Increasing Marginal Costs, Firm Heterogeneity, and the Gains from "Deep" International Trade Agreements

Bergstrand, Jeffrey H. and Cray, Stephen R. and Gervais, Antoine

University of Notre Dame, University of Notre Dame, Université de Sherbrooke

17 October 2022

Increasing Marginal Costs, Firm Heterogeneity, and the Gains from "Deep" International Trade Agreements*

Jeffrey H. Bergstrand,† Stephen R. Cray‡ and Antoine Gervais§

October 17, 2022

Abstract: Two parameters are central to several modern quantitative models of bilateral international trade flows: the elasticity of substitution in consumption (σ) and the inverse index of heterogeneity of firms' productivities (θ). However, structural parameter estimation applications using the seminal Feenstra econometric methodology typically focus on estimates of only σ and a bilateral export supply elasticity – which we will term γ. Separately, modern trade agreements are increasingly "deep," meaning they reduce fixed trade costs alongside variable trade costs (such as tariffs). Although Melitz models of international trade recognize both trade costs theoretically, very little is known quantitatively about their *relative* impacts on trade and welfare. In this paper, we offer three contributions. First, in the spirit of Arkolakis (2010), we extend the canonical Melitz model of trade to allow for increasing marginal market-penetration costs, alongside fixed marketing costs, to show theoretically the importance of accounting for increasing marginal costs (via γ) – in the presence of firm heterogeneity – in understanding the relative impacts on trade, extensive margins, intensive margins, and welfare of reducing fixed trade costs and variable trade costs. Second, we provide a microeconomic foundation for estimating all three parameters using the Feenstra econometric methodology alongside a gravity equation. Third, we demonstrate the importance of increasing marginal costs using two counterfactual exercises. One illustrative quantitative implication for U.S. trade policy is that, under (empirically rejected) constant marginal costs, fixed trade costs would have to be reduced by 57 percent for a welfare-equivalent reduction in variable trade costs of 3 percent; by contrast, under (empirically supported) increasing marginal costs, fixed trade costs would have to be reduced by only 14 percent.

Keywords: International trade, deep trade agreements, Melitz models, increasing marginal costs, gravity equations

JEL Classification Numbers: F1

**Acknowledgements*: We are grateful to editor Costas Arkolakis, two referees, Scott Baier, Matthew Clance, William Powers, David Riker, Anson Soderbery, and conference and seminar participants at the U.S. International Trade Commission, the University of Cape Town, and the University of Pretoria for very helpful comments on earlier drafts. All remaining errors are our own.

†*Affiliation:* Department of Finance, Department of Economics, and Kellogg Institute for International Studies, University of Notre Dame, and CESifo Munich. *Address:* Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556 USA. *E-mail:* bergstrand.1@nd.edu.

‡*Affiliation:* Department of Finance, University of Notre Dame, Notre Dame, IN 46556 USA. *E-mail:* craystephen@gmail.com.

§*Affiliation:* Department of Economics, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1. *E-mail:* Antoine.Gervais2@USherbrooke.ca.

1 Introduction

Central to the post-2000 modern quantitative models of international trade are two parameters. The first – and arguably most visible – is the elasticity of substitution in consumption among differentiated products, σ. This parameter is key in the seminal theoretical foundation for the gravity equation with Armington preferences in Anderson (1979), monopolistic competition model of intra-industry trade with Dixit-Stiglitz preferences in Krugman (1980), analysis of optimal tariffs in Broda et al. (2008) and Ossa (2016), and a vast array of applied computable general equilibrium (CGE) models used for trade-policy analyses, cf., U.S. International Trade Commission (2019). The second parameter, which surfaced over the last 20 years, is a (inverse) measure of heterogeneity of firms' productivities, which we denote θ. Motivated by theoretical models of Eaton and Kortum (2002) and Melitz (2003), θ is the key parameter in modern quantitative trade models with heterogeneous firms for capturing the infamous "trade elasticity" (i.e., elasticity of bilateral trade with respect to *ad valorem* bilateral variable trade costs), one of two sufficient statistics to measure welfare effects of trade liberalizations in a broad set of quantitative trade models (cf., Arkolakis et al. (2012), henceforth, ACR).

