

By substituting the definition of $\bar{\rho}_{ij}(n)$, we can write the free entry condition as

$$\sigma w_i \bar{F}_i = \sum_j n_{ij}\bar{x}_{ij} - \sum_j n_{ij}\bar{x}_{ij} \frac{\bar{\epsilon}_{ij}(n_{ij})}{\int_0^{n_{ij}} \rho_{ij}(n)\ dn} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn. \tag{42}$$

Using the market clearing condition in (19), we have that

$$\frac{1}{N_i} = \sigma \frac{\bar{F}_i}{\bar{L}_i} + \sum_j \frac{n_{ij}\bar{x}_{ij}}{w_i \bar{L}_i} \frac{\bar{\epsilon}_{ij}(n_{ij})}{\int_0^{n_{ij}} \rho_{ij}(n)\ dn} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn,$$

which immediately yields equation (21).

A.2 Proof of Proposition 1

Part 1. We start by pointing out that equation (17) implies that knowledge of $\bar{\rho}_{ij}(n)$ implies knowledge of $\rho_{ij}(n)$. We then use the equilibrium conditions in Proposition 1 to obtain a system of equations for the changes in $\{\{n_{ij}, \bar{x}_{ij}\}_j, P_i, N_i, E_i, w_i\}$ given changes in $\left\{\bar{T}_i, \bar{L}_i, \bar{F}_i, \{\bar{r}_{ij}, \bar{f}_{ij}\}_j\right\}_i$.

1. The extensive and intensive margins of firm-level sales, n_{ij} and $\bar{x}_{ij}$, in (12) and (15) imply

$$\frac{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\epsilon}_{ij}(n_{ij})} = \frac{\hat{\bar{f}}_{ij}}{\hat{\bar{r}}_{ij}} \left[\left(\frac{\hat{w}_i}{\hat{P}_j}\right)^{\sigma} \frac{\hat{P}_j}{\hat{E}_j}\right], \tag{43}$$

$$\hat{\bar{x}}_{ij} = \hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \left[\left(\frac{\hat{w}_i}{\hat{P}_j}\right)^{1-\sigma} \hat{E}_j\right]. \tag{44}$$

2. Let $\iota_i \equiv w_i L_i / E_i = \left(\sum_d X_{id}\right) / \left(\sum_o X_{oi}\right)$ be the output-spending ratio in country i in the initial equilibrium.

The spending equation in (9) implies

$$\hat{E}_i = \iota_i \left(\hat{w}_i \hat{\bar{L}}_i \right) + (1 - \iota_i) \hat{\bar{T}}_i, \tag{45}$$

3. Let $y_{ij} \equiv (N_i n_{ij} \bar{x}_{ij}) / (w_i L_i) = X_{ij} / \left(\sum_{j'} X_{ij'} \right)$ be the share of i's revenue from sales to j. The labor market clearing condition in (19) implies

$$\hat{w}_i \hat{\bar{L}}_i = \sum_j y_{ij} \left(\hat{N}_i \hat{n}_{ij} \hat{\bar{x}}_{ij} \right). \tag{46}$$

4. The price index (20) implies

$$\begin{aligned} \hat{P}_j^{1-\sigma} &= \sum_i \frac{\bar{r}_{ij} w_i^{1-\sigma} \bar{\rho}_{ij}(n_{ij}) n_{ij} N_i}{P_j^{1-\sigma}} \left(\hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{w}_i^{1-\sigma} \hat{n}_{ij} \hat{N}_i \right) \\ &= \sum_i \frac{\bar{r}_{ij} w_i^{1-\sigma} \bar{\rho}_{ij}(n_{ij}) n_{ij} N_i E_j P_j^{\sigma-1}}{\sum_o \bar{r}_{oj} w_o^{1-\sigma} \bar{\rho}_{oj}(n_{oj}) n_{oj} N_o E_j P_j^{\sigma-1}} \left(\hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{w}_i^{1-\sigma} \hat{n}_{ij} \hat{N}_i \right) \\ &= \sum_i \frac{\bar{x}_{ij} n_{ij} N_i}{\sum_o \bar{x}_{oj} n_{oj} N_o} \left(\hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{w}_i^{1-\sigma} \hat{n}_{ij} \hat{N}_i \right) \end{aligned}$$

Let $x_{ij} \equiv (N_i n_{ij} \bar{x}_{ij}) / (\sum_o \bar{x}_{oj} n_{oj} N_o) = X_{ij} / (\sum_o X_{oj})$ be the spending share of country j on country i. Thus,

$$\hat{P}_j^{1-\sigma} = \sum_i x_{ij} \left(\hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{w}_i^{1-\sigma} \hat{n}_{ij} \hat{N}_i \right). \tag{47}$$

5. The free entry condition in (21) implies

$$N_i \hat{N}_i = \left[\sigma \frac{\bar{F}_i}{\bar{L}_i} \frac{\hat{\bar{F}}_i}{\hat{\bar{L}}_i} + \sum_j \frac{n_{ij} \bar{x}_{ij}}{w_i \bar{L}_i} \frac{\hat{n}_{ij} \hat{\bar{x}}_{ij}}{\hat{w}_i \hat{\bar{L}}_i} \frac{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})} \, dn} \right]^{-1}$$

Using (21) to substitute for $\sigma \frac{\bar{F}_i}{\bar{L}_i}$,

$$\hat{N}_i = \left[\left(1 - \sum_j y_{ij} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})} \, d} \right) \frac{\hat{\bar{F}}_i}{\hat{\bar{L}}_i} + \sum_j y_{ij} \frac{\hat{n}_{ij} \hat{\bar{x}}_{ij}}{\hat{w}_i \hat{\bar{L}}_i} \frac{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})} \, dn} \right]^{-1}$$

$$\hat{N}_i = \left[\left(1 - \sum_j y_{ij} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})} \, d} \right) \frac{\hat{\bar{F}}_i}{\hat{\bar{L}}_i} + \sum_j y_{ij} \frac{\hat{n}_{ij} \hat{\bar{x}}_{ij}}{\hat{w}_i \hat{\bar{L}}_i} \frac{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})} \, dn} \right]^{-1}. \tag{48}$$

A.3 Proof of Proposition 2

Part 1. We start by totally differentiating the equilibrium equations in Lemma 1. Equation (12) implies

$$\varepsilon_{ij}(n_{ij}) d \ln n_{ij} = (\sigma - 1) d \ln \bar{r}_{ij} + \sigma d \ln w_i - (\sigma - 1) d \ln P_j - d \ln E_j$$

Since equation 9 implies $d \ln E_j = \iota_j d \ln w_j$,

$$\varepsilon_{ij}(n_{ij}) d \ln n_{ij} = (\sigma - 1) d \ln \bar{r}_{ij} + \sigma d \ln w_i - (\sigma - 1) d \ln P_j - \iota_j d \ln w_j. \tag{49}$$

Using again $d \ln E_j = \iota_j d \ln w_j$, equation (15) yields

$$d\ln\bar{x}_{ij} = (1-\sigma)d\ln\bar{\tau}_{ij} + \varrho_{ij}(n_{ij})d\ln n_{ij} - (\sigma-1)d\ln w_i + (\sigma-1)d\ln P_j + \iota_j d\ln w_j.$$

The sum of the two equations above implies that

$$d\ln\bar{x}_{ij} = d\ln w_i + (\varrho_{ij}(n_{ij}) - \varepsilon_{ij}(n_{ij}))\, d\ln n_{ij} \tag{50}$$

By combining the market clearing condition in (19) with the version of the free entry condition in (41), we have that

$$\frac{\sigma\bar{F}_i}{\bar{L}_i} = \frac{1}{N_i} - \sum_j \frac{\sigma\bar{f}_{ij}}{\bar{L}_i}\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn,$$

which implies that

$$\begin{aligned} -\tfrac{1}{N_i}d\ln N_i &= \textstyle\sum_j \frac{\sigma\bar{f}_{ij}}{\bar{L}_i}\frac{\rho_{ij}(n_{ij})}{\bar{\epsilon}_{ij}(n_{ij})} n_{ij} d\ln n_{ij} \\ &= \textstyle\sum_j \frac{\bar{x}_{ij}n_{ij}}{w_i\bar{L}_i}\frac{\rho_{ij}(n_{ij})}{\bar{\rho}_{ij}(n_{ij})} d\ln n_{ij} \\ &= \textstyle\sum_j \frac{\bar{x}_{ij}n_{ij}}{w_i\bar{L}_i}\left(1+\varrho_{ij}(n_{ij})\right) d\ln n_{ij} \\ &= \textstyle\frac{1}{N_i}\sum_j y_{ij}\left(1+\varrho_{ij}(n_{ij})\right) d\ln n_{ij}, \end{aligned}$$

where the second equality uses $\bar{x}_{ij} = \frac{\bar{\rho}_{ij}(n_{ij})}{\bar{\epsilon}_{ij}(n_{ij})}\sigma\bar{f}_{ij}w_i$, the third equality uses (17), and the fourth uses$y_{ij} \equiv N_i\bar{x}_{ij}n_{ij}/w_i\bar{L}_i$.

Thus,

$$d\ln N_i = -\sum_j y_{ij}\left(1+\varrho_{ij}(n_{ij})\right) d\ln n_{ij} \tag{51}$$

Equation (19) implies

$$d\ln w_i = \sum_j y_{ij}\left(d\ln N_i + d\ln n_{ij} + d\ln\bar{x}_{ij}\right),$$

which combined with (50) implies

$$-d\ln N_i = \sum_j y_{ij}\left(1+\varrho_{ij}(n_{ij}) - \varepsilon_{ij}(n_{ij})\right) d\ln n_{ij}.$$

The combination of this equation and (51) implies that

$$\sum_j y_{ij}\varepsilon_{ij}(n_{ij})d\ln n_{ij} = 0. \tag{52}$$

Finally, equation (20) implies

$$(1-\sigma)d\ln P_j = \sum_i x_{ij}\left((1-\sigma)d\ln\bar{\tau}_{ij} - (\sigma-1)d\ln w_i + \left(1+\varrho_{ij}(n_{ij})\right)d\ln n_{ij} + d\ln N_i\right) \tag{53}$$

Equations (49), (51), (52) and (53) form a system that determines $\{d\ln n_{ij}, d\ln N_i, d\ln P_i, d\ln w_i\}_{i.j}$ for any arbitrary set of trade cost shocks $\{d\ln\bar{\tau}_{ij}\}_{i,j}$. We now establish Part 1 of Proposition 2 by reducing this system to two sets of equations determining $\{d\ln P_i, d\ln w_i\}_i$ in terms of σ, $\{\theta_{ij}(n_{ij}), n_{ij}, X_{ij}\}_{i,j}$, and $\{d\ln\bar{\tau}_{ij}\}_{i,j}$. To this end, note that the definition of $\theta_{ij}(n_{ij})$ in (18) implies that $\frac{1+\varrho_{ij}(n_{ij})}{\varepsilon_{ij}(n_{ij})} = 1 + \frac{\theta_{ij}(n_{ij})}{1-\sigma}$.

Thus, equations (53) and (51) imply

$$\begin{array}{rcl}(1-\sigma)d\ln P_j &=& \sum_i x_{ij}\left((1-\sigma)d\ln\bar{\tau}_{ij}-(\sigma-1)d\ln w_i+\varepsilon_{ij}(n_{ij})d\ln n_{ij}\right)\\ &+& \sum_i x_{ij}\left[\left(\frac{\theta_{ij}(n_{ij})}{1-\sigma}\right)\varepsilon_{ij}(n_{ij})d\ln n_{ij}+d\ln N_i\right]\end{array} \tag{54}$$

$$d\ln N_i=\sum_j y_{ij}\left(\frac{\theta_{ij}(n_{ij})}{\sigma-1}\right)\varepsilon_{ij}(n_{ij})d\ln n_{ij}. \tag{55}$$

where the derivation of (55) uses (52).

By substituting the second equation into the first, we get that

$$\begin{array}{rcl}(1-\sigma)d\ln P_j &=& \sum_i x_{ij}\left((1-\sigma)d\ln\bar{\tau}_{ij}-(\sigma-1)d\ln w_i+\varepsilon_{ij}(n_{ij})d\ln n_{ij}\right)\\ &-& \sum_i x_{ij}\left[\left(\frac{\theta_{ij}(n_{ij})}{\sigma-1}\right)\varepsilon_{ij}(n_{ij})d\ln n_{ij}-\sum_d y_{id}\left(\frac{\theta_{id}(n_{id})}{\sigma-1}\right)\varepsilon_{id}(n_{id})d\ln n_{id}\right].\end{array}$$

By substituting (49) into this expression,

$$\begin{array}{rcl}\iota_j d\ln w_j-\sum_i x_{ij}d\ln w_i &=& -\sum_i x_{ij}\left[\theta_{ij}(n_{ij})\left(d\ln\bar{\tau}_{ij}+\frac{\sigma}{\sigma-1}d\ln w_i-d\ln P_j-\frac{\iota_j}{\sigma-1}d\ln w_j\right)\right]\\ &+& \sum_i x_{ij}\sum_d y_{id}\theta_{id}(n_{id})\left(d\ln\bar{\tau}_{id}+\frac{\sigma}{\sigma-1}d\ln w_i-d\ln P_d-\frac{\iota_d}{\sigma-1}d\ln w_d\right)\end{array}$$

Thus,

$$\sum_j v^p_{ij}d\ln P_j-\sum_j v^w_{ij}d\ln w_j=d\ln\tau^p_i \tag{56}$$

$$\begin{array}{rcl}v^w_{ij} &\equiv& 1[i=j]\left(1-\sum_j\frac{x_{ji}\theta_{ji}(n_{ji})}{\sigma-1}\right)\iota_i\\ &+& \left[\left(\sum_o x_{oi}y_{oj}\theta_{oj}(n_{oj})\right)\left(\frac{\iota_j}{\sigma-1}\right)-x_{ji}\left(1-\frac{\sigma}{\sigma-1}\left(\theta_{ji}(n_{ji})-\sum_d y_{jd}\theta_{jd}(n_{jd})\right)\right)\right]\end{array} \tag{57}$$

$$v^p_{ij}\equiv 1[i=j]\left(\sum_o x_{oi}\theta_{oi}(n_{oi})\right)-\left(\sum_o x_{oi}y_{oj}\theta_{oj}(n_{oj})\right) \tag{58}$$

$$d\ln\tau^p_i\equiv\sum_j x_{ji}\left(\theta_{ji}(n_{ji})d\ln\bar{\tau}_{ji}-\sum_d y_{jd}\theta_{jd}(n_{jd})d\ln\bar{\tau}_{jd}\right) \tag{59}$$

Equations (52) and (49) imply

$$\sum_j y_{ij}\left(d\ln\bar{\tau}_{ij}+\frac{\sigma}{\sigma-1}d\ln w_i-d\ln P_j-\frac{\iota_j}{\sigma-1}d\ln w_j\right)=0$$

Thus,

$$\sum_j m^w_{ij}d\ln w_j-\sum_j m^p_{ij}d\ln P_j=d\ln\tau^w_j \tag{60}$$

$$m^w_{ij}\equiv 1[i=j]\frac{\sigma}{\sigma-1}-y_{ij}\frac{\iota_j}{\sigma-1} \tag{61}$$

$$m^p_{ij}\equiv y_{ij} \tag{62}$$

$$d\ln\tau^w_j\equiv-\sum_j y_{ij}d\ln\bar{\tau}_{ij}. \tag{63}$$

Let us use bold letters to denote vectors, $\boldsymbol{v}=[v_i]_i$ and bold bar variables to denote matrices, $\bar{\boldsymbol{v}}=[v_{ij}]_{i,j}$.

Thus, equations (56)–(60) imply

$$\begin{aligned} \bar{\boldsymbol{v}}^p d\ln \boldsymbol{P} - \bar{\boldsymbol{v}}^w d\ln \boldsymbol{w} &= d\ln \boldsymbol{\tau}^p \\ -\bar{\boldsymbol{m}}^p d\ln \boldsymbol{P} + \bar{\boldsymbol{m}}^w d\ln \boldsymbol{w} &= d\ln \boldsymbol{\tau}^w \end{aligned}$$

We then use the first equation to solve for the price index change,

$$d\ln \boldsymbol{P} = \left(\bar{\boldsymbol{v}}^p\right)^{-1}\left(\bar{\boldsymbol{v}}^w d\ln \boldsymbol{w} + d\ln \boldsymbol{\tau}^1\right), \tag{64}$$

which we then substitute into the second equation to obtain,

$$\left[\bar{\boldsymbol{m}}^w - \bar{\boldsymbol{m}}^p\left(\bar{\boldsymbol{v}}^p\right)^{-1}\bar{\boldsymbol{v}}^w\right] d\ln \boldsymbol{w} = d\ln \boldsymbol{\tau}^w + \bar{\boldsymbol{m}}^p\left(\bar{\boldsymbol{v}}^p\right)^{-1} d\ln \boldsymbol{\tau}^p. \tag{65}$$

Notice that, because of the numeraire choice, the solving (65) requires dropping one row and one column by setting $d\ln w_n = 0$ for some arbitrary country n.

Recall that $\{X_{ij}\}_{ij}$ immediately yields $\{\iota_j, x_{ij}, y_{ij}\}_{i,j}$. Thus, the system (64)–(65) determines $\{d\ln P_i, d\ln w_i\}_i$ as a function of $\{d\ln \bar{\tau}_{ij}\}_{i,j}$ where all coefficients depend only on σ and $\{\theta_{ij}(n_{ij}), X_{ij}\}_{i,j}$. This establishes that $\frac{d\ln w_i}{d\ln \bar{\tau}_{od}}$ and $\frac{d\ln P_i}{d\ln \bar{\tau}_{od}}$ are functions of $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$. Notice that $d\ln E_i = \iota_i d\ln w_i$, so $\frac{d\ln E_i}{d\ln \bar{\tau}_{od}}$ is also a function of $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$.

To obtain changes in the number of entrants, we combine equations (55) and (49):

$$d\ln N_i = \sum_j y_{ij}\left(\frac{\theta_{ij}(n_{ij})}{\sigma-1}\right)\left((\sigma-1)d\ln \bar{\tau}_{ij} + \sigma d\ln w_i - (\sigma-1)d\ln P_j - \iota_j d\ln w_j\right).$$

This implies that $d\ln N_i$ is a function of $\{d\ln \bar{\tau}_{ij}\}_j$, $\{\theta_{ij}(n_{ij})\}_j$, $\{d\ln P_j, d\ln w_j\}_j$, and $\{X_{ij}\}_{ij}$. Thus, given that $\{d\ln P_j, d\ln w_j\}_j$ is a function of $\{d\ln \bar{\tau}_{km}\}_{km}$ and $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$, $\frac{d\ln N_i}{d\ln \bar{\tau}_{od}}$ is a function of $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$.

Finally,

$$\begin{aligned} d\ln X_{ij} &= d\ln N_i + d\ln n_{ij} + d\ln \bar{x}_{ij} \\ &= d\ln N_i + d\ln w_i + \left(1 + \varrho_{ij}(n_{ij}) - \varepsilon_{ij}(n_{ij})\right) d\ln n_{ij} \\ &= d\ln N_i + d\ln w_i - \theta_{ij}(n_{ij})\tfrac{\varepsilon_{ij}(n_{ij})}{\sigma-1} d\ln n_{ij} \end{aligned}$$

where the first equality follows from $X_{ij} \equiv N_i n_{ij} \bar{x}_{ij}$, the second equality follows from (50), and the third equality follows from the definition of $\theta_{ij}(n_{ij})$ in (18).

Using (49),

$$d\ln X_{ij} = d\ln N_i + d\ln w_i - \theta_{ij}(n_{ij})\left(d\ln \bar{\tau}_{ij} + \frac{\sigma}{\sigma-1} d\ln w_i - d\ln P_j - \frac{\iota_j}{\sigma-1} d\ln w_j\right).$$

Thus, since $\{d\ln P_j, d\ln w_j, d\ln N_j\}_j$ is a function of $\{d\ln \bar{\tau}_{km}\}_{km}$ and $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$, $\frac{d\ln X_{ij}}{d\ln \bar{\tau}_{od}}$ is a function of $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$.

Part 2. From equation (49),

$$\frac{d\ln n_{ij}}{d\ln \bar{\tau}_{od}} = \frac{1}{\varepsilon_{ij}(n_{ij})}\left[(\sigma-1)1[od=id] + \sigma\frac{d\ln w_i}{d\ln \bar{\tau}_{od}} - (\sigma-1)\frac{d\ln P_j}{d\ln \bar{\tau}_{od}} - \iota_j \frac{d\ln w_j}{d\ln \bar{\tau}_{od}}\right]$$

Since $\frac{d\ln w_i}{d\ln \bar{\tau}_{od}}$ and $\frac{d\ln P_j}{d\ln \bar{\tau}_{od}}$ are functions of $(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$, then $\frac{d\ln n_{ij}}{d\ln \bar{\tau}_{od}}$ is a function of

$(\sigma, \{\theta_{km}(n_{km})\}_{km}, \{X_{km}\}_{km})$ and $\varepsilon_{ij}(n_{ij})$.

A.4 Proof of the expressions in Section 3.2

Equation (24). Assume that $\iota_i = 1$. If $\hat{\bar{r}}_{ii} = \hat{\bar{f}}_{ii} = 1$, then equation (12) implies that

$$\frac{\bar{\epsilon}_{ii}(n_{ii}\hat{n}_{ii})}{\bar{\epsilon}_{ii}(n_{ii})} = \left(\frac{w_i}{P_i}\right)^{\sigma-1},$$

which immediately yields the expression in (24).

Equation (25). Recall that $\hat{x}_{ii} \equiv \hat{N}_i\hat{n}_{ii}\hat{\bar{x}}_{ii}/\hat{E}_i$. If $\hat{\bar{r}}_{ii} = 1$, then equation (44) implies that

$$\hat{x}_{ii} = \hat{N}_i\hat{n}_{ii}\frac{\bar{\rho}_{ii}(n_{ii}\hat{n}_{ii})}{\bar{\rho}_{ii}(n_{ii})}\left(\frac{\hat{w}_i}{\hat{P}_i}\right)^{1-\sigma},$$

which immediately yields the expression in (25).

Equation (26). For the case of balanced trade with $\iota_i = 1$, equation (49) implies that

$$\varepsilon_{ii}(n_{ii})d\ln n_{ii} = (\sigma-1)d\ln\bar{\tau}_{ii} + (\sigma-1)d\ln w_i/P_i.$$

Equation (50) implies that

$$\begin{aligned} d\ln x_{ii} &= d\ln N_i + d\ln n_{ii} + d\ln\bar{x}_{ii} - d\ln E_i \\ &= d\ln N_i + (1+\varrho_{ij}(n_{ij}) - \varepsilon_{ij}(n_{ij}))\, d\ln n_{ij} \\ &= d\ln N_i + \tfrac{(1+\varrho_{ij}(n_{ij})-\varepsilon_{ij}(n_{ij}))}{\varepsilon_{ij}(n_{ij})}\varepsilon_{ij}(n_{ij})d\ln n_{ij} \end{aligned}$$

Using the fact that $\varepsilon_{ii}(n_{ii})d\ln n_{ii} = (\sigma-1)d\ln w_i/P_i$,

$$-\left[(\sigma-1)\left(1-\frac{1+\varrho_{ij}(n_{ij})}{\varepsilon_{ij}(n_{ij})}\right)\right]d\ln w_i/P_i = d\ln x_{ii}/N_i$$

Together with the definition of $\theta_{ii}(n_{ii})$ in (18), this immediately yields the expression in (26).

A.5 Gains from Trade

We now compute the gains from trade in our model. We assume that $\hat{\bar{\tau}}_{ij} \to \infty$ for all $i \neq j$, and that $\hat{\bar{a}}_i = \hat{\bar{F}}_i = \hat{\bar{f}}_{ij} = \hat{\bar{\tau}}_{ii} = \hat{\bar{L}}_i = 1$ for all i and j.

Corollary 2. *Consider an economy moving from the trade equilibrium to the autarky equilibrium with $\hat{T}_i^A = 0$. The change in the real wage is given by (24) or (25) where $\hat{n}_{ii}^A$ and $\hat{N}_i^A$ solve*

$$\frac{\epsilon_{ii}\left(n_{ii}\hat{n}_{ii}^A\right)}{\epsilon_{ii}\left(n_{ii}\right)} = \left(\frac{x_{ii}}{\iota_i}\right)\left(\hat{n}_{ii}^A\hat{N}_i^A\right)\frac{\bar{\rho}_{ii}\left(n_{ii}\hat{n}_{ii}^A\right)}{\bar{\rho}_{ii}\left(n_{ii}\right)}, \tag{66}$$

$$\left(1-\sum_j \frac{X_{ij}}{\sum_{j'} X_{ij'}} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\, d}\right) \hat{N}_i^A = 1 - \frac{\int_0^{n_{ii}\hat{n}_{ii}^A} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ii}\hat{n}_{ii}^A} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ii}\hat{n}_{ii}^A)}\, dn}. \tag{67}$$

In order to compute the gains from trade using (24) or (25), we need to compute changes in n_{ii} and N_i when the economy moves to autarky (i.e, $\hat{\bar{\tau}}_{ij} \to \infty$ for all $i \neq j$). Equation (66) captures the change in the profitability of the domestic market that determines the change in the domestic survival rate of firms (given $\hat{N}_i^A$). This expression follows immediately from equalizing the two expressions for changes in the real wage in (24) and (25). Equation (67) is the free entry condition that determines the change in the number of entrants when the country moves to autarky. The left hand size of (67) is the profit/revenue ratio (inclusive of entry costs) that firms have in different markets in the initial equilibrium. The right hand size is the profit/revenue ratio that entrants have in the domestic market in the autarky equilibrium.

A.5.1 Proof of Corollary 2

To simplify the notation, we drop the superscript A and use "hat" variables to denote the change from the initial equilibrium to the autarky equilibrium. We assume that $\hat{\bar{\tau}}_{ij} \to \infty$ for all $i \neq j$, and that $\hat{\bar{a}}_i = \hat{\bar{F}}_i = \hat{\bar{f}}_{ij} = \hat{\bar{\tau}}_{ii} = \hat{\bar{L}}_i = 1$ for all i and j. We set the wage of i to be the numerarie, $w_i \equiv 1$, so that $\hat{w}_i = 1$. Equation (47) implies that

$$\left(\hat{P}_i\right)^{1-\sigma} = x_{ii} \frac{\bar{\rho}_{ii}\left(n_{ii}\hat{n}_{ii}\right)}{\bar{\rho}_{ii}\left(n_{ii}\right)} \left(\hat{n}_{ii}\hat{N}_i\right) \tag{68}$$

From equation (43), we get that, for all $i \neq j$, $\epsilon_{ij}\left(n_{ij}\hat{n}_{ij}\right) \to \infty$ and, therefore, $\hat{n}_{ij} = 0$. In addition, it implies that

$$\frac{\epsilon_{ii}\left(n_{ii}\hat{n}_{ii}\right)}{\epsilon_{ii}\left(n_{ii}\right)} = \frac{\left(\hat{P}_i\right)^{1-\sigma}}{\hat{E}_i} \tag{69}$$

Using the fact that $\hat{E}_i = \iota_i$, (68) and (69) imply that

$$\frac{\epsilon_{ii}\left(n_{ii}\hat{n}_{ii}\right)}{\epsilon_{ii}\left(n_{ii}\right)} = \frac{x_{ii}}{\iota_i} \frac{\bar{\rho}_{ii}\left(n_{ii}\hat{n}_{ii}\right)}{\bar{\rho}_{ii}\left(n_{ii}\right)} \left(\hat{n}_{ii}\hat{N}_i\right). \tag{70}$$

From expression (48),

$$\hat{N}_i = \left[\left(1-\sum_j y_{ij} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\, d}\right) + y_{ii}\hat{n}_{ii}\hat{\bar{x}}_{ii} \frac{\int_0^{n_{ii}\hat{n}_{ii}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ii}\hat{n}_{ii}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ii}\hat{n}_{ii})}\, dn}\right]^{-1}$$

$$1 = \hat{N}_i \left(1-\sum_j y_{ij} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\, d}\right) + y_{ii}\hat{N}_i\hat{n}_{ii}\hat{\bar{x}}_{ii} \frac{\int_0^{n_{ii}\hat{n}_{ii}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ii}\hat{n}_{ii}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ii}\hat{n}_{ii})}\, dn}$$

Recall that $\hat{x}_{ii} = \frac{\hat{N}_i\hat{n}_{ii}\hat{\bar{x}}_{ii}}{\hat{E}_i} = 1/x_{ii}$. Thus, $y_{ii}\hat{N}_i\hat{n}_{ii}\hat{\bar{x}}_{ii} = y_{ii}\frac{\iota_i}{x_{ii}} = \frac{X_{ii}}{w_i \bar{L}_i}\frac{E_i}{X_{ii}}\frac{w_i\bar{L}_i}{E_i} = 1$ and, therefore,

$$\left(1-\sum_j y_{ij} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\, d}\right) \hat{N}_i = 1 - \frac{\int_0^{n_{ii}\hat{n}_{ii}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ii}\hat{n}_{ii}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ii}\hat{n}_{ii})}\, dn}.$$

A.6 Proof of Part 3 of Corollary 1

Assume that that $\varepsilon_{ij}(n) = \varepsilon_i$ and $\varrho_{ij}(n) = \varrho_i$ for all n, i and j. By the definition of θ_{ij}, we immediately get that $\theta_{ij} \equiv \theta_i = (\sigma - 1)\left(1 - \frac{1+\varrho_i}{\varepsilon_i}\right)$. Note also that, by equation (17), $\rho_i(n) = (1 + \varrho_i(n))\, \bar{\rho}_i(n)$ and, therefore, $\rho_i(n) = (1 + \varrho)\, n^{\varrho_i}$. Consider the free entry condition in equation (42):

$$\begin{aligned} \sigma w_i \bar{F}_i &= \sum_j n_{ij}\bar{x}_{ij}\left(1 - \frac{\int_0^{n_{ij}} n^{\varrho_i - \varepsilon_i}\, dn}{n_{ij}^{-\varepsilon_i}\int_0^{n_{ij}} n^{\varrho_i}\, dn}\right) \\ &= \sum_j n_{ij}\bar{x}_{ij}\left(1 - \frac{1+\varrho_i}{1+\varrho_i-\varepsilon_i}\frac{n^{1+\varrho_i-\varepsilon_i}}{n^{1+\varrho_i-\varepsilon_i}}\right) = \left(\frac{-\varepsilon_i}{1+\varrho_i-\varepsilon_i}\right)\sum_j n_{ij}\bar{x}_{ij}. \end{aligned}$$

The market clearing condition in (19) implies that $\sum_j n_{ij}\bar{x}_{ij} = w_i\bar{L}_i/N_i$ and, therefore,

$$N_i = \frac{\bar{L}_i}{\sigma\bar{F}_i}\left(\frac{-\varepsilon_i}{1+\varrho_i-\varepsilon_i}\right) \implies \frac{d\ln N_i}{d\ln\bar{\tau}_{od}} = 0.$$

A.7 Derivation of Equation (36)

By plugging (33) into (12)–(15) we have that

$$\ln\epsilon_{ij}(n_{ij}) = z_{ij}/\kappa^{\epsilon} + \left(\tilde{\sigma}\eta_{ij}^{\tau} + \eta_{ij}^{f}\right) + \left[\ln\sigma w_i\left(\frac{\sigma}{\sigma-1}\frac{w_i}{\bar{a}_i}\right)^{\sigma-1} + \tilde{\sigma}\delta_i^{\tau} + \delta_i^{f}\right] - \left[\ln\left(P_j^{\sigma-1}E_j\right) - \tilde{\sigma}\zeta_j^{\tau} - \zeta_j^{f}\right]$$

$$\ln\bar{x}_{ij} - \ln\bar{\rho}_{ij}(n_{ij}) = -\tilde{\sigma}\kappa^{\tau}z_{ij} - \tilde{\sigma}\eta_{ij}^{\tau} + \left[\ln\left(\frac{\sigma}{\sigma-1}\frac{w_i}{\bar{a}_i}\right)^{1-\sigma} - \tilde{\sigma}\delta_j^{\tau}\right] + \left[\ln\left(P_j^{\sigma-1}E_j\right) - \tilde{\sigma}\zeta_j^{\tau}\right]$$

where $\tilde{\sigma} \equiv \sigma - 1$ and $\kappa^{\epsilon} \equiv 1/(\tilde{\sigma}\kappa^{\tau} + \kappa^{f})$.

This implies that

$$\begin{aligned} -\kappa^{\epsilon}\left(\tilde{\sigma}\eta_{ij}^{\tau} + \eta_{ij}^{f}\right) &= z_{ij} - \kappa^{\epsilon}\ln\epsilon_{ij}(n_{ij}) + \kappa^{\epsilon}\left[\ln\sigma w_i\left(\frac{\sigma}{\sigma-1}\frac{w_i}{\bar{a}_i}\right)^{\sigma-1} + \tilde{\sigma}\delta_i^{\tau} + \delta_i^{f}\right] \\ &\quad -\kappa^{\epsilon}\left[\ln\left(P_j^{\sigma-1}E_j\right) - \tilde{\sigma}\zeta_j^{\tau} - \zeta_j^{f}\right], \end{aligned}$$

$$-\tilde{\sigma}\eta_{ij}^{\tau} = \ln\bar{x}_{ij} + \tilde{\sigma}\kappa^{\tau}z_{ij} - \ln\bar{\rho}_{ij}(n_{ij}) - \left[\ln\left(\frac{\sigma}{\sigma-1}\frac{w_i}{\bar{a}_i}\right)^{1-\sigma} - \tilde{\sigma}\delta_j^{\tau}\right] - \left[\ln\left(P_j^{\sigma-1}E_j\right) - \tilde{\sigma}\zeta_j^{\tau}\right].$$

We can then write

$$\begin{aligned} v_{ij}^{\epsilon} &= z_{ij} - \kappa^{\epsilon}\ln\epsilon_{ij}(n_{ij}) - \delta_i^{\epsilon} - \zeta_j^{\epsilon} \\ v_{ij}^{\rho} &= \ln\bar{x}_{ij} + \tilde{\sigma}\kappa^{\tau}z_{ij} - \ln\bar{\rho}_{ij}(n_{ij}) - \delta_i^{\rho} - \zeta_j^{\epsilon} \end{aligned} \tag{71}$$

where

$$v_{ij}^{\epsilon} \equiv -\kappa^{\epsilon}\left(\tilde{\sigma}\eta_{ij}^{\tau} + \eta_{ij}^{f}\right) \quad \text{and} \quad v_{ij}^{\rho} \equiv -\tilde{\sigma}\eta_{ij}^{\tau},$$

$$\delta_i^{\epsilon} \equiv -\kappa^{\epsilon}\left[\ln\sigma w_i\left(\frac{\sigma}{\sigma-1}\frac{w_i}{\bar{a}_i}\right)^{\sigma-1} + \tilde{\sigma}\delta_i^{\tau} + \delta_i^{f}\right] \quad \text{and} \quad \zeta_j^{\epsilon} \equiv \kappa^{\epsilon}\left[\ln\left(P_j^{\sigma-1}E_j\right) - \tilde{\sigma}\zeta_j^{\tau}\right] - \kappa^{\epsilon}\zeta_j^{f},$$

$$\delta_i^{\rho} \equiv \ln\left(\frac{\sigma}{\sigma-1}\frac{w_i}{\bar{a}_i}\right)^{1-\sigma} - \tilde{\sigma}\delta_j^{\tau} \quad \text{and} \quad \zeta_j^{\rho} \equiv \ln\left(P_j^{\sigma-1}E_j\right) - \tilde{\sigma}\zeta_j^{\tau}.$$

We obtain expression (36) by plugging the functional form assumptions in (34) into (71). Finally, notice that the definitions of ζ_j^{ϵ} and ζ_j^{ρ} above immediately imply that

$$\kappa^{\epsilon}\zeta_j^{f} = \kappa^{\epsilon}\zeta_j^{\rho} - \zeta_j^{\epsilon}.$$

B Extensions

This appendix presents three extensions of the equilibrium characterization of the baseline framework in Section 2.