A common assumption to these quantitative trade models is constant marginal costs. By contrast, the most widely respected structural method for estimating σ – introduced by Feenstra (1994) and further developed by Broda and Weinstein (2006) (henceforth, F/BW) and Broda et al. (2008) – assumes bilateral export supply prices are positive functions of the level of exports to foreign markets, which implies increasing marginal costs of exporting to each destination market. We will refer to the parameter that governs the bilateral export supply elasticity as γ. Although σ and θ currently play central roles in trade theory and calibration exercises of new quantitative trade models, γ has been largely ignored. Moreover, the bilateral export supply elasticity has typically been incorporated in these econometric analyses in an *ad hoc* manner. For instance, in Feenstra (1994), Broda and Weinstein (2006), and Soderbery (2015, 2018), positively-sloped bilateral export supply curves were simply assumed. More recently, Feenstra et al. (2018) extend the method of Feenstra (1994) allowing firm heterogeneity based upon a standard Melitz model with constant marginal costs, but still introduce an equation that "plays the role of a supply curve" (p. 140).

Separately, modern international trade agreements – such as free trade agreements (FTAs) – are increasingly "deep," meaning that, beyond the typical reductions in *ad valorem* tariff rates found in "shallow" agreements, they reduce *fixed* trade costs. The World Bank has recently compiled a large data set on deep trade agreements' (DTAs) provisions. The database, summarized comprehensively in Hofmann et al. (2017), documents the extensive growth in DTAs over the past twenty years. A notable economic difference concerning these deep provisions is that they relate to regulatory convergences and administrative liberalizations that are unrelated to the quantity of goods exported and are more readily interpreted as reducing fixed trade costs. For instance, the most popular non-tariff measures included in modern trade agreements are customs administration (often referred to as trade facilitation measures), competition policy, sanitary and photosanitary (SPS) regulations, and technical barriers to trade (TBT) regulations.

Recent empirical work using gravity equations indicates economically and statistically significant effects of indexes of DTAs' provisions on trade flows, cf., Kohl et al. (2016), Baier and Regmi (2020), Crowley et al. (2020), Breinlich et al. (2021) and Fontagne et al. (2022). By contrast, there has been a dearth in numerical analyses of variable *versus* fixed bilateral trade costs in either standard CGE models (such as GTAP) or in the new quantitative trade models. Zhai (2008) is one of the earliest – and rare – studies to introduce a standard Melitz model (with constant marginal costs) into a global CGE model of world trade and to contrast the trade and welfare effects of a 5 percent variable trade-cost reduction relative to a 50 percent fixed trade-cost reduction.[1] In Zhai (2008), it would take a *29 percent* reduction in bilateral fixed trade costs to achieve the equivalent gain in welfare as a 4 percent reduction in *ad valorem* variable trade costs (a ratio of 7.25:1). More recently, however, Arkolakis et al. (2021) extend the canonical Melitz model of trade to allow multiproduct firms facing constant marginal costs in core-product production, but allowing increasing marginal market-penetration costs and increasing marginal costs in non-core products. Among several findings, one counterfactual implies that it would take a 13 percent reduction in fixed trade costs with countries to generate the same welfare gain as a 4 percentage point reduction in tariff rates (or a ratio of 3.25:1). Such estimates suggest evaluating the role of increasing versus constant marginal costs to address the question: Why have countries increasingly pursued deep trade agreements?

Given these considerations, we now summarize our paper's contributions. Our first contribution, motivated by Arkolakis (2010), is to introduce increasing marginal costs (IMC) into the Melitz model via an empirically-tractable formulation of increasing marginal market-penetration costs. To get a sense of the impact of IMC on the trade elasticity, consider a simple Armington trade model. Figure 1 illustrates the attenuation of the intensive margin elasticity in the presence of a positively-sloped bilateral export supply curve, consistent with IMC. In the standard case of constant marginal costs (CMC), a one percent increase in *ad valorem* variable trade costs, $\Delta \ln \tau_{ij} = \overline{AD}$, lowers bilateral imports from country i to country j (IM_{ij}) by $\Delta \ln IM_{ij} = (1-\sigma)\Delta \ln \tau_{ij} = \overline{AB}$, where σ is the elasticity of substitution in consumption. However, with IMC, the same one percent increase in *ad valorem* variable trade costs lowers bilateral imports by less, $\Delta \ln IM_{ij} = \overline{AC} < \overline{AB}$. Figure

[1] We will discuss Balistreri et al. (2011) and Dixon et al. (2016) below in section 6.

1 clearly illustrates that under CMC the trade elasticity is a function solely of the elasticity of substitution, whereas under IMC the trade elasticity also depends on an index of the shape of the supply curve.

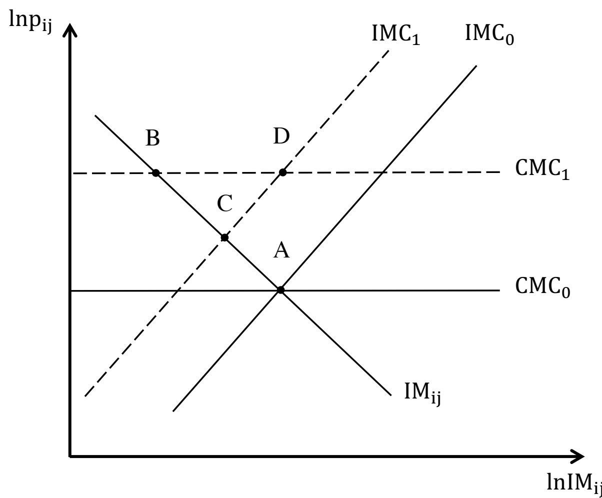

Figure 1: Increasing Marginal Costs vs. Constant Marginal Costs

Our extended model yields several analytical results. First, we derive a gravity equation similar to that in ACR except that the extensive margin elasticity and the trade elasticity with respect to (*ad valorem*) variable trade costs are magnified; yet, the variable trade-cost intensive margin elasticity is diminished, consistent with Figure 1. An implication is that variable trade-cost liberalizations with IMC will have more firm entry and exit and more labor reallocations than under CMC. Second, the fixed trade-cost trade elasticity – which is a function of the variable trade-cost extensive margin elasticity relative to the variable trade-cost intensive margin elasticity – is magnified under IMC. Moreover, a further implication of IMC is that the fixed trade-cost trade elasticity is magnified relative to the variable trade-cost trade elasticity, which will be important in understanding the welfare-equivalent impacts of fixed trade-cost liberalizations relative to variable trade-cost liberalizations in deep FTAs. Third, allowing IMC diminishes the welfare effect of a given change in the domestic trade share (for a given θ). The intuition is that real wage gains from a trade liberalization can be traced to changes in average productivity. In the Melitz model, changes in average productivity are proportionate to changes in output of the zero-cutoff-profit (ZCP) productivity firm. In the CMC case, the latter are directly proportionate to productivity changes of the ZCP firm. However, with increasing marginal costs ($\gamma < \infty$), output of the ZCP firm rises less than proportionately to the change in the ZCP firm's productivity. The gains to average productivity are diminished at a rate of $1 + 1/\gamma$.