B.1 Multi-Sector, Multi-Factor Heterogeneous Firm Model with Input-Output Links

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

B.1.1 Environment

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

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

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

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

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

where the consumption price index is

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

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 input good 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 $\sigma_j^s > 1$ 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, the price is the CES price index of intermediate inputs:

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

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

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

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

$$\bar{p}_i^s = \left[\alpha_i^s \left(W_i^s\right)^{1-\mu_i^s} + \left(1-\alpha_i^s\right)\left(V_i^s\right)^{1-\mu_i^s}\right]^{\frac{1}{1-\mu_i^s}}, \tag{75}$$

where

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

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

$$l_i^{s,f} = \beta_i^{s,f}\left(\frac{w_i^f}{W_i^s}\right)^{1-\eta_i^s} \alpha_i^s \left(\frac{W_i^s}{\bar{p}_i^s}\right)^{1-\mu_i^s} \quad \text{and} \quad m_i^{ks} = \theta_i^{ks}\left(\frac{P_i^k}{V_i^s}\right)^{1-\kappa_i^s}\left(1-\alpha_i^s\right)\left(\frac{V_i^s}{\bar{p}_i^s}\right)^{1-\mu_i^s}. \tag{77}$$

Production of traded intermediate varieties ω. Assume that sector s has a continuum of monopolitic firms that produce output using only a non-tradable input q_i^s. In order to sell q in market j, variety ω of country i faces a cost function given by

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

where $\bar{p}_i^s$ is the price of the non-tradable input q_i^s in country i.

Given this production technology, the optional price is $p_{ij}^s(\omega) = \frac{\sigma_j^s}{\sigma_j^s-1}\frac{\tau_{ij}^s(\omega)}{a_i^s(\omega)}\frac{\bar{\tau}_{ij}^s}{\bar{a}_i^s}\bar{p}_i^s$ and the associated revenue is

$$R_{ij}^s(\omega) = r_{ij}^s(\omega)\,\bar{r}_{ij}^s\left[\left(\frac{\bar{p}_i^s}{P_j^s}\right)^{1-\sigma^s} E_j^s\right] \tag{78}$$

where

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

Firm ω of country i chooses to enter a foreign market j if, and only if, $\pi_{ij}^s(\omega) = (1/\sigma_j^s)R_{ij}^s(\omega) -$

$\bar{p}_i^s \bar{f}_{ij}^s f_{ij}^s(\omega) \geq 0$. This condition determines the set of firms from country i that operate in sector s of country j:

$$\omega \in \Omega_{ij}^s \quad \Leftrightarrow e_{ij}^s(\omega) \geq \sigma^s \frac{\bar{f}_{ij}^s}{\bar{r}_{ij}^s} \left[\left(\frac{\bar{p}_i^s}{P_j^s} \right)^{\sigma^s} \frac{P_j^s}{E_j^s} \right], \tag{80}$$

where

$$e_{ij}^s(\omega) \equiv \frac{r_{ij}^s(\omega)}{f_{ij}^s(\omega)}. \tag{81}$$

Entry of traded intermediate varieties ω. 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 this case, they take a draw of the variety characteristics from an arbitrary distribution:

$$v_i(\omega) \equiv \left\{ a_i^s(\omega), b_{ij}^s(\omega), \tau_{ij}^s(\omega), f_{ij}^s(\omega) \right\}_j \sim G_i^s(v). \tag{82}$$

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,

$$\sum_j E\left[\max\left\{\pi_{ij}^s(\omega);\ 0\right\}\right] = \bar{p}_i^s \bar{F}_i^s. \tag{83}$$

Market clearing. We 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

$$E_i^s = c_j^s (w_i L_i + T_i) + \sum_k m_i^{ks} \left(\bar{p}_i^k q_i^k\right). \tag{84}$$

The market clearing conditions for factor f in country i is

$$w_i^f \bar{L}_i^f = \sum_s l_i^{s,f} \left(\bar{p}_i^s q_i^s\right). \tag{85}$$

Since all the revenue of the sectoral intermediate good comes from sales to the firms producing the varieties ω, we have that

$$\frac{\bar{p}_i^s q_i^s}{N_i^s} = \underbrace{\sum_j \left(1 - \frac{1}{\sigma^s}\right) Pr[\omega \in \Omega_{ij}^s] E[R_{ij}^s(\omega) | \omega \in \Omega_{ij}^s]}_{\text{final good production}} + \underbrace{\sum_j \bar{p}_i^s \bar{f}_{ij}^s Pr[\omega \in \Omega_{ij}^s] E[f_{ij}^s(\omega) | \omega \in \Omega_{ij}^s]}_{\text{fixed-cost of entering markets}} + \underbrace{\bar{p}_i^s \bar{F}_i^s}_{\text{fixed cost of entry}}.$$

The free entry condition in (83) implies that

$$\bar{p}_i^s \bar{F}_i^s = \sum_j E\left[\max\left\{\pi_{ij}^s(\omega);\ 0\right\}\right] = \sum_j Pr[\omega \in \Omega_{ij}^s] \left(\frac{1}{\sigma^s} E[R_{ij}^s(\omega) | \omega \in \Omega_{ij}^s] - \bar{p}_i^s \bar{f}_{ij}^s E[f_{ij}^s(\omega) | \omega \in \Omega_{ij}^s] \right).$$

Thus, $\bar{p}_i^s q_i^s = \sum_j N_i^s Pr[\omega \in \Omega_{ij}^s] E[R_{ij}^s(\omega) | \omega \in \Omega_{ij}^s]$ and, therefore,

$$\bar{p}_i^s q_i^s = \sum_j \bar{r}_{ij}^s \left[\left(\frac{\bar{p}_i^s}{P_j^s} \right)^{1-\sigma^s} E_j^s \right] \left[\int_{\omega \in \Omega_{ij}^s} r_{ij}^s(\omega) \, d\omega \right]. \tag{86}$$

Equilibrium. Given the distribution in (82), the equilibrium is P_i, $\{\Omega_{ij}^s\}_{j,s}$, $\{P_i^s, N_i^s, \bar{p}_i^s, W_i^s, V_i^s \bar{p}_i^s q_i^s, E_i^s, c_i^s\}_s$, $\{m_i^{sk}\}_{k,s}$, $\{l_i^{s,f}\}_{f,s}$ and $\{w_i^f\}_f$ for all i that satisfy equations (73), (72), (74), (80), (83), (75), (76), (77), (84), (86).

B.1.2 Extensive and Intensive margin of Firm-level Export

We now turn to the characterization of the bilateral levels of entry and sales in each sector. As before, we consider the marginal distribution of $\left(r_{ij}^s(\omega), e_{ij}^s(\omega)\right)$ implied by G_i^s, which can be decomposed without loss of generality as

$$r_{ij}^s(\omega) \sim H_{ij}^{r,s}\left(r|e\right), \qquad \text{and} \qquad e_{ij}^s(\omega) \sim H_{ij}^{e,s}(e), \tag{87}$$

where $H_{ij}^{e,s}$ 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 $\bar{\epsilon}_{ij}^s(n) \equiv \left(H_{ij}^{e,s}\right)^{-1}(1-n)$ such that

$$\ln \bar{\epsilon}_{ij}^s(n_{ij}^s) = \ln\left(\sigma^s \bar{f}_{ij}^s / \bar{r}_{ij}^s\right) + \ln\left(\bar{p}_i^s\right)^{\sigma^s} - \ln E_j^s \left(P_j^s\right)^{\sigma^s - 1}. \tag{88}$$

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 $\bar{x}_{ij}^s \equiv E\left[R_{ij}^s(\omega)\,|\omega \in \Omega_{ij}^s\right]$. Define the average revenue potential of exporters when $n_{ij}^s\%$ of $i's$ firms in sector s export to j as $\bar{\rho}_{ij}^s\left(n_{ij}^s\right) \equiv \frac{1}{n_{ij}^s}\int_0^{n_{ij}^s} \rho_{ij}(n)\, dn$ where $\rho_{ij}^s(n) \equiv E[r|e = \bar{\epsilon}_{ij}^s(n)]$ is the average revenue potential in quantile n of the entry potential distribution. Using the transformation $n = 1 - H_{ij}^{e,s}(e)$ such that $e = \bar{\epsilon}_{ij}^s(n)$ and $dH_{ij}^{e,s}(e) = -dn$, we can follow the same steps as in the baseline model to show that

$$\ln \bar{x}_{ij}^s - \ln \bar{\rho}_{ij}^s(n_{ij}^s) = \ln\left(\bar{r}_{ij}^s\right) + \ln\left(\bar{p}_i^s\right)^{1-\sigma^s} + \ln E_j^s \left(P_j^s\right)^{\sigma^s - 1}. \tag{89}$$

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

B.1.3 General Equilibrium

We now write the equilibrium conditions in terms of $\bar{\rho}_{ij}^s(n)$ and $\bar{\epsilon}_{ij}^s(n)$. We start by writing the price index P_j^s in (74) in terms of $\bar{\rho}_{ij}^s(n)$. Using the expression for $p_{ij}^s(\omega)$ and (74), we have that $(P_j^s)^{1-\sigma^s} = \sum_i \bar{r}_{ij}^s \left(\bar{p}_i^s\right)^{1-\sigma^s} \int_{\Omega_{ij}^s} r_{ij}^s(\omega)\; d\omega$. Since $\int_{\Omega_{ij}^s} r_{ij}^s(\omega)\; d\omega = N_i^s Pr[\omega \in \Omega_{ij}^s] E[r|\omega \in \Omega_{ij}^s] = N_i^s n_{ij}^s \bar{\rho}_{ij}^s(n_{ij}^s)$, we can write P_j^s as

$$(P_j^s)^{1-\sigma^s} = \sum_i \bar{r}_{ij}^s \left(\bar{p}_i^s\right)^{1-\sigma^s} \bar{\rho}_{ij}^s(n_{ij}^s) n_{ij}^s N_i^s. \tag{90}$$

We then turn to the free entry condition in (83). Following the same steps as in Appendix A.1, it is straight forward to show that

$$\mathbb{E}\left[\max\left\{\pi_{ij}^s(\omega);\ 0\right\}\right] = \frac{1}{\sigma^s} n_{ij}^s \bar{x}_{ij}^s - \bar{p}_i^s \bar{f}_{ij}^s \int_0^{n_{ij}} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n)}\ dn,$$

which implies that the free entry condition in (83) is equivalent to

$$\sigma^s \bar{p}_i^s F_i^s = \sum_j n_{ij}^s \bar{x}_{ij}^s - \sum_j \left(\sigma^s \bar{p}_i^s \bar{f}_{ij}^s\right) \int_0^{n_{ij}} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n)}\ dn.$$

Notice that the summation of (88) and (89) implies that $\ln\left(\sigma^s \bar{p}_i^s \bar{f}_{ij}^s\right) = \ln \bar{x}_{ij}^s - \ln \bar{\rho}_{ij}^s(n_{ij}^s) + \ln \bar{\epsilon}_{ij}^s(n_{ij}^s)$. Thus, following again the same steps as in Appendix A.1, it is straight forward to show that

$$\sigma^s \bar{p}_i^s \bar{F}_i^s = \sum_j n_{ij}^s \bar{x}_{ij}^s \left(1 - \frac{\int_0^{n_{ij}^s} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n)}\ dn}{\int_0^{n_{ij}^s} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n_{ij}^s)}\ dn}\right). \tag{91}$$

Finally, we established above that $\bar{p}_i^s q_i^s = \sum_j N_i^s Pr[\omega \in \Omega_{ij}^s] E[R_{ij}^s(\omega) | \omega \in \Omega_{ij}^s]$. Since $n_{ij}^s \equiv Pr[\omega \in \Omega_{ij}^s]$ and $\bar{x}_{ij}^s \equiv E[R_{ij}^s(\omega) | \omega \in \Omega_{ij}^s]$, then

$$\bar{p}_i^s q_i^s = \sum_j N_i^s n_{ij}^s \bar{x}_{ij}^s. \tag{92}$$

The following proposition summarizes the conditions determining all aggregate variables in general equilibrium.

Proposition 3. *Given* $\left\{\{L_i^f\}_f, \{F_i^s, \alpha_i^s, \gamma_i^s, \eta_i^s, \mu_i^s, \kappa_i^s\}_s, \{\bar{r}_{ij}^s, \bar{f}_{ij}^s\}_{j,s}, \{\beta_i^{s,f}\}_{f,s}, \{\theta_i^{sk}\}_{k,s}, \lambda_i\right\}_i$*, an equilibrium vector* $\left\{P_i, \{n_{ij}^s, \bar{x}_{ij}^s\}_{j,s}, \{P_i^s, N_i^s, \bar{p}_i^s, W_i^s, V_i^s \bar{p}_i^s q_i^s, E_i^s, c_i^s\}_s, \{m_i^{sk}\}_{k,s}, \{l_i^{s,f}\}_{f,s}, \{w_i^f\}_f\right\}_i$ *satisfies the following conditions.*

1. The extensive and intensive margins of firm-level sales, n_{ij}^s *and* $\bar{x}_{ij}^s$*, satisfy (88) and (89) for all* s*,* i *and* j*.*

2. The price of the final sectoral good P_j^s *is given by (90). The final consumption good price* P_i *is given by (73).*

3. The number of entrants in sector s *of country* i N_i^s *satisfies the free entry condition in (91).*

4. The price of the the intermediate sector good $\bar{p}_i^s$ *is given by (75) where* W_i^s *and* V_i^s *are given by (76).*

5. The total revenue of the intermediate sectoral good $\bar{p}_i^s q_i^s$ *is given by (92).*

6. Spending on the final sectoral good E_i^s *is (84) with final consumption spending share* c_i^s *given by (72) and the intermediate consumption spending share* m_i^{sk} *given by (77).*

7. Factor price w_i^s *implies that the factor market clearing in (85) holds with* l_i^f *given by (77).*

B.1.4 Nonparametric Counterfactual Predictions

We now use the equilibrium characterization above to compute counterfactual changes in aggregate outcomes using the functions $\bar{\epsilon}_{ij}^s(n)$ and $\bar{\rho}_{ij}^s(n)$. As in our baseline model, this implies that we do not need any parametric restrictions in the distribution of firm heterogeneity G_i. The implications of firm heterogeneity for the model's aggregate counterfactual predictions are summarized by $\bar{\epsilon}_{ij}^s(n)$ and $\bar{\rho}_{ij}^s(n)$.

The extensive and intensive margins of firm-level sales in (88) and (89) imply that

$$\ln \frac{\bar{\epsilon}_{ij}^s(n_{ij}^s \hat{n}_{ij}^s)}{\bar{\epsilon}_{ij}^s(n_{ij}^s)} = \ln\left(\hat{\bar{f}}_{ij}^s / \hat{\bar{r}}_{ij}^s\right) + \ln\left(\hat{\bar{p}}_i^s\right)^{\sigma^s} - \ln \hat{E}_j^s \left(\hat{P}_j^s\right)^{\sigma^s - 1}. \tag{93}$$

$$\ln \hat{\bar{x}}_{ij}^s - \ln \frac{\bar{\rho}_{ij}^s(n_{ij}^s \hat{n}_{ij}^s)}{\bar{\rho}_{ij}^s(n_{ij}^s)} = \ln\left(\hat{\bar{r}}_{ij}^s\right) + \ln\left(\hat{\bar{p}}_i^s\right)^{1-\sigma^s} + \ln \hat{E}_j^s \left(\hat{P}_j^s\right)^{\sigma^s - 1}. \tag{94}$$

The price of the final sectoral good P_j^s in (90) implies that

$$(\hat{P}_j^s)^{1-\sigma^s} = \sum_i \frac{\bar{x}_{ij}^s n_{ij}^s N_i^s}{E_j^s} \left(\hat{\bar{r}}_{ij}^s \left(\hat{\bar{p}}_i^s\right)^{1-\sigma^s} \frac{\bar{\rho}_{ij}^s(n_{ij}^s \hat{n}_{ij}^s)}{\bar{\rho}_{ij}^s(n_{ij}^s)} \hat{n}_{ij}^s \hat{N}_i^s \right).$$

Let $x_{ij}^s \equiv \bar{x}_{ij}^s n_{ij}^s N_i^s / E_j^s = X_{ij}^s / \left(\sum_o X_{oj}^s\right)$ be the spending share of country j on country i. Thus,

$$(\hat{P}_j^s)^{1-\sigma^s} = \sum_i x_{ij}^s \left(\hat{\bar{r}}_{ij}^s \left(\hat{\bar{p}}_i^s\right)^{1-\sigma^s} \frac{\bar{\rho}_{ij}^s(n_{ij}^s \hat{n}_{ij}^s)}{\bar{\rho}_{ij}^s(n_{ij}^s)} \hat{n}_{ij}^s \hat{N}_i^s \right). \tag{95}$$

The final consumption good price P_i in (73) implies that

$$\hat{P}_j^{1-\lambda_j} = \sum_k c_j^s \left(\hat{P}_j^k\right)^{1-\lambda_j}. \tag{96}$$

The free entry condition in (91) implies that

$$\hat{\bar{p}}_i^s \hat{\bar{\bar{F}}}_i^s \sum_j n_{ij}^s \bar{x}_{ij}^s \left(1 - \frac{\int_0^{n_{ij}^s} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n_{ij})}\, dn} \right) = \sum_j n_{ij}^s \bar{x}_{ij}^s (\hat{n}_{ij}^s \hat{\bar{x}}_{ij}^s) \left(1 - \frac{\int_0^{n_{ij}^s \hat{n}_{ij}^s} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n)}\, dn}{\int_0^{n_{ij}^s \hat{n}_{ij}^s} \frac{\rho_{ij}^s(n)}{\bar{\epsilon}_{ij}^s(n_{ij}^s \hat{n}_{ij}^s)}\, dn} \right). \tag{97}$$

The price of the the intermediate sector good $\bar{p}_i^s$ in (75) implies that

$$\hat{\bar{p}}_i^s = \left[\tilde{\alpha}_i^s \left(\hat{W}_i^s\right)^{1-\mu_i^s} + (1 - \tilde{\alpha}_i^s) \left(\hat{V}_i^s\right)^{1-\mu_i^s} \right]^{\frac{1}{1-\mu_i^s}}, \tag{98}$$

where $\tilde{\alpha}_i^s$ is the share of labor in total cost of sector s in country i.