Our second contribution is to further develop the microeconomic foundation for the F/BW econometric approach to estimate σ and γ by accounting explicitly for firm heterogeneity. Unlike F/BW, our approach distinctly recognizes the importance of differences in the masses of exporting firms, which depend on the exporting country's labor-force size and the zero-cutoff-profit productivity threshold. In the context of the heterogeneous-firm models, one must account for both new import varieties from trade liberalizations as well as declining numbers of domestic varieties. The F/BW reduced-form estimating equation includes two variables and one interaction term. Our extension of the F/BW approach to account for firm heterogeneity motivates the inclusion potentially of 6 additional variables, for a total of 8 variables and 28 interaction terms. While we can show all 35 (right-hand-side) coefficients are functions of the three structural parameters only, the large number of nonlinear constraints precludes estimation of σ, γ, and θ simultaneously. Instead, we pursue a two-pronged approach, composed of two reduced-form estimating equations. In the first part of our estimation, we implement our extension of the F/BW reduced-form equation that allows us to estimate σ and γ while controlling explicitly for firm heterogeneity. In the second part, we use the gravity equation generated from our theoretical model to identify θ using the trade elasticity alongside the first part's estimate of γ. Our novel estimation approach yields median estimates across the distribution of industries of σ and γ of 6.45 and 6.00, respectively – approximately *35 and 50 percent larger*, respectively, than the comparable F/BW estimates ignoring firm heterogeneity. Moreover, our median estimate of θ from the second step is 8.50 – which is very close to Eaton and Kortum (2002)'s and Arkolakis (2010)'s preferred estimate of 8.28.

Our third contribution is to illustrate the impact of recognizing increasing marginal costs on the estimated effects of DTAs in the world. Goldberg and Pavcnik (2016) emphasized that economists have not paid sufficient attention to the study of the effects of trade-policy changes other than *ad valorem* tariff-rate changes, and that a better understanding of the effects of reduced fixed trade costs on international trade and economic welfare is critical. In this spirit, we conduct two numerical analyses. In the first exercise, we show that – even under IMC – the welfare gains from trade for an economy can be captured by a function of an economy's current intra-national trade share and the trade elasticity. This result is fully consistent with the main conclusion in ACR that the trade elasticity (independent of its structural interpretation) and the intra-national trade share are sufficient statistics to measure the welfare effect of a change in bilateral variable or fixed trade costs (τ_{ij} or f_{ij}, respectively). However, in the presence of IMC, the trade elasticity is higher (in absolute terms) and consequently the welfare gains lower, owing to a "welfare diminution effect" attributable to diminishing marginal returns. In a second exercise, we examine the *relative* impacts of variable trade-cost changes and fixed trade-cost changes. We show that, for typical values of σ and θ, under CMC ($\gamma = \infty$) the degree of liberalization of fixed trade costs needed to generate an equivalent increase in welfare is very large relative to the degree of liberalization of variable trade costs, questioning the increasing effort toward deep trade agreements. By contrast, under increasing marginal costs ($\gamma < \infty$), the degree of liberalization of fixed trade costs needed to generate an equivalent increase in welfare is *dramatically reduced* relative to the degree of liberalization of variable trade costs, which helps explain the attractiveness of deep trade agreements. For instance, we show for the United States that, under CMC, fixed trade costs would have to be reduced by 57 percent to provide the same increase in welfare as a reduction in variable trade costs of 3 percent. By contrast, under the empirically supported assumption of IMC, it would take only a 14 percent reduction in fixed trade costs to increase U.S. welfare by the same respective variable trade-cost reduction.

The remainder of this paper is as follows. In section 2, we introduce and solve our Melitz model allowing increasing marginal costs, asymmetric countries, and a Pareto distribution of productivities. In section 3, we solve for our gravity equation and trade elasticity, derive the variable- and fixed-trade-cost elasticities of extensive and (for variable trade costs) intensive margins, discuss welfare implications, and provide the intuition behind our "welfare diminution effect." In section 4, we discuss our econometric methodology, empirical specifications, and data sources. In section 5, we provide estimates of σ, γ, θ, and the variable- and fixed-trade-cost trade elasticities. In section 6, we provide numerical estimates of a counterfactual analysis of the impact of introducing increasing marginal costs on the welfare effects from trade and another counterfactual analysis demonstrating the importance of recognizing empirically-justified increasing marginal costs toward evaluating the quantitative welfare significance of liberalizations of fixed trade costs relative to those of variable trade costs, two components of (modern) deep trade agreements. In section 7, we offer some conclusions.

2 Theory

Our theoretical framework builds on the Melitz (2003) heterogeneous firms model. As in Chaney (2008) and Redding (2011), we allow for differences in countries' labor endowments and bilateral trade barriers and we assume a Pareto distribution for productivity draws. The Pareto distribution is particularly useful because it yields closed-form solutions that we can use to obtain clear theoretical predictions and to develop our novel econometric approach for the estimation. A key difference with the Melitz (2003) model is that our framework features an empirically tractable adaptation of the increasing marginal market-penetration cost aspect of Arkolakis (2010) to allow for the possibility of increasing marginal costs of providing output to any market. It seems reasonable to study the more general version of the model – especially one that motivates the econometrically tractable structural bilateral import demand and bilateral export supply functions in F/BW – and let the data determine the slope of the bilateral export supply curve, instead of imposing CMC *ex ante*.

2.1 Consumer Behavior

Our modeling of consumer behavior is standard. We assume a world with $j = 1, 2, ..., N$ countries. In each country, there is a mass of consumers, L_j, each endowed with one unit of labor (or a composite input we call "labor"). The preferences of the representative consumer in country j are a constant-elasticity-of-substitution (CES) function of the consumption of a continuum of differentiated varieties:

$$U_j = \left[\sum_{i=1}^{N} \int_{\nu \in \Omega_{ij}} b_i^{\frac{1-\sigma}{\sigma}} c_{ij}(\nu)^{\frac{\sigma-1}{\sigma}} d\nu \right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where $c_{ij}(\nu)$ is the quantity consumed of variety ν from country i, Ω_{ij} is the (endogenous) mass of varieties produced in country i and available for consumption in country j, $b_i > 0$ is an exogenous (inverse) preference parameter for country i's varieties (c.f., Anderson and van Wincoop (2003)) and $\sigma > 1$ is the elasticity of substitution between varieties.

The representative consumer maximizes utility subject to the standard income constraint, such that the optimal aggregate demand function for each variety is given by:

$$c_{ij}(\nu) = E_j P_j^{\sigma-1} b_i^{1-\sigma} p_{ij}^c(\nu)^{-\sigma}, \tag{2}$$

where E_j denotes aggregate expenditures in country j, $p_{ij}^c(\nu)$ is the price of a unit of variety ν from country i facing the consumer in country j, and P_j defined as:

$$P_j = \left[\sum_{i=1}^{N} \int_{\nu \in \Omega_{ij}} b_i^{1-\sigma} p_{ij}^c(\nu)^{1-\sigma} d\nu \right]^{\frac{1}{1-\sigma}} \tag{3}$$

is the price index dual to the consumption index $C_j \equiv U_j$. Because consumers have no taste for leisure, they always supply their unit of labor to the market at the prevailing wage rate, w_j. Hence, the equilibrium labor supply is L_j.

2.2 Cost Function

The F/BW approach assumes upward-sloping bilateral export supply curves to estimate bilateral import demand elasticities (σ) for various industries. Starting with Feenstra (1994), the bilateral supply curve for j's imports from country i was specified as $p_{ij} = q_{ij}^{\omega} \xi_{ij}$, where q_{ij} is the quantity produced in i and exported to j, and ξ_{ij} was assumed to be a random (technology) factor (cf., his equation (8)).[2] More than twenty years later, Soderbery (2018) also assumed the same positively sloped export supply curve stating "an upward sloping constant elasticity (bilateral) export supply curve of this nature was championed by Feenstra (1994), and has become standard with Broda and Weinstein (2006) and Broda et al. (2008) for structurally estimating (bilateral) import demand and export supply elasticities. Additionally, recent deviations from Feenstra (1994) by Feenstra and Weinstein (2017) and Hottman et al. (2016) model a tighter link between exporter cost functions and export supply, but effectively assume that (bilateral) export supply is isoelastic and upward sloping" (p. 47). The "tighter link" that Soderbery (2018) refers to is Feenstra and Weinstein (2017) specifying that "marginal costs from each exporting country" to an importing country are an exponential function $mc_{ij} = \omega_{ij0} q_{ij}^{\omega}$, where ω_{ij0} is an undefined term.[3]