From (76), $\hat{W}_i^s$ and $\hat{V}_i^s$ are given by

$$\hat{W}_i^s = \left[\sum_f l_i^{s,f} \left(\hat{w}_i^f\right)^{1-\eta_i^s} \right]^{\frac{1}{1-\eta_i^s}} \quad \text{and} \quad \hat{V}_i^s = \left[\sum_k m_i^{ks} \left(\hat{P}_i^k\right)^{1-\kappa_i^s} \right]^{\frac{1}{1-\kappa_i^s}}. \tag{99}$$

The total revenue of the intermediate sectoral good $\bar{p}_i^s q_i^s$ in (92) implies

$$\widehat{\bar{p}_i^s q_i^s} = \sum_j x_{ij}^s \hat{N}_i^s \hat{n}_{ij}^s \hat{\bar{x}}_{ij}^s.$$

Let $\iota_i \equiv w_i L_i / (w_i L_i + T_i)$. Spending on the final sectoral good E_i^s in (84) implies that

$$\hat{E}_i^s = \hat{c}_j^s \frac{w_i L_i + T_i}{E_j} \left(\iota_i \hat{w}_i + (1 - \iota_i) \hat{T}_i \right) + \sum_k \frac{\tilde{M}_i^{ks}}{E_i^s} \widehat{\bar{p}_i^s q_i^s}.$$

where $\tilde{M}_i^{ks}$ is the value of intermediate sales of sector k to s in country i.

The final consumption spending share c_i^s in (72) implies that

$$\hat{c}_j^s = \left(\frac{\hat{P}_j^s}{\hat{P}_j}\right)^{1-\lambda_j}. \tag{100}$$

The intermediate consumption spending share m_i^{sk} in (77) implies that

$$\hat{m}_i^{ks} = \left(\frac{\hat{P}_i^k}{\hat{V}_i^s}\right)^{1-\kappa_i^s}\left(\frac{\hat{V}_i^s}{\hat{\bar{p}}_i^s}\right)^{1-\mu_i^s}. \tag{101}$$

The labor spending share l_i^f in (77).

$$\hat{l}_i^{s,f} = \left(\frac{\hat{w}_i^f}{\hat{W}_i^s}\right)^{1-\eta_i^s}\left(\frac{\hat{W}_i^s}{\hat{\bar{p}}_i^s}\right)^{1-\mu_i^s}. \tag{102}$$

The factor market clearing in (85) implies that

$$\hat{w}_i^f = \sum_s \zeta_i^{f,s}\hat{l}_i^{s,f}\left(\widehat{\bar{p}_i^s q_i^s}\right) \tag{103}$$

where $\zeta_i^{f,s}$ is the share of factor f income coming from sector s in country i.

This system determines counterfactual predictions in the model. The following proposition summarizes this.

Proposition 4. *Consider any change in $\hat{\bar{\tau}}_{ij}$ for $i \neq j$. For every i, the changes in $\hat{P}_i$, $\{\hat{n}_{ij}^s, \hat{\bar{x}}_{ij}^s\}_{j,s}$, $\{\hat{P}_i^s, \hat{N}_i^s, \hat{\bar{p}}_i^s, \hat{W}_i^s, \hat{V}_i^s \widehat{\bar{p}_i^s q_i^s}, \hat{E}_i^s, \hat{c}_i^s\}_s$, $\{\hat{m}_i^{sk}\}_{k,s}$, $\{\hat{l}_i^{s,f}\}_{f,s}$ and $\{\hat{w}_i^f\}_f$ must satisfy the system in (93)–(102).*

B.2 Allowing for zero bilateral trade

In this section, we extend our baseline framework to allow zero trade flows between two countries.

B.2.1 Environment

Consider the same environment described in Section 2.1.

B.2.2 Extensive and Intensive Margin of Firm Export

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

$$r_{ij}(\omega) \sim H_{ij}^r(r|e), \qquad \text{and} \qquad e_{ij}(\omega) \sim H_{ij}^e(e). \tag{104}$$

To allow for zero trade flows, we follow Helpman et al. (2008) by allows the support of the entry potential distribution to be bounded. Specifically, assume that $H_{ij}(e)$ has full support over $[0, \bar{e}_{ij}]$.

Extensive margin of firm-level exports. Recall that $n_{ij} \equiv Pr[\omega \in \Omega_{ij}]$ where Ω_{ij} is given by (5). It implies that

$$n_{ij} = \begin{cases} 1 - H_{ij}^e(e_{ij}^*) & \text{if} \quad e_{ij}^* \leq \bar{e}_{ij} \\ 0 & \text{if} \quad e_{ij}^* > \bar{e}_{ij} \end{cases} \quad \text{where} \quad e_{ij}^* \equiv \sigma\frac{\bar{f}_{ij}}{\bar{r}_{ij}}\left[\left(\frac{w_i}{P_j}\right)^{\sigma}\frac{P_j}{E_j}\right].$$

Let us now define the extensive margin function as

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

Using this definition and the expression for n_{ij} above, we get that

$$\epsilon_{ij}(n_{ij}) = \min\left\{e^*_{ij}, \bar{e}_{ij}\right\}.$$

We then define $\bar{\epsilon}_{ij}(n) \equiv \epsilon_{ij}(n_{ij})/\bar{e}_{ij}$ and $\tilde{f}_{ij} \equiv \bar{f}_{ij}/\bar{e}_{ij}$. Then,

$$\ln\bar{\epsilon}_{ij}(n_{ij}) = \min\left\{\ln\left(\sigma\tilde{f}_{ij}\bar{r}_{ij}\right) + \ln\left(w_i^{\sigma}\right) - \ln\left(E_j P_j^{\sigma-1}\right), 0\right\}. \tag{105}$$

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

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

We consider the transformation $n = 1 - H^e_{ij}(e)$ such that $e = \epsilon_{ij}(n)$ and $dH_{ij}(e) = -dn$. By defining $\rho_{ij}(n) \equiv E[r|e = \epsilon_{ij}(n)]$ and $\bar{\rho}_{ij}(0) = 0$,

$$\ln\bar{x}_{ij} - \ln\bar{\rho}_{ij}(n_{ij}) = \ln\left(\bar{r}_{ij}\right) + \ln\left(w_i^{1-\sigma}\right) + \ln\left(E_j P_j^{\sigma-1}\right). \tag{106}$$

B.2.3 General Equilibrium

We now write the equilibrium conditions in terms of $\bar{\rho}_{ij}(.)$ and $\bar{\epsilon}_{ij}(.)$. Since $x_{ij} = \bar{x}_{ij} n_{ij} N_i / E_j$ and $\sum_i x_{ij} = 1$, the expression above immediately implies that

$$P_j^{1-\sigma} = \sum_{i:n_{ij}>0} \bar{r}_{ij}\left(w_i\right)^{1-\sigma}\bar{\rho}_{ij}(n_{ij})\left(n_{ij}N_i\right) \tag{107}$$

We then turn to the free entry condition in (8). Following the same steps as in Appendix A.1, it is straight forward to show that

$$\sigma w_i F_i = \sum_j n_{ij}\bar{x}_{ij} - \sum_j\left(\sigma w_i \bar{f}_{ij}\right)\int_0^{n_{ij}}\frac{\rho_{ij}(n)}{\epsilon_{ij}(n)}\; dn.$$

For $n_{ij} > 0$, the ratio of (105) and (106) implies that $\sigma\bar{f}_{ij}w_i = \bar{x}_{ij}\epsilon_{ij}(n_{ij})/\bar{\rho}_{ij}(n_{ij})$. Thus,

$$\sigma w_i F_i = \sum_{j:n_{ij}>0} n_{ij}\bar{x}_{ij}\left(1 - \frac{\int_0^{n_{ij}}\frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\; dn}{\int_0^{n_{ij}}\frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\; dn}\right). \tag{108}$$

Following the same steps as in Appendix A.1, it is straight forward to show that

$$w_i L_i = \sum_{j:n_{ij}>0} N_i n_{ij}\bar{x}_{ij} \tag{109}$$

The following proposition summarizes the conditions determining all aggregate variables in general equilibrium.

Proposition 5. *Given* $\left\{\bar{L}_i, \bar{F}_i, \{\bar{r}_{ij}, \tilde{f}_{ij}\}_j\right\}_i$*, an equilibrium vector* $\{\{n_{ij}, \bar{x}_{ij}\}_j, P_i, N_i, E_i, w_i\}_i$ *satisfies the following conditions.*

1. The extensive and intensive margins of firm-level sales, n_{ij} *and* $\bar{x}_{ij}$*, satisfy (105) and (106) for all* i *and* j*.*

2. For all i*, the price index is given by (107).*

3. For all i*, free entry is given by (108).*

4. For all i*, total spending,* E_i*, satisfies (9).*

5. For all i*, the labor market clearing is given by (109).*

B.2.4 Nonparametric Counterfactual Predictions

We now use the equilibrium characterization above to compute counterfactual changes in aggregate outcomes using the functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$. We further assume that, in bilateral pairs for which initially $n_{ij} = 0$, we still have that $n'_{ij} = 0$. Thus, (105) implies that

$$\frac{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\epsilon}_{ij}(n_{ij})} = \frac{\hat{\tilde{f}}_{ij}}{\hat{\bar{r}}_{ij}} \left[\left(\frac{\hat{w}_i}{\hat{P}_j} \right)^{\sigma} \frac{\hat{P}_j}{\hat{E}_j} \right] \quad \text{for} \quad n_{ij} > 0 \tag{110}$$

$$n'_{ij} = 0 \quad \text{for} \quad n_{ij} = 0. \tag{111}$$

The intensive margin equation remains the same:

$$\hat{\bar{x}}_{ij} = \hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \left[\left(\frac{\hat{w}_i}{\hat{P}_j} \right)^{1-\sigma} \hat{E}_j \right] \quad \text{for} \quad n_{ij} > 0. \tag{112}$$

The price index equation in (107) implies that

$$\hat{P}_j^{1-\sigma} = \sum_{i:n_{ij}>0} x_{ij} \left(\hat{\bar{r}}_{ij} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{w}_i^{1-\sigma} \hat{n}_{ij} \hat{N}_i \right). \tag{113}$$

The spending equation in (9) implies that

$$\hat{E}_i = \iota_i \left(\hat{w}_i \hat{\bar{L}}_i \right) + (1 - \iota_i) \hat{\bar{T}}_i, \tag{114}$$

The labor market clearing condition in (109) implies

$$\hat{w}_i \hat{\bar{L}}_i = \sum_{j:n_{ij}>0} y_{ij} \left(\hat{N}_i \hat{n}_{ij} \hat{\bar{x}}_{ij} \right). \tag{115}$$

The free entry condition in (108) implies that

$$\hat{w}_i \sum_{j:n_{ij}>0} n_{ij}\bar{x}_{ij} \left(1 - \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\, dn} \right) = \sum_{j:n_{ij}>0} n_{ij}\bar{x}_{ij}(\hat{n}_{ij}\hat{\bar{x}}_{ij}) \left(1 - \frac{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}\, dn} \right). \tag{116}$$

This system determines counterfactual predictions in the model. Notice that, as in our baseline model, it only depends on data in the initial equilibrium, the elasticity of substitution σ, and the two elasticity functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ in the gravity equations (105) and (106). The following proposition summarizes the result in this section.

Proposition 6. *Consider any change in $\hat{\bar{\tau}}_{ij}$ for $i \neq j$. Assume that, if $n_{ij} = 0$, then $n'_{ij} = 0$. The change in aggregate outcomes $\{\hat{\bar{\boldsymbol{n}}}, \hat{\bar{\boldsymbol{X}}}, \hat{\boldsymbol{P}}, \hat{\boldsymbol{N}}, \hat{\boldsymbol{E}}, \hat{\boldsymbol{w}}\}$ must satisfy the system in (110)–(116).*

B.3 Model with Import Tariffs

In this section, we follow Costinot and Rodriguez-Clare (2013) to extend our baseline framework to allow for import tariffs.

B.3.1 Environment

We assume that country j charges an ad-valorem tariff of t_{ij} such that the total trade costs between country i and j is $\bar{\tau}_{ij}(1+t_{ij})$. We consider a monopolistic competitive environment in which firms maximize profits given the demand in (1). For firm ω of country i, the optimal price in market j is $p_{ij}(\omega) = \frac{\sigma}{\sigma-1}\frac{\bar{\tau}_{ij}(1+t_{ij})w_i}{\bar{a}_i}\frac{\tau_{ij}(\omega)}{a_i(\omega)}$ with an associated revenue of

$$R_{ij}(\omega) = \bar{r}_{ij} r_{ij}(\omega) \left[\left(\frac{w_i}{P_j} \right)^{1-\sigma} E_j \right], \tag{117}$$

where

$$r_{ij}(\omega) \equiv b_{ij}(\omega) \left(\frac{\tau_{ij}(\omega)}{a_i(\omega)} \right)^{1-\sigma} \quad \text{and} \quad \bar{r}_{ij} \equiv \bar{b}_{ij} \left(\frac{\sigma}{\sigma-1} \frac{\bar{\tau}_{ij}(1+t_{ij})}{\bar{a}_i} \right)^{1-\sigma}. \tag{118}$$

The firm's entry decision depends on the profit generated by the revenue in (117), $\sigma^{-1}(1+t_{ij})^{-1}R_{ij}(\omega)$, and the fixed-cost of entry, $w_i \bar{f}_{ij} f_{ij}(\omega)$. Specifically, firm ω of i enters j if, and only if, $\pi_{ij}(\omega) = \sigma^{-1}(1+t_{ij})^{-1}R_{ij}(\omega) - w_i \bar{f}_{ij} f_{ij}(\omega) \geq 0$. This yields the set of firms from i selling in j:

$$\omega \in \Omega_{ij} \quad \Leftrightarrow e_{ij}(\omega) \geq \sigma(1+t_{ij}) \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right], \tag{119}$$

where

$$e_{ij}(\omega) \equiv \frac{r_{ij}(\omega)}{f_{ij}(\omega)}. \tag{120}$$

The aggregate trade flows (including tariff) is still given by

$$X_{ij} = \int_{\omega \in \Omega_{ij}} R_{ij}(\omega)\, d\omega. \tag{121}$$

As before, free entry implies that N_i satisfies

$$\sum_j E\left[\max\{\pi_{ij}(\omega);\ 0\}\right] = w_i \bar{F}_i. \tag{122}$$

Market clearing. As shown by Costinot and Rodriguez-Clare (2013), the country's spending now also includes the tariff revenue:

$$E_i = w_i \bar{L}_i + \bar{T}_i + \sum_i \frac{t_{ij}}{1+t_{ij}} X_{ij}. \tag{123}$$

Now a fraction $t_{ij}/(1+t_{ij})$ of total revenue goes to the government of country j. So, labor in country i only receive a fraction $1/(1+t_{ij})$ of the sales revenue. Thus, $w_i L_i = (1+t_{ij})^{-1} \int_{\omega \in \Omega_{ij}} R_{ij}(\omega)\, d\omega$ and, by (118),

$$w_i \bar{L}_i = \frac{\bar{r}_{ij}}{1+t_{ij}} \left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j \left[\int_{\omega \in \Omega_{ij}} r_{ij}(\omega)\, d\omega\right]. \tag{124}$$

B.3.2 Extensive and Intensive Margin of Firm Export

Using the same definitions of the baseline model, expression (119) yields

$$\ln \bar{\epsilon}_{ij}(n_{ij}) = \ln\left(\sigma(1+t_{ij})\bar{f}_{ij}/\bar{r}_{ij}\right) + \ln\left(w_i^{\sigma}\right) - \ln\left(E_j P_j^{\sigma-1}\right). \tag{125}$$

Again, following the same steps of the baseline model, equation (117) implies the same intensive margin equation:

$$\ln \bar{x}_{ij} - \ln \bar{\rho}_{ij}(n_{ij}) = \ln\left(\bar{r}_{ij}\right) + \ln\left(w_i^{1-\sigma}\right) + \ln\left(E_j P_j^{\sigma-1}\right). \tag{126}$$

B.3.3 General Equilibrium

Part 1. To derive the labor market clearing condition notice that there are three sources of demand for labor: production of goods, fixed-cost of entering a market and fixed-cost of creating a variety. Thus,

$$w_i \bar{L}_i = \sum_j N_i Pr[\omega \in \Omega_{ij}] \left(1 - \frac{1}{\sigma}\right) \frac{1}{1+t_{ij}} E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] + \sum_j N_i Pr[\omega \in \Omega_{ij}] w_i \bar{f}_{ij} E\left[f_{ij}(\omega) | \omega \in \Omega_{ij}\right] + N_i w_i \bar{F}_i$$

From the free entry condition, we know that

$$w_i \bar{F}_i = \sum_j \mathbb{E}\left[\max\{\pi_{ij}(\omega);\ 0\}\right] = \sum_j Pr[\omega \in \Omega_{ij}] \left(\frac{1}{\sigma} \frac{1}{1+t_{ij}} E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] - w_i \bar{f}_{ij} E\left[f_{ij}(\omega) | \omega \in \Omega_{ij}\right]\right),$$

which implies that

$$w_i \bar{L}_i = \sum_j N_i Pr[\omega \in \Omega_{ij}] E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] \frac{1}{1+t_{ij}}.$$

Thus, since $\bar{x}_{ij} \equiv E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right]$ and $n_{ij} = Pr[\omega \in \Omega_{ij}]$, this immediately implies that

$$w_i \bar{L}_i = \sum_j \frac{N_i n_{ij} \bar{x}_{ij}}{1+t_{ij}}. \tag{127}$$

Part 2. Since $p_{ij}(\omega) = \frac{\sigma}{\sigma-1} \frac{\bar{\tau}_{ij} w_i}{\bar{a}_i} \frac{\tau_{ij}(\omega)}{a_i(\omega)}$, the expression for $P_j^{1-\sigma}$ in (2) implies that

$$P_j^{1-\sigma} = \sum_i \left[\bar{b}_{ij} \left(\frac{\sigma}{\sigma-1} \frac{(1+t_{ij})\bar{\tau}_{ij}}{\bar{a}_i}\right)^{1-\sigma}\right] \left(w_i^{1-\sigma}\right) \int_{\Omega_{ij}} \left(b_{ij}(\omega)\right) \left(\frac{\tau_{ij}(\omega)}{a_i(\omega)}\right)^{1-\sigma} d\omega$$

Using the definitions in (4), we can write this expression as

$$P_j^{1-\sigma} = \sum_i \bar{r}_{ij} \left(w_i^{1-\sigma}\right) \int_{\Omega_{ij}} r_{ij}(\omega)\ d\omega$$

Notice that $\int_{\Omega_{ij}} r_{ij}(\omega)\ d\omega = N_i Pr[\omega \in \Omega_{ij}] E[r|\omega \in \Omega_{ij}] = N_i n_{ij} \bar{\rho}_{ij}(n_{ij})$. This immediately yields

$$P_j^{1-\sigma} = \sum_i \bar{r}_{ij} w_i^{1-\sigma} \bar{\rho}_{ij}(n_{ij}) n_{ij} N_i. \tag{128}$$

Part 3. We start by writing

$$\begin{aligned}
\mathbb{E}\left[\max\{\pi_{ij}(\omega);\ 0\}\right] &= Pr[\omega \in \Omega_{ij}] E\left[\pi_{ij}(\omega) | \omega \in \Omega_{ij}\right] + Pr[\omega \notin \Omega_{ij}] 0 \\
&= Pr[\omega \in \Omega_{ij}] \left(\tfrac{1}{\sigma} \tfrac{1}{1+t_{ij}} E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right] - w_i \bar{f}_{ij} E\left[f_{ij}(\omega) | \omega \in \Omega_{ij}\right]\right) \\
&= n_{ij} \left(\tfrac{1}{\sigma} \tfrac{1}{1+t_{ij}} \bar{x}_{ij} - w_i \bar{f}_{ij} E\left[r_{ij}(\omega)/e_{ij}(\omega) | \omega \in \Omega_{ij}\right]\right)
\end{aligned}$$

where the second equality follows from the expression for $\pi_{ij}(\omega) = \frac{1}{\sigma} \frac{1}{1+t_{ij}} R_{ij}(\omega) - w_i \bar{f}_{ij} f_{ij}(\omega)$, and the third equality follows from the definitions of $\bar{x}_{ij} \equiv E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right]$ and $e_{ij}(\omega) \equiv r_{ij}(\omega)/f_{ij}(\omega)$.

By defining $e_{ij}^* \equiv \sigma(1+t_{ij}) \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j}\right)^{\sigma} \frac{P_j}{E_j}\right]$, we can write

$$E\left[r_{ij}(\omega)/e_{ij}(\omega) | \omega \in \Omega_{ij}\right] = \int_{e_{ij}^*}^{\infty} \frac{1}{e} \left[\int_0^{\infty} r dH_{ij}^r(r|e)\right] \frac{dH^e(e)}{1 - H^e(e_{ij}^*)}$$

Consider the transformation $n = 1 - H_{ij}(e)$ such that $e = \bar{\epsilon}_{ij}(n)$. In this case, $dH_{ij}(e) = -dn$ and $n_{ij} = 1 - H_{ij}(e_{ij}^*)$, which implies that

$$E\left[r_{ij}(\omega)/e_{ij}(\omega) | \omega \in \Omega_{ij}\right] = \frac{1}{n_{ij}} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} dn.$$

Thus,

$$\mathbb{E}\left[\max\{\pi_{ij}(\omega);\ 0\}\right] = \frac{1}{\sigma} \frac{1}{1+t_{ij}} n_{ij} \bar{x}_{ij} - w_i \bar{f}_{ij} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn.$$

Thus, the free entry condition is

$$\sigma w_i \bar{F}_i = \sum_j \frac{n_{ij} \bar{x}_{ij}}{1+t_{ij}} - \sum_j \left(\sigma w_i \bar{f}_{ij}\right) \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn. \tag{129}$$

Notice that the summation of (125) and (126) implies that

$$\ln\left(\sigma(1+t_{ij}) w_i \bar{f}_{ij}\right) = \ln \bar{x}_{ij} - \ln \bar{\rho}_{ij}(n_{ij}) + \ln \bar{\epsilon}_{ij}(n_{ij})$$

which yields

$$\sigma w_i \bar{F}_i = \sum_j \frac{n_{ij} \bar{x}_{ij}}{1+t_{ij}} - \sum_j \frac{n_{ij} \bar{x}_{ij}}{1+t_{ij}} \frac{\bar{\epsilon}_{ij}(n_{ij})}{\bar{\rho}_{ij}(n_{ij})} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\ dn.$$

By substituting the definition of $\bar{\rho}_{ij}(n)$, we can write the free entry condition as

$$\sigma w_i \bar{F}_i = \sum_j \frac{n_{ij}\bar{x}_{ij}}{1+t_{ij}} - \sum_j \frac{n_{ij}\bar{x}_{ij}}{1+t_{ij}} \frac{\bar{\epsilon}_{ij}(n_{ij})}{\bar{\rho}_{ij}(n_{ij})} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn.$$

Using the market clearing condition in (19), we have that

$$\frac{1}{N_i} = \sigma \frac{\bar{F}_i}{\bar{L}_i} + \sum_j \frac{n_{ij}\bar{x}_{ij}}{(1+t_{ij})w_i \bar{L}_i} \frac{\bar{\epsilon}_{ij}(n_{ij})}{\bar{\rho}_{ij}(n_{ij})} \int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn. \tag{130}$$

which immediately yields equation (21).

Part 4. Equation (123) implies that

$$E_i = w_i \bar{L}_i + \bar{T}_i + \sum_j \frac{t_{ji}}{1+t_{ji}} (N_j n_{ji} \bar{x}_{ji}). \tag{131}$$

The following proposition summarizes the conditions determining all aggregate variables in general equilibrium.

Proposition 7. *Given* $\left\{\bar{L}_i, \bar{F}_i, \{t_{ij}, \bar{r}_{ij}, \bar{f}_{ij}\}_j\right\}_i$*, an equilibrium vector* $\{\{n_{ij}, \bar{x}_{ij}\}_j, P_i, N_i, E_i, w_i\}_i$ *satisfies the following conditions.*

1. *The extensive and intensive margins of firm-level sales,* n_{ij} *and* $\bar{x}_{ij}$*, satisfy (125) and (126) for all* i *and* j.
2. *For all* i*, the price index is given by (128).*
3. *For all* i*, free entry is given by (130).*
4. *For all* i*, total spending,* E_i*, satisfies (131).*
5. *For all* i*, the labor market clearing is given by (127).*

B.3.4 Nonparametric Counterfactual Predictions

We now use the equilibrium characterization above to compute counterfactual changes in aggregate outcomes using the functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$.

From (125) and (126),

$$\frac{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\epsilon}_{ij}(n_{ij})} = \widehat{(1+t_{ij})} \frac{\hat{\bar{f}}_{ij}}{\hat{\bar{r}}_{ij}} \frac{\hat{w}_i^{\sigma}}{\hat{E}_j \hat{P}_j^{\sigma-1}}. \tag{132}$$

$$\hat{\bar{x}}_{ij} = \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{\bar{r}}_{ij} \frac{\hat{E}_j \hat{P}_j^{\sigma-1}}{\hat{w}_i^{\sigma-1}} \tag{133}$$

From (128),

$$\hat{P}_j^{1-\sigma} = \sum_i x_{ij} \hat{\bar{r}}_{ij} \hat{w}_i^{1-\sigma} \frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})} \hat{n}_{ij} \hat{N}_i. \tag{134}$$

From (130),

$$N_i \hat{N}_i = \left[\sigma \frac{\bar{F}_i}{\bar{L}_i} \frac{\hat{\bar{F}}_i}{\hat{\bar{L}}_i} + \sum_j \frac{n_{ij}\bar{x}_{ij}}{(1+t_{ij})w_i\bar{L}_i} \frac{\hat{n}_{ij}\hat{\bar{x}}_{ij}}{\widehat{(1+t_{ij})}\hat{w}_i \hat{\bar{L}}_i} \frac{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\, dn}{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}\, dn} \right]^{-1}$$

Using (130) to substitute for $\sigma \frac{\bar{F}_i}{\bar{L}_i}$,

$$\hat{N}_i = \left[\left(1 - \sum_j y_{ij} \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})} \, d} \right) \frac{\hat{\bar{F}}_i}{\hat{\bar{L}}_i} + \sum_j y_{ij} \frac{\hat{n}_{ij} \hat{\bar{x}}_{ij}}{\widehat{(1+t_{ij})} \hat{w}_i \hat{\bar{L}}_i} \frac{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)} \, dn}{\int_0^{n_{ij}\hat{n}_{ij}} \frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})} \, dn} \right]^{-1}. \tag{135}$$

where $y_{ij} = \frac{X_{ij}}{(1+t_{ij}) w_i \bar{L}_i}$ is the share of income in i from sales to j.

From (131),

$$\hat{E}_i = \iota_i \hat{w}_i \hat{\bar{L}}_i + \vartheta_i \hat{\bar{T}}_i + \sum_j \left(\frac{t_{ji}}{1+t_{ji}} \frac{\hat{t}_{ji}}{\widehat{(1+t_{ji})}} \right) \frac{X_{ji}}{E_i} (\hat{N}_j \hat{n}_{ji} \hat{\bar{x}}_{ji}). \tag{136}$$

where $\iota_i \equiv Y_i / E_i$ and $\vartheta_i \equiv \bar{T}_i / E_i$.

This system determines counterfactual predictions in the model. Notice that, as in our baseline model, it only depends on data in the initial equilibrium, the elasticity of substitution σ, and the two elasticity functions $\bar{\epsilon}_{ij}(n)$ and $\bar{\rho}_{ij}(n)$ in the gravity equations (125) and (126). The following proposition summarizes the result in this section.

Proposition 8. *Consider any change in $\hat{\bar{\tau}}_{ij}$ for $i \neq j$. The change in aggregate outcomes $\{\hat{\bar{\boldsymbol{n}}}, \hat{\bar{\boldsymbol{X}}}, \hat{\boldsymbol{P}}, \hat{\boldsymbol{N}}, \hat{\boldsymbol{E}}, \hat{\boldsymbol{w}}\}$ must satisfy the system in (132)–(136).*

B.4 Multi-product Firms

In this section, we extend the framework of Arkolakis et al. (2019b) to formulate a model of multi-product firms without parametric assumptions on the distribution of firm fundamentals.[24] The setup adapts the Bernard et al. (2011) framework to a finite products for each firm.

B.4.1 Environment

Demand. 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

$$q_{ij}(\omega) = \left(\bar{b}_{ij} b_{ij}(\omega) \right) \left(\frac{p_{ij}(\omega)}{P_j} \right)^{-\sigma} \frac{E_j}{P_j}, \tag{137}$$

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,

$$P_j^{1-\sigma} = \sum_i \int_{\Omega_{ij}} \left(\bar{b}_{ij} b_{ij}(\omega) \right) \left(p_{ij}(\omega) \right)^{1-\sigma} \, d\omega, \tag{138}$$

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

Production. We consider a monopolistic competitive environment in which firms maximize profits given the demand in (137). We continue to assume that each variety is produced by a single firm, η. But firms can

[24] While the outline follows Arkolakis et al. (2019b), this approach simplifies certain steps for both expositional clarity and brevity. In particular, here we limit the maximal number of products a firm can produce. See Arkolakis et al. (2019b) for conditions on which the constraint can be limited.

now produce an exogenous large number of varieties $\omega \in \Omega(\eta)$. If the firm produces $q_{ij}(\omega) > 0$ units of each variety ω the labor cost is

$$C_{ij}(\omega) = \int_{\omega \in \Omega(\eta)} \left(w_i \frac{\tau_{ij}(\omega)}{a_i(\omega)} \frac{\bar{\tau}_{ij}}{\bar{a}_i} q_{ij}(\omega) + w_i \bar{f}_{ij} f_{ij}(\omega) \mathbb{I}_{ij}(\omega) \right),$$

The terms are the same as in our baseline. They capture the fixed and variable cost of selling variety ω in country j by firm η from country i and $\mathbb{I}_{ij}(\omega)$ is an indicator function that determines which goods are produced in location i to be sold in j in positive quantities, $q_{ij}(\omega) > 0$. In summary, there are four sources of heterogeneity, productivity $a_i(\omega)$, demand appeal $b_{ij}(\omega)$, variable trade costs $\tau_{ij}(\omega)$ and fixed production costs $f_{ij}(\omega)$. There product-specific sources of heterogeneity may be arbitrarily correlated each other and with other products from the same firm.

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

$$R_{ij}(\omega) = \bar{r}_{ij} r_{ij}(\omega) \left[\left(\frac{w_i}{P_j} \right)^{1-\sigma} E_j \right], \tag{139}$$

where

$$r_{ij}(\omega) \equiv b_{ij}(\omega) \left(\frac{\tau_{ij}(\omega)}{a_i(\omega)} \right)^{1-\sigma} \quad \text{and} \quad \bar{r}_{ij} \equiv \bar{b}_{ij} \left(\frac{\sigma}{\sigma-1} \frac{\bar{\tau}_{ij}}{\bar{a}_i} \right)^{1-\sigma}. \tag{140}$$

Firm η of i sells variety ω in j if, and only if profits are positive, $\pi_{ij}(\omega) = \frac{1}{\sigma} R_{ij}(\omega) - w_i \bar{f}_{ij} f_{ij}(\omega) \geq 0$. This yields the set of products from firm η of country i that are sold in j:

$$\omega \in \Omega_{ij}(\eta) \quad \Leftrightarrow e_{ij}(\omega) \geq \sigma \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right] \tag{141}$$

where

$$e_{ij}(\omega) \equiv \frac{r_{ij}(\omega)}{f_{ij}(\omega)}. \tag{142}$$

Following the entry condition in equation (141). Firm revenues are given by

$$\tilde{R}_{ij}(\eta) = \int_{\omega \in \Omega_{ij}(\eta)} R_{ij}(\omega) = \bar{r}_{ij} \tilde{r}_{ij}(\eta) \left[\left(\frac{w_i}{P_j} \right)^{1-\sigma} E_j \right], \tag{143}$$

where

$$\tilde{r}_{ij}(\eta) \equiv \sum_{\omega \in \Omega_{ij}(\eta)} r_{ij}(\omega). \tag{144}$$

The set of entrants $\mathcal{N}_{ij}$ corresponds to the set of firms that sell at least one product. This is determined by the entry of the firm's variety with the maximum entry potential since (141) has the same threshold in the right-hand side for all varieties. Thus,

$$\eta \in \mathcal{N}_{ij} \Leftrightarrow \tilde{e}_{ij} \geq \sigma \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right] \tag{145}$$

where

$$\tilde{e}_{ij}(\eta) \equiv \max_{\omega \in \Omega(\eta)} e_{ij}(\omega).$$

This implies that, if $\eta \notin \mathcal{N}_{ij}$, then $\pi_{ij}(\eta) = 0$. If $\eta \in \mathcal{N}_{ij}$, then

$$\pi_{ij}(\eta) = \frac{1}{\sigma}\tilde{R}_{ij}(\eta) - \int_{\omega\in\Omega_{ij}(\eta)} w_i \bar{f}_{ij} f_{ij}(\omega).$$

The set of varieties from country i sold in country j can be written as $\Omega_{ij} = \cup_\eta \Omega_{ij}(\eta)$.

Entry. An entrant firm pays a fixed labor cost $\bar{F}_i$ to draw its type η. The firm's type determines its set of varieties $\Omega(\eta)$. The firm then draws variety characteristics from an arbitrary distribution for a finite set of products ω:

$$v_i(\omega) \equiv \{a_i(\omega), b_{ij}(\omega), \tau_{ij}(\omega), f_{ij}(\omega)\}_j \sim G_i^v(v|\eta), \tag{146}$$

$$\eta \sim G_i^\eta(\eta). \tag{147}$$

In equilibrium, N_i firms pay the fixed cost of entry in exchange for an ex-ante expected profit of zero. The free entry implies that

$$\sum_j E\left[\max\{\pi_{ij}(\eta);\ 0\}\right] = w_i \bar{F}_i. \tag{148}$$

Market clearing. We follow Dekle et al. (2008) by introducing exogenous international transfers, so that spending is

$$E_i = w_i \bar{L}_i + \bar{T}_i, \quad \sum_i \bar{T}_i = 0. \tag{149}$$

Since labor is the only factor of production, labor income in i equals the total revenue of firms from i: $w_i L_i = \sum_j \int_{\eta\in\mathcal{N}_{ij}} \int_{\omega\in\Omega_{ij}(\eta)} R_{ij}(\omega)\, d\omega d\eta$. Given the expression in (139),

$$w_i \bar{L}_i = \sum_j \bar{r}_{ij} \left(\frac{w_i}{P_j}\right)^{1-\sigma} E_j \left[\int_{\eta\in\mathcal{N}_{ij}} \int_{\omega\in\Omega_{ij}(\eta)} r_{ij}(\omega)\, d\omega d\eta\right]. \tag{150}$$

Equilibrium. Given the arbitrary distribution in (146) and (147), the equilibrium is defined as the vector $\{P_i, \{\Omega_{ij}(\eta), \mathcal{N}_{ij}\}_{ij}, N_i, E_i, w_i\}_i$ satisfying equations (138), (141), (145), (148), (149), (150) for all i.

B.4.2 Extensive and Intensive Margin of Firm Exports

The extensive and intensive margin of firm exports follows the single-product case. We now use the definitions of entry and revenue potentials to characterize firm-level entry and sales in different markets in general equilibrium. We consider the CDF of $(r_{ij}(\omega), e_{ij}(\omega), \tilde{e}_{ij}(\eta))$ implied by $G_i^v(v|\eta)$. We assume that

$$r_{ij}(\omega) \sim H_{ij}^r(r|\tilde{e}), \qquad e_{ij}(\omega) \sim H_{ij}^e(e|\tilde{e}), \quad \text{and} \quad \tilde{e}_{ij}(\eta) \sim H_{ij}^{\tilde{e}}(\tilde{e}). \tag{151}$$

As in our baseline, we assume that $H_{ij}^{\tilde{e}}(e)$ is continuous and strictly increasing in $\mathbb{R}_+$ with $\lim_{e\to\infty} H_{ij}^e(e) = 1$. We also assume that $\bar{H}_{ij}^e(\tilde{e}) \equiv \int_e^\infty H_{ij}^e(e|\tilde{e}) dH_{ij}^{\tilde{e}}(\tilde{e})$ is invertible.

We derive three types of semiparametric gravity equations. We first derive two firm-specific gravity equations determining per-product average sales, $\bar{x}_{ij} \equiv E_\eta\left[E\left[R_{ij}(\omega)|\omega\in\Omega_{ij}(\eta)\right]|\eta\in\mathcal{N}_{ij}\right]$, and product entry across markets, $\bar{n}_{ij} = E_\eta\left[Pr\left[\omega\in\Omega_{ij}(\eta)\right]|\eta\in\mathcal{N}_{ij}\right]$. We then derive a third gravity equation for firm entry across destinations, $n_{ij} = Pr\left[\eta\in\mathcal{N}_{ij}\right]$.

Extensive margin of firm entry. The share of firms of country i entering market j is given by:

$$n_{ij} = 1 - H_{ij}^{\tilde{e}} \left(\frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right] \right).$$

Given the restrictions imposed on H^e, we can proceed as in the baseline to define $\bar{\epsilon}_{ij}(n) \equiv \left(H_{ij}^{e}\right)^{-1}(1-n)$ where $\bar{\epsilon}_{ij}(n)$ is strictly decreasing, $\bar{\epsilon}_{ij}(1) = 0$, and $\lim_{n\to 0} \bar{\epsilon}_{ij}(n) = \infty$. In any equilibrium, we can then re-write the expression above as

$$\ln \bar{\epsilon}_{ij}(n_{ij}) = \ln \left(\sigma \bar{f}_{ij}/\bar{r}_{ij}\right) + \ln \left(w_i^{\sigma}\right) - \ln \left(E_j P_j^{\sigma-1}\right). \tag{152}$$

Extensive margin of number of products for exporting firms. We define $\bar{\mu}_{ij}(n|\tilde{e}) \equiv \left(\bar{H}_{ij}^{e}\right)^{-1}(1-n)$ where $\bar{\mu}_{ij}(n)$ is strictly decreasing, $\bar{\mu}_{ij}(1) = 0$, and $\lim_{n\to 0} \bar{\mu}_{ij}(n) = \infty$. In any equilibrium, expression (141) yields

$$\bar{n}_{ij} = \int_{\frac{\bar{f}_{ij}}{\bar{r}_{ij}}\left[\left(\frac{w_i}{P_j}\right)^{\sigma}\frac{P_j}{E_j}\right]}^{\infty} H^e \left(\frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right] |\tilde{e} \right) dH_{ij}^{\tilde{e}} = \bar{H}_{ij}^{e} \left(\frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right] \right)$$

Thus,

$$\ln \bar{\mu}_{ij}(\bar{n}_{ij}) = \ln \left(\sigma \bar{f}_{ij}/\bar{r}_{ij}\right) + \ln \left(w_i^{\sigma}\right) - \ln \left(E_j P_j^{\sigma-1}\right). \tag{153}$$

Intensive margin of exports per product for exporting firms. As in our baseline model, we define the average revenue potential when a share n_{ij} of $i's$ firms sell in j as

$$\bar{\rho}_{ij}\left(n_{ij}\right) \equiv \frac{1}{n_{ij}} \int_0^{n_{ij}} \rho_{ij}(n)\ dn. \tag{154}$$

The definition of sales per product, averaged across exporters, $\bar{x}_{ij} \equiv E_{\eta}\left[E\left[R_{ij}\left(\omega\right)|\omega \in \Omega_{ij}(\eta)\right]|\eta \in \mathcal{N}_{ij}\right],$ implies that

$$\bar{x}_{ij} = \bar{r}_{ij} \left[\left(\frac{w_i}{P_j} \right)^{1-\sigma} E_j \right] \int_{e_{ij}^*}^{\infty} \left[\int_{e_{ij}^*}^{\infty} E\left[r|\tilde{e}\right] \frac{dH^e(e|\tilde{e})}{1 - H^e(e_{ij}^*|\tilde{e})} \right] \frac{d\bar{H}^e(e)}{1 - \bar{H}^e(e_{ij}^*)}, \quad e_{ij}^* \equiv \sigma \frac{\bar{f}_{ij}}{\bar{r}_{ij}} \left[\left(\frac{w_i}{P_j} \right)^{\sigma} \frac{P_j}{E_j} \right].$$

Notice that assumption (151) implies that we can write the average revenue potential conditional only on $\tilde{e}$.

We then consider again the transformation $n = 1 - H_{ij}^{e}(e)$ such that $\tilde{e} = \bar{\epsilon}_{ij}(n)$ and $dH_{ij}^{e}(\tilde{e}) = -dn$. Since $1 - H_{ij}^{e}(e_{ij}^*) = n_{ij}$ and $\lim_{e\to\infty} H_{ij}^{e}(e) = \infty$,

$$\bar{x}_{ij} = \bar{r}_{ij} \left[\left(\frac{w_i}{P_j} \right)^{1-\sigma} E_j \right] \int_0^{n_{ij}} E\left[r|\tilde{e} = \bar{\epsilon}_{ij}(n)\right] \frac{dn}{n_{ij}}.$$

Thus,

$$\ln \bar{x}_{ij} - \ln \bar{\rho}_{ij}(n_{ij}) = \ln \left(\bar{r}_{ij}\right) + \ln \left(w_i^{1-\sigma}\right) + \ln \left(E_j P_j^{\sigma-1}\right). \tag{155}$$

B.5 Non-CES Preferences

B.5.1 Environment

Demand. In country j with income y_j, we assume that the Marshallian demand function for product ω that can be written as

$$q_j\left(p\left(\omega\right);\boldsymbol{p}_j,y_j\right)=q\left(p\left(\omega\right);P_j\left(\boldsymbol{p}_j,y_j\right),y_j\right) \tag{156}$$

where $P_j\left(\boldsymbol{p}_j,y_j\right)$ is a price aggregator and $\boldsymbol{p}_j$ is the vector of all prices in market j.This class of demand functions includes a number of homothetic and non-homothetic examples, as discussed in Arkolakis et al. (2019a) and Matsuyama and Ushchev (2017).

We make two assumptions following Arkolakis et al. (2019a). First, we assume that the demand function features a choke price or in other words, for each $P_j\left(\boldsymbol{p},y\right)$ there exists $a\in\mathbb{R}$ such that for all $x\geq a$. This way we can abstract from the fixed cost of entry – that is, we assume that $\bar{f}_{ij}=0$ for all i and j. Second, we assume that the demand elasticity $\varepsilon_j\left(p\left(\omega\right);P,y\right)=-\partial\ln q\left(p\left(\omega\right);P,y\right)/\partial\ln p$ is decreasing in price. For exposition, we suppress the dependence of the demand function and its elasticity on on P and y.

Production. We assume that the production function is

$$C_{ij}(\omega,q)=w_i\frac{\tau_{ij}(\omega)}{a_i(\omega)}\frac{\bar{\tau}_{ij}}{\bar{a}_i}q.$$

Notice that, relative to the baseline, we abstract from the fixed cost of entry. In this case, the extensive margin of firm exports arises from the chock price in demand. We define the firm-specific cost shifter as

$$c_{ij}\left(\omega\right)\equiv\frac{\tau_{ij}\left(\omega\right)}{a_i\left(\omega\right)}$$

Since the production function is constant returns to scale, the quantity for each firm ω can be defined for each pair of markets separately. Thus, given aggregates P and y, the problem of firm ω from i when selling in j is

$$\pi_{ij}\left(\omega\right)=\max_{p(\omega)}\left\{\left(p\left(\omega\right)-w_i\frac{\bar{\tau}_{ij}}{\bar{a}_i}c_{ij}\left(\omega\right)\right)q_{ij}\left(p\left(\omega\right)\right)\right\}$$

The associated first order condition is given by

$$\left(1-w_i\frac{\bar{\tau}_{ij}}{\bar{a}_i}\frac{c_{ij}\left(\omega\right)}{p_{ij}\left(\omega\right)}\right)=-1/\left(\partial\ln q_j\left(p_{ij}\left(\omega\right)\right)/\partial\ln p\right)=1/\varepsilon_j\left(p_{ij}\left(\omega\right)\right).$$

Thus, markups are inversely related to the elasticity of demand:

$$m_{ij}\left(\omega\right)\equiv\frac{p_{ij}\left(\omega\right)}{w_i\frac{\bar{\tau}_{ij}}{\bar{a}_i}c_{ij}\left(\omega\right)}=\frac{\varepsilon_j\left(p_{ij}\left(\omega\right)\right)}{\varepsilon_j\left(p_{ij}\left(\omega\right)\right)-1}.$$

Furthermore, our second assumption guarantees that the markup is strictly decreasing on the marginal cost, that the price is strictly increasing on marginal cost, and that quantities and sales are strictly decreasing on the marginal cost (see related arguments in Arkolakis et al. (2019a)). This implies that we can perform a change of variables to express all variables in terms of the marginal cost of production:

$$\pi\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega)\right) = \left(\frac{m\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega)\right) - 1}{m\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega)\right)}\right) R\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega)\right). \tag{157}$$

Since a higher marginal cost lowers markups and sales, the profit function is strictly decreasing on $w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega)$. Therefore, given every P and y, there exists a unique revenue potential threshold that determines entry into a market:

$$\omega \in \Omega_{ij} \Leftrightarrow w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega) \leq c_j^*(P_j, y_j) \quad \text{such that} \quad \pi\left(c_j^*(P_j, y_j); P_j, y_j\right) = 0. \tag{158}$$

Conditional on entering, the revenues and profits are

$$R_{ij}(\omega) = R_{ij}\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega); P_j, y_j\right) \quad \text{and} \quad \pi_{ij}(\omega) = \pi\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega); P_j, y_j\right). \tag{159}$$

Entry. Let us assume that firms pay a fixed entry cost $\bar{F}_i$ in domestic labor to get a draw of variety characteristics from a distribution:

$$v_i(\omega) = \{a_i(\omega), \tau_{ij}(\omega)\}_j \sim G_i(v) \tag{160}$$

In expectation, firms only pay the fixed entry cost if ex-ante profits exceed entry them:

$$\sum_j E\left[\max\{\pi_{ij}(\omega);\ 0\}\right] = w_i \bar{F}_i, \tag{161}$$

Market clearing. We follow Dekle et al. (2008) by introducing exogenous international transfers, so that spending is

$$E_i = w_i \bar{L}_i + \bar{T}_i, \quad \sum_i \bar{T}_i = 0. \tag{162}$$

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

$$y_i = w_i \bar{L}_i = \int_{\omega \in \Omega_{ij}} R_{ij}\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c_{ij}(\omega); P, y\right). \tag{163}$$

Equilibrium. Given the arbitrary distribution in (160) , the equilibrium is defined as the vector $\{P_i, y_i, \{\Omega_{ij}\}_j, N_i, E_i, w_i\}_i$ satisfying the price index in (156), (158), (161), (162), (163) for all i.

B.5.2 Extensive and Intensive margin of Firm-level Export

We now turn to the characterization of the semiparametric gravity equations. We consider the distribution of firm-specific shifts of marginal costs implies by G_i:

$$c_{ij}(\omega) \sim H_{ij}(c), \tag{164}$$

where H_{ij} has full support in $\mathbb{R}_+$.

Extensive margin of firm-level exports. The share of firms of country i serving market j is $n_{ij} = Pr\left[\omega \in \Omega_{ij}\right] = Pr\left[c_{ij}(\omega) \leq \frac{\bar{a}_i}{\bar{\tau}_{ij} w_i} c_j^* \left(P_j, y_j\right)\right]$.

$$n_{ij} = H_{ij}\left(\frac{\bar{a}_i}{\bar{\tau}_{ij} w_i} c_j^* \left(P_j, y_j\right)\right)$$

We define $\bar{\epsilon}_{ij}(n) \equiv \left(H_{ij}\right)^{-1}(n)$. Notice that it is now strictly increasing in n. Thus,

$$\ln \bar{\epsilon}_{ij}(n_{ij}) = \ln\left(\bar{a}_i / \bar{\tau}_{ij}\right) - \ln w_i + \ln c_j^* \left(P_j, y_j\right). \tag{165}$$

Thus, in this case, we obtain again a semiparametric gravity equation for the extensive margin of firm exports.

Intensive margin of firm-level exports. The average revenue of firms from country i in country j is $\bar{x}_{ij} \equiv E\left[R_{ij}(\omega) | \omega \in \Omega_{ij}\right]$.

$$\bar{x}_{ij} = \int_0^{\frac{\bar{a}_i}{\bar{\tau}_{ij} w_i} c_j^*(P_j, y_j)} R_{ij}\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} c; P_j, y_j\right) dH_{ij}(c)$$

We can then use the transformation, $n = H_{ij}(c)$ such that $dn = dH_{ij}(c)$, $c = \bar{\epsilon}_{ij}(n)$, and $n_{ij} = H_{ij}\left(\frac{\bar{a}_i}{\bar{\tau}_{ij} w_i} c_j^* \left(P_j, y_j\right)\right)$. Thus,

$$\bar{x}_{ij} = \int_0^{n_{ij}} R_{ij}\left(w_i \frac{\bar{\tau}_{ij}}{\bar{a}_i} \bar{\epsilon}_{ij}(n); P_j, y_j\right) dn.$$

Using (165),

$$\bar{x}_{ij} = \int_0^{n_{ij}} R_{ij}\left(c_j^* \left(P_j, y_j\right) \frac{\bar{\epsilon}_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}; P_j, y_j\right) dn. \tag{166}$$

In this case, we can derive an expression for average sales as function of $\bar{\epsilon}_{ij}(n)$. So, although we do not have a gravity equation for average firm exports, this is entirely determined by the function governing the semiparametric gravity equation for the extensive margin of firm exports.

C Estimation

C.1 Data

Table 4: Data Availability

Country Name	$\{\bar{x}_{ij}, n_{ij}, z_{ij}\}$ Origin	Dest.	$\{\tau_{ij}\}$ Origin	Dest.	Developed
(1)	(2)	(3)	(4)	(5)	(6)
AUS	42	13	3	0	1
AUT	34	34	2	0	1
BEL	36	33	3	0	1
BGR	42	29	3	0	0
BRA	42	13	3	0	0
CAN	40	29	3	0	1
CHE	0	33	0	0	1
CHN	0	32	0	0	0
CYP	16	31	1	0	1
CZE	36	33	2	28	0
DEU	26	35	1	0	1
DNK	42	34	3	0	1
ESP	42	34	3	0	1
EST	36	31	3	0	0
FIN	34	33	3	0	1
FRA	36	35	3	0	1
GBR	36	34	3	0	1
GRC	25	33	2	0	1
HRV	42	11	3	0	0
HUN	36	33	3	0	0
IDN	0	13	0	0	0
IND	0	32	0	0	0
IRL	35	35	3	0	1
ITA	36	34	3	0	1
JPN	0	29	0	0	1
KOR	41	13	3	0	1
LTU	34	31	3	0	0
LUX	30	29	3	0	1
LVA	30	29	3	0	0
MEX	41	27	3	0	0
MLT	22	27	3	0	1
NLD	36	35	3	0	1
NOR	42	31	3	0	1
POL	36	34	2	0	0
PRT	42	31	3	0	1
RUS	0	33	0	0	0
SVK	34	34	2	34	0
SVN	22	31	2	0	1
SWE	29	34	3	0	1
TUR	42	32	3	0	0
TWN	0	10	0	0	1
USA	34	32	1	31	1
Count > 0	35	42	35	3	26
Observations	1229		93		

C.2 Restricted Cubic Spline Implementation

We follow Harrell Jr (2001) in setting up our restricted cubic splines.[25] Formally we use a restricted cubic spline with knot values u_k for $k = 1, ..., K$:

$$f_1(\ln n) = \ln n$$

$$f_{k+1}(\ln n) = \frac{(\ln n - \ln u_k)_+^3 - \frac{(\ln n - \ln u_{k-1})_+^3 (\ln u_K - \ln u_{k+1}) - (\ln n - \ln u_k)_+^3 (\ln u_{k-1} - \ln u_k)}{(\ln u_k - \ln u_{k-1})}}{(\ln u_K - \ln u_1)^2},$$

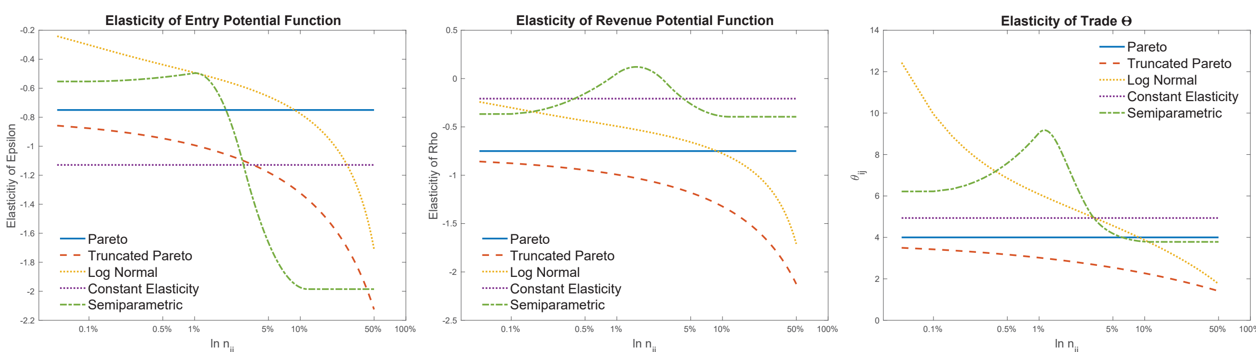
with the auxiliary function $(n)_+ = n$ if $n > 0$ and zero otherwise.

In our main specification, we choose $K = 4$ knots. To determine the values u_i, we follow Harrell Jr (2001) and divide the data into the 5th, 35th, 65th, and 95th percentiles.

C.3 Comparison with Literature

In Figure 7, we compare the baseline estimates in Figure 3 to calibrated elasticity functions obtained from estimates in literature of parametric distributions of firm fundamentals.

Figure 7: Baseline Estimates and Parametric Distributions in the Literature



(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Pareto is the Melitz-Pareto model in Chaney (2008) with a trade elasticity of four. Truncated Pareto uses the productivity distribution in Melitz and Redding (2015). Log-normal uses the baseline estimate of the productivity distribution in Head et al. (2014). Constant elasticity and Splice correspond to the baseline estimates in Section 5.4.

C.4 Robustness of Baseline Estimates in Section 5.4

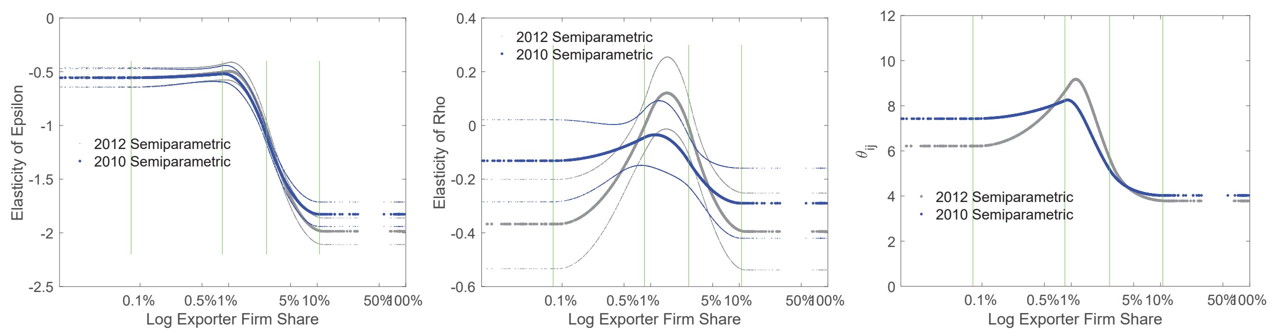
In this section, we investigate the robustness of the baseline estimates of $\bar{\rho}(n)$ and $\bar{\epsilon}(n)$ presented in Section 5.4. First, we show that results are similar when we use data for different years that have a similar country coverage. Second, we show that results are similar when we ignore observations associated with domestic sales. Third, we investigate how our results depend on the assumptions used to compute the number of domestic entrants, N_i.

[25] We avoid using a standard cubic spline, as they display degenerate out of sample properties.

C.4.1 Alternative Sample Years

Our baseline estimates use the sample of country pairs for 2012. Our data has a similar country coverage for all years between 2010 and 2014. We thus estimate the model with alternative samples for 2010 and 2014. Figures 8 and 9 show that results are broadly consistent with the baseline estimates obtained from the sample for 2012.

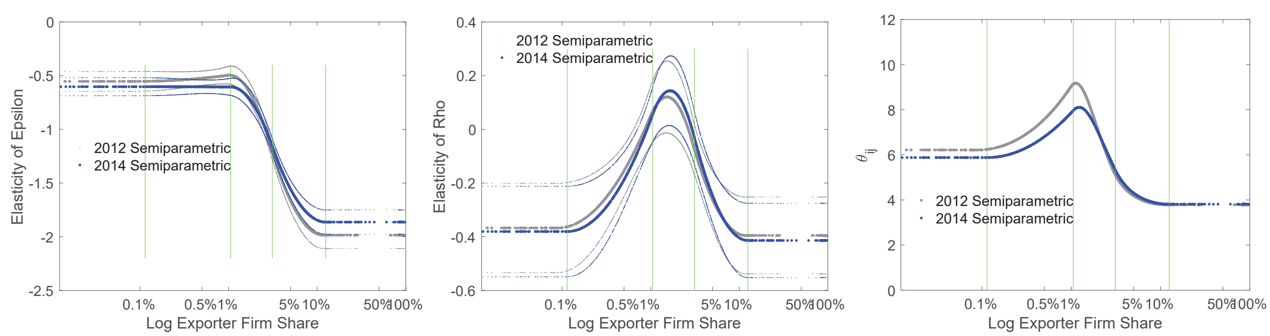
Figure 8: Semiparametric gravity estimation – 2010 Sample



(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2010 sample of 1,390 origin-destination pairs. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

Figure 9: Semiparametric gravity estimation – 2014 Sample



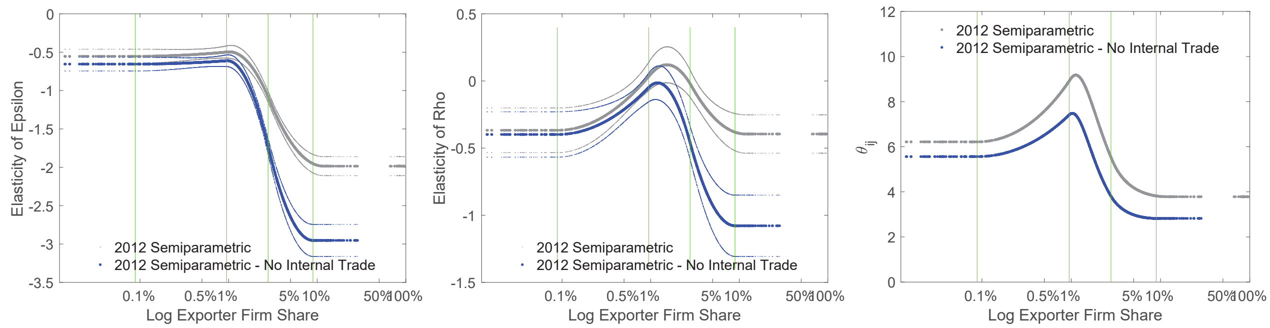
(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2010 sample of 1,496 origin-destination pairs. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

C.4.2 Baseline Sample Excluding Domestic Trade Observations

Our main estimation combines data on international trade with domestic sales. This requires not only n_{ij} and z_{ij} for $i \neq j$, but also n_{ii} and z_{ii}. Importantly, domestic sales are a high fraction of observations in the top knot where the trade elasticity is lower. To access whether these estimates depend on domestic sales, we re-run our estimation procedure in an alternative sample without domestic trade observations. Figure 10 shows that this has only a small impact on our baseline estimates of the trade elasticity function.

Figure 10: Semiparametric gravity estimation – 2012 Sample excluding domestic trade observations



(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

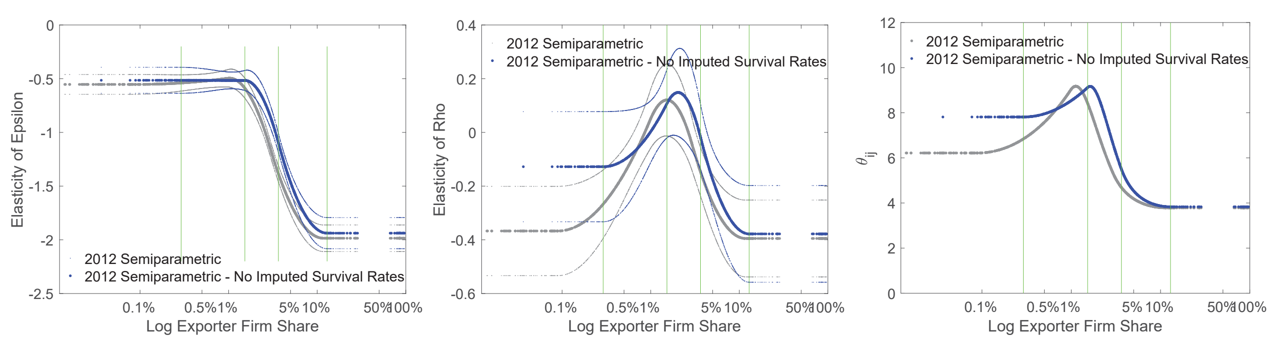
Note. Estimates obtained with GMM estimator in (38) in the 2012 sample excluding domestic trade observation. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

C.4.3 Alternative Measures of the Number of Entrants

In our data construction, we measure the share of successful entrants using one-year survival rates for manufacturing firms. This accounts for the fact that not all entrant firms are successful in entry and may leave the market (in the spirit of Melitz (2003)). We now conduct four robustness checks with respect to the construction of n_{ii}. We first exclude all countries with imputed values of n_{ii} from our baseline sample. We also re-estimate the elasticity functions under the alternative assumptions that n_{ii} is either one (survival rate of 100%), the 2-year firm survival rate, or the 3-year firm survival rate.

Alternative sample excluding origin countries with imputed survival rate. In Figure 11, we replicate our baseline estimation in a sample that excludes all observations for origin countries with imputed values of the one-year survival rate. Results are roughly similar to our baseline estimates.

Figure 11: Semiparametric gravity estimation – 2012 Sample excluding countries with imputed survival rates

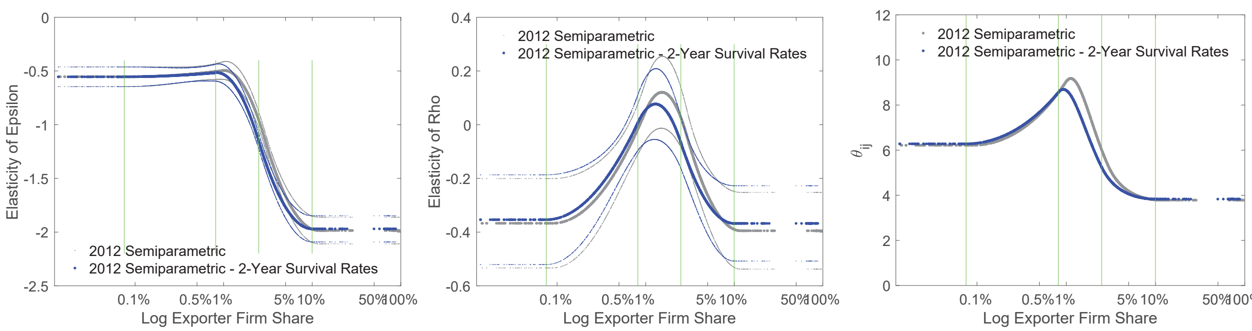


(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 Sample excluding countries with imputed survival rates. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

Alternative sample with n_{ii} measured with survival rates over different horizons. We now investigate how our baseline estimates change when we measure n_{ii} using 2-year and 3-year survival rates in manufacturing. Figures 12 and 13 show that the estimated elasticity functions are almost identical in both cases.

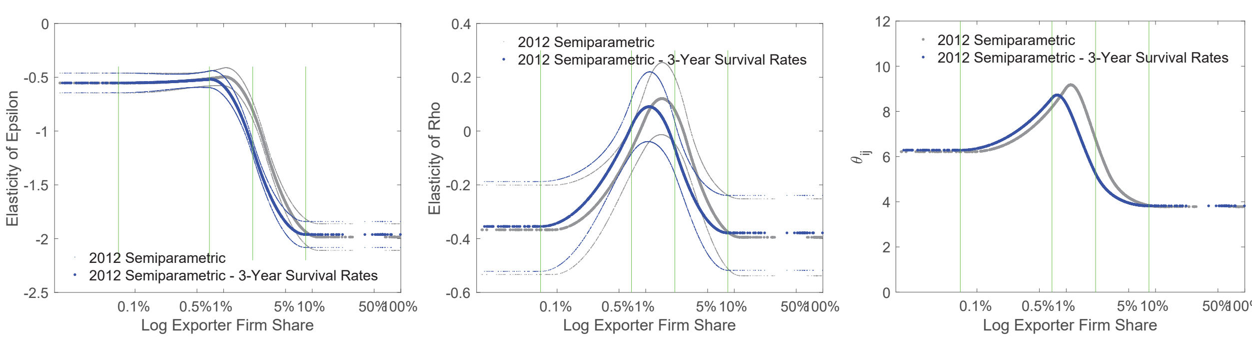
Figure 12: Semiparametric gravity estimation – 2012 Sample with n_{ii} measured as two-year survival rate in manufacturing



(a) Elasticity of $\bar{\epsilon}\,(n)$ (b) Elasticity of $\bar{\rho}\,(n)$ (c) Implied $\theta\,(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 of Appendix C.1. n_{ii} measured with two-year survival rate in manufacturing. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

Figure 13: Semiparametric gravity estimation – 2012 Sample with n_{ii} measured as three-year survival rate in manufacturing

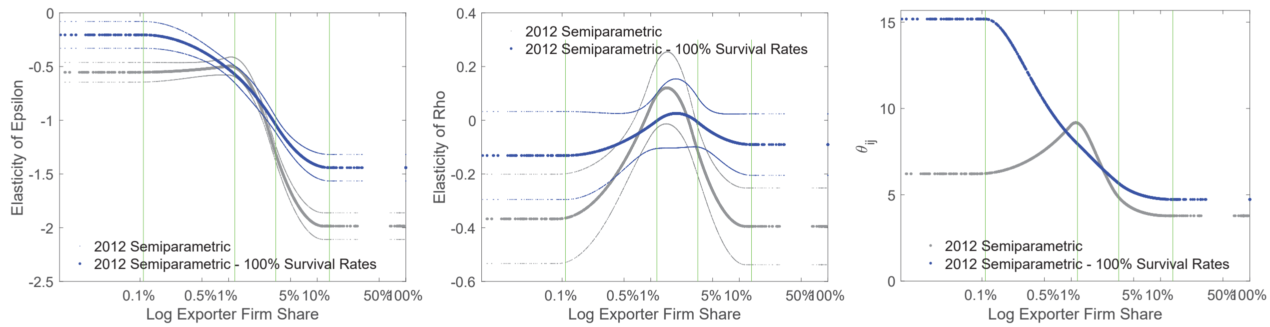


(a) Elasticity of $\bar{\epsilon}\,(n)$ (b) Elasticity of $\bar{\rho}\,(n)$ (c) Implied $\theta\,(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 of Appendix C.1. n_{ii} measured with three-year survival rate in manufacturing. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

Alternative sample with $n_{ii} = 1$ (survival rate of 100%). In Figure 14, we replicate our baseline estimates under the assumption that all entrants produce for the domestic market (i.e., $\bar{f}_{ii} = 0$ and $n_{ii} = 1$). In panel (a), we find that $\bar{\epsilon}\,(n)$ displays a similar shape, however point estimates in the bottom range of the support are near zero. This implies a much higher trade elasticity $\theta\,(n)$ for low levels of firm export share since the elasticity of $\bar{\epsilon}\,(n)$ enters the denominator of the trade elasticity definition in (18).

Figure 14: Semiparametric gravity estimation – 2012 Sample with $n_{ii} = 1$



(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 of Appendix C.1. Sample construction imposes $n_{ii} = 1$. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a single group ($G = 1$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

C.5 Additional Results

Our baseline estimates impose that the elasticity functions are identical for all exporter-destination pairs ($G = 1$). In this section, we estimate alternative specifications where we allow the elasticity function to vary across groups of countries.

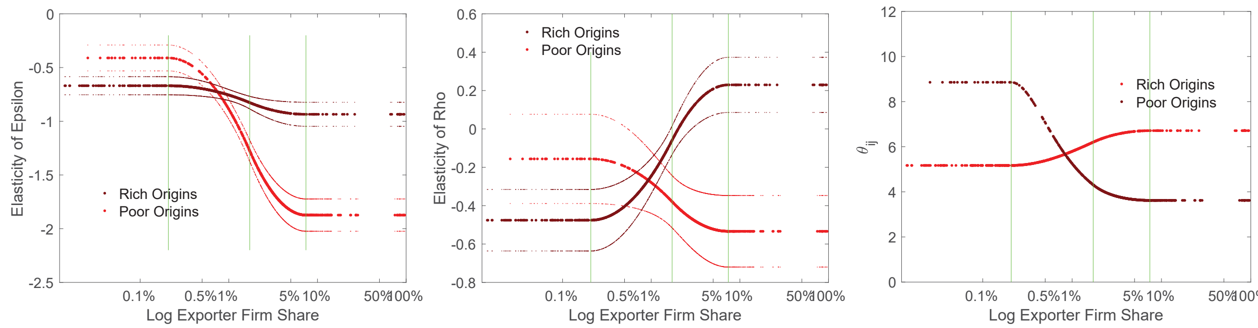
C.5.1 Heterogeneity with respect to per capita income

We first implement our estimation procedure with country groups defined defined in terms of per capita GDP. This type of heterogeneity in trade elasticity has been explored by Adao et al. (2017). We use a cutoff of \$9,000 of per capita GDP in 2002 to divide our sample into developed and developing nations. Column (6) of Table 4 in Appendix C.1 shows the list of developed and developing countries in our sample.

Figure 15 reports estimates for two groups defined in terms of development of the origin country. Panel (a) indicates that the extensive margin elasticity varies less with the exporter firm share in developed origin countries. In addition, Panel (b) indicates that, for low levels of exporter firm share, selection forces into exporting are stronger in developed countries. This pattern is reversed for high levels of the exporter firm share. Panel (c) shows the offsetting effects of $\bar{\epsilon}(n)$ and $\bar{\rho}(n)$ on the trade elasticity $\theta(n)$. For developed countries, the trade elasticity remains around six in the entire support. However, for developing countries, the trade elasticity falls with the exporter firm share. It is around nine when n_{ij} is low, but it is only four when n_{ij} is above 10%.

We also implement our estimation procedure for four groups defined in terms of per capita income of both the origin and destination countries. Figure 16 shows that the per capita income of the destination country does not have a large impact on the estimates reported in Figure 15.

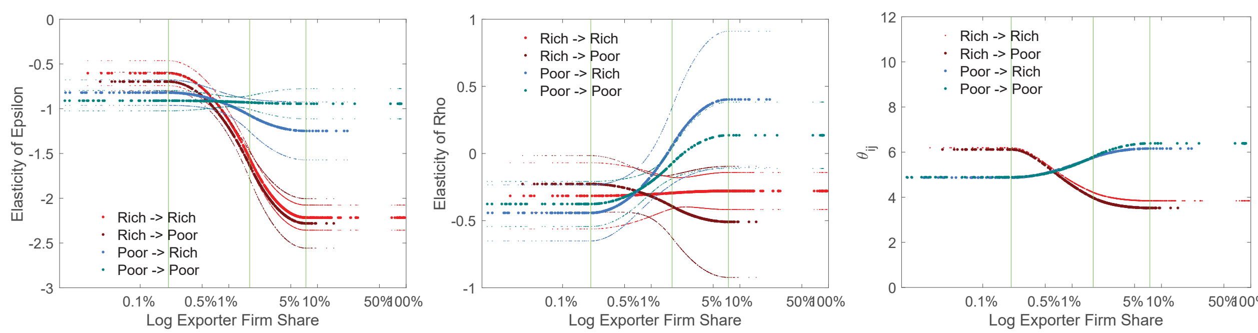
Figure 15: Semiparametric gravity estimation – Country groups defined in terms of per capita income of origin country



(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 of Appendix C.1. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a two groups ($G = 2$). Groups defined in terms of origin country per capita income – see column (6) of Table 4 of Appendix C.1. Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

Figure 16: Semiparametric gravity estimation – Country groups defined in terms of per capita income of origin and destination countries



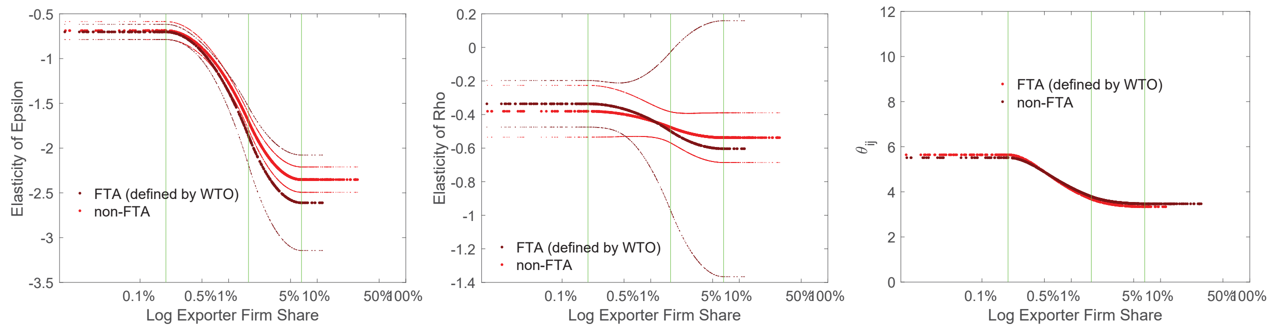
(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,479 origin-destination pairs described in Table 4 in Appendix C.1. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a four groups ($G = 4$). Groups defined in terms of per capita income of origin and destination countries – see column (6) of Table 4 of Appendix C.1. Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Standard errors clustered by origin-destination pair.

C.5.2 Heterogeneity with respect to participation in free trade area

We implement our estimation procedure for two groups of exporter-importer pairs defined as country pairs inside and outside a common Free Trade Areas (FTA) (using the CEPII bilateral gravity dataset). A large body of literature has documented that membership in free trade areas reduces trade costs – see Head and Mayer (2013). We investigate here if it also affects the different elasticity margins of trade flows. Figure 17 indicates that there are only small differences in the estimates for countries inside and outside common free trade areas.

Figure 17: Semiparametric gravity estimation – Country groups defined in terms of membership in free trade areas



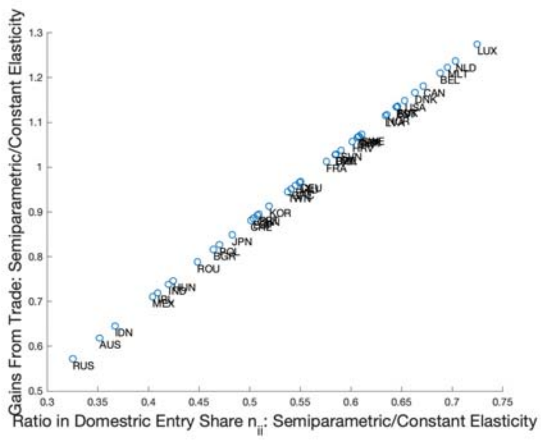
(a) Elasticity of $\bar{\epsilon}(n)$ (b) Elasticity of $\bar{\rho}(n)$ (c) Implied $\theta(n)$

Note. Estimates obtained with GMM estimator in (38) in the 2012 sample of 1,443 origin-destination pairs described in Table 4 of Appendix C.1. Estimates obtained with a cubic spline over four intervals ($K = 4$) for a two groups ($G = 2$). Calibration of $\tilde{\sigma} = 2.9$ from Hottman et al. (2016). Groups defined as country pairs inside and outside a common Free Trade Areas (FTA). We omit domestic trade. Standard errors clustered by origin-destination pair.

D Counterfactual Analysis

D.1 Additional Results

Figure 18: Importance of Firm Heterogeneity for the Gains from Trade



Note. Gains from Trade is the percentage change in the real wage implied by moving from autarky to the observed equilibrium in 2012. Gains from trade for semiparametric specification computed with the formula in Section 3.2 for $\hat{n}_{ii}$ and $\hat{N}_i$ solving the system in Appendix A.5 and the baseline spline estimates in Figure 3. Gains from trade for the constant elasticity specification computed with the formula in Section 3.3 and the trade elasticity of five reported in Table 3.

Table 5: Gains From Trade

Country	Trade	$\hat{W}$		
		Generalized	Non-linear Spline	
Origin	Penetration	Pareto	$\hat{n}_{ii}$	$\hat{n}_{ii}$ and $\hat{N}_i$
AUS	0.04 %	0.83 %	1.09 %	0.42 %
AUT	0.17 %	3.96 %	5.20 %	4.32 %
BEL	0.26 %	6.34 %	8.35 %	8.10 %
BGR	0.20 %	4.57 %	6.00 %	3.87 %
BRA	0.04 %	0.80 %	1.04 %	0.99 %
CAN	0.10 %	2.06 %	2.70 %	2.50 %
CHE	0.14 %	3.15 %	4.13 %	2.39 %
CHN	0.03 %	0.70 %	0.92 %	0.49 %
CYP	0.16 %	3.66 %	4.80 %	4.85 %
CZE	0.23 %	5.48 %	7.21 %	5.50 %
DEU	0.15 %	3.23 %	4.23 %	2.73 %
DNK	0.13 %	2.90 %	3.80 %	3.19 %
ESP	0.09 %	1.85 %	2.43 %	1.68 %
EST	0.30 %	7.40 %	9.76 %	8.85 %
FIN	0.13 %	2.89 %	3.78 %	2.67 %
FRA	0.11 %	2.35 %	3.08 %	2.53 %
GBR	0.11 %	2.37 %	3.11 %	2.94 %
GRC	0.10 %	2.11 %	2.77 %	2.39 %
HRV	0.17 %	3.89 %	5.10 %	4.71 %
HUN	0.31 %	7.85 %	10.36 %	5.43 %
IDN	0.06 %	1.31 %	1.71 %	0.64 %
IND	0.04 %	0.86 %	1.13 %	0.64 %
IRL	0.17 %	3.91 %	5.13 %	0.74 %
ITA	0.08 %	1.71 %	2.24 %	1.50 %
JPN	0.04 %	0.86 %	1.13 %	0.67 %
KOR	0.11 %	2.48 %	3.24 %	2.05 %
LTU	0.30 %	7.56 %	9.97 %	7.60 %
LUX	0.22 %	5.27 %	6.93 %	7.45 %
LVA	0.20 %	4.60 %	6.04 %	5.94 %
MEX	0.13 %	2.89 %	3.79 %	2.03 %
MLT	0.51 %	15.62 %	20.84 %	25.22 %
NLD	0.24 %	5.74 %	7.56 %	8.02 %
NOR	0.09 %	2.02 %	2.64 %	1.83 %
POL	0.17 %	3.74 %	4.90 %	3.12 %
PRT	0.11 %	2.48 %	3.25 %	2.13 %
ROU	0.14 %	3.19 %	4.18 %	2.59 %
RUS	0.07 %	1.54 %	2.01 %	0.54 %
SVK	0.29 %	7.22 %	9.52 %	7.95 %
SVN	0.26 %	6.27 %	8.25 %	6.48 %
SWE	0.14 %	3.13 %	4.10 %	3.18 %
TUR	0.10 %	2.24 %	2.93 %	2.96 %
TWN	0.17 %	3.96 %	5.20 %	3.28 %
USA	0.05 %	1.08 %	1.42 %	1.47 %
Average	0.16 %	3.68 %	4.84 %	3.92 %

D.2 Computation Algorithms

D.2.1 Hat Algebra

We now describe an algorithm to compute the changes in aggregate outcomes that solve the system in Appendix A.2 for an arbitrary trade cost change from $\{\bar{\tau}_{ij}\}_{ij}$ to $\{\bar{\tau}'_{ij}\}_{ij}$.

1. Compute the partition of the shock with length R: $d\ln\bar{\tau}_{ij} = \frac{1}{R}\left(\ln\bar{\tau}'_{ij} - \ln\bar{\tau}_{ij}\right)$. Consider the initial equilibrium with $\theta_{ij}(n^0_{ij})$ and $\{x^0_{ij}, y^0_{ij}, \iota^0_j\}$.

2. For each step r, we consider the initial conditions $\left(\varepsilon_{ij}(n^{r-1}_{ij}), \varrho_{ij}(n^{r-1}_{ij}), \theta_{ij}(n^{r-1}_{ij})\right)$ and $\{x^{r-1}_{ij}, y^{r-1}_{ij}, \iota^{r-1}_j\}$.

   (a) Compute $(d\ln\boldsymbol{\tau}^{w,r}, d\ln\boldsymbol{\tau}^{p,r})$ and $d\ln\boldsymbol{w}^r$ using (65) (for a given numerarie with $d\ln w^r_m = 0$).

   (b) Solve $d\ln\boldsymbol{P}^r$ using (64).

   (c) Solve $d\ln n^r_{ij}$ and $d\ln\bar{x}^r_{ij}$ using (49) and (50).

   (d) Solve for $d\ln N^r_i$ using (55).

   (e) Compute the change in bilateral trade flows: $d\ln X^r_{ij} = d\ln\bar{x}^r_{ij} + d\ln n^r_{ij} + d\ln N^r_i$.

   (f) Compute the initial conditions for the next step: $X^r_{ij} = X^{r-1}_{ij}e^{d\ln X^r_{ij}}$ and $n^r_{ij} = n^{r-1}_{ij}e^{d\ln n^r_{ij}}$.

   (g) Compute $x^r_{ij} = X^r_{ij}/\sum_o X^r_{oj}$, $y^r_{ij} = X^r_{ij}/\sum_d X^r_{id}$, $\iota^r_i = \left(\sum_d X^r_{id}\right)/\left(\sum_o X^r_{oi}\right)$, and $\left(\varepsilon_{ij}(n^r_{ij}), \varrho_{ij}(n^r_{ij}), \theta_{ij}(n^r_{ij})\right)$.

3. Compute changes in aggregate variables as

$$\hat{\boldsymbol{w}}^{linear} = \exp\left(\sum_{r=1}^{R} d\ln\boldsymbol{w}^r\right), \quad \hat{\boldsymbol{P}}^{linear} = \exp\left(\sum_{r=1}^{R} d\ln\boldsymbol{P}^r\right), \quad \hat{\boldsymbol{N}}^{linear} = \exp\left(\sum_{r=1}^{R} d\ln\boldsymbol{N}^r\right).$$

4. Use $\{\hat{w}^{linear}_i, \hat{P}^{linear}_i\}$ as an initial guess is the solution of the hat algebra system. Set the same numerarie as above, $\hat{w}_m \equiv 1$.

   (a) Given guess of $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{P}})$, compute

$$\frac{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\epsilon}_{ij}(n_{ij})} = \left(\hat{\bar{\tau}}_{ij}\right)^{\sigma-1}\left[\left(\frac{\hat{w}_i}{\hat{P}_j}\right)^{\sigma}\frac{\hat{P}_j}{\iota_j\hat{w}_j}\right],$$

$$\hat{\bar{x}}_{ij} = \left(\hat{\bar{\tau}}_{ij}\right)^{1-\sigma}\frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})}\left[\left(\frac{\hat{w}_i}{\hat{P}_j}\right)^{1-\sigma}\iota_j\hat{w}_j\right].$$

   From (48),

$$\hat{N}_i = \left[1 - \sum_j y_{ij}\left(\frac{\int_0^{n_{ij}}\frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\,dn}{\int_0^{n_{ij}}\frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij})}\,dn}\right) + \sum_j y_{ij}\frac{\hat{n}_{ij}\hat{\bar{x}}_{ij}}{\hat{w}_i}\left(\frac{\int_0^{n_{ij}\hat{n}_{ij}}\frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n)}\,dn}{\int_0^{n_{ij}\hat{n}_{ij}}\frac{\rho_{ij}(n)}{\bar{\epsilon}_{ij}(n_{ij}\hat{n}_{ij})}\,dn}\right)\right]^{-1}.$$

   (b) Compute the functions:

$$F^P_j\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{P}}\right) \equiv \hat{P}_j^{1-\sigma} - \sum_i x_{ij}\left(\left(\hat{\bar{\tau}}_{ij}\right)^{1-\sigma}\frac{\bar{\rho}_{ij}(n_{ij}\hat{n}_{ij})}{\bar{\rho}_{ij}(n_{ij})}\hat{w}_i^{1-\sigma}\hat{n}_{ij}\hat{N}_i\right)$$

$$F_i^w\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{P}}\right) \equiv \hat{w}_i - \sum_j y_{ij}\left(\hat{N}_i \hat{n}_{ij} \hat{\hat{x}}_{ij}\right).$$

(c) Find $\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{P}}\right)$ that minimizes $\left\{|F_j^P\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{P}}\right)|, |F_i^w\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{P}}\right)|\right\}$.

D.2.2 Gains from trade

We now describe an algorithm to compute the gains from trade described in Appendix A.5.

1. Define the uni-dimensional function

$$F_i(\hat{n}_{ii}^A) = \sum_j y_{ij}\left(1 - \frac{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n)}\, dn}{\int_0^{n_{ij}} \frac{\rho_{ij}(n)}{\epsilon_{ij}(n_{ij})}\, dn}\right) - \frac{1}{\hat{N}_i^A(\hat{n}_{ii}^A)} \frac{y_{ii}}{x_{ii}}\left(1 - \frac{\int_0^{n_{ii}\hat{n}_{ii}^A} \frac{\rho_{ii}(n)}{\epsilon_{ii}(n)}\, dn}{\int_0^{n_{ii}\hat{n}_{ii}^A} \frac{\rho_{ii}(n)}{\epsilon_{ii}(n_{ii}\hat{n}_{ii}^A)}\, dn}\right)$$

   where

$$\hat{N}_i^A(\hat{n}_{ii}^A) = \frac{\bar{\rho}_{ii}\left(n_{ii}\right)}{\epsilon_{ii}\left(n_{ii}\right)} \frac{\epsilon_{ii}\left(n_{ii}\hat{n}_{ii}^A\right)}{\bar{\rho}_{ii}\left(n_{ii}\hat{n}_{ii}^A\right)} \frac{1}{\hat{n}_{ii}^A} \frac{1}{x_{ii}}.$$

2. For each i, we find $\hat{n}_{ii}^A$ such that $F_i(\hat{n}_{ii}^A) = 0$. We then compute the gains from trade using equation (24).