We note two issues in the extant literature. First, because the studies cited focus on demand considerations, they do not provide a micro-foundation for the supply-side of the model. Second, these studies ignore the heterogeneity in firms' productivities that is now well documented. In this paper, we address both of these issues by extending the Melitz trade model to allow for fixed and variable marketing costs in the spirit of Arkolakis (2010), adapted to an empirically tractable framework.[4] The core idea put forward in Arkolakis (2010) is that "firms reach individual consumers rather than the market in its entirety" (p. 1152). Arkolakis (2010) introduced a variable cost component to the fixed export (marketing) component that yielded that only the most productive firms would enter a foreign market (selling to the "first consumer"), but to reach additional consumers (i.e., marginal market penetration) the firm faced "increasing *marginal* penetration costs" (p. 1151; italics added). The model in Arkolakis (2010) provides a rich extension of the Melitz model that matches empirical regularities in the data, such as the observation in Eaton et al. (2011) that the typical destination market for exporters has a large number of smaller firms. Specifically, Eaton et al. (2011) note that the size distribution of exporters within each destination market exhibits a Pareto distribution for relatively larger exporters, but they also note *deviations* from the Pareto distribution for a large proportion of French exporters in each market selling small amounts. Moreover, the rationale for introducing marginal marketing costs is supported empirically. As noted in Arkolakis (2010), Bagwell (2007) reviewed the literature on the economics of advertising and notes that most studies found that advertising's effectiveness

[2] We have modified his notation to be consistent with that of the current paper.

[3] Once again, we have modified notation in Feenstra and Weinstein (2017) to be consistent with that of the current paper; see that paper's equation (23) on page 1059. Also, Fajgelbaum et al. (2020) asume the same bilateral increasing marginal cost function.

[4] For simplicity here, we assume a single industry as in Melitz (2003). As common to the literature, we could instead have multiple industries with Cobb-Douglas preferences. Nevertheless, our estimation method recognizes that the structural parameters vary across industries.

is subject to diminishing returns.[5]

For our purposes, the specific features of the model in Arkolakis (2010) are constraining. First, by introducing variable marketing costs via the export fixed cost term, the model in Arkolakis (2010) yields a pricing function where price is a function of *constant* marginal costs, independent of destination output and consequently inconsistent with the typical F/BW positively sloped supply curve. Second, as Anderson (2011) pointed out, the marketing element in Arkolakis (2010) effectively has a "fixed-cost component and a variable-cost component *subject to diminishing returns*" (p. 140; italics added). Third, one of the benefits of Arkolakis embedding the variable cost marketing component inside export fixed costs is that – for his calibrations – he avoids having to specify "as many [export] fixed costs as destinations" (p. 1164). However, as Anderson (2011) noted, the introduction of numerous additional parameters is useful for his simulations, but "is not econometrically tractable" (p. 140). Consequently, we introduce in our model a simple explicit variable marketing cost in the production function, similar in spirit to iceberg transport costs, that captures increasing marginal market-penetration costs in an econometrically tractable manner consistent with the F/BW approach.

Let m_{ij} denote an *ad valorem* factor representing the additional output that must be produced by firms in country i to cover variable marketing costs of "marginal market penetration" from selling in country j, like iceberg trade costs. Hence, variable marketing costs are a function of the quantity sold within the destination market, $m_{ij}(q_{ij})$. However, unlike iceberg trade costs, variable marketing costs are an exponential function $m_{ij}(q_{ij}) = q_{ij}^{1/\gamma}$ where $0 < \gamma < \infty$ (and hence $0 < 1/\gamma < 1$), capturing that previous empirical studies noted above suggest that marketing expenditures exhibit diminishing returns to reach more consumers *within* market j.[6] Having defined all the components of costs, we can now introduce the cost function. Production uses only one input, labor. The labor required by a country-i firm with productivity φ to produce q_{ij} units of output for sale to country j is given by:

$$l_{ij}(\varphi) = \frac{1}{A_i}\left(f_{ij} + \frac{m_{ij}(q_{ij})q_{ij}}{\varphi}\right) = \frac{1}{A_i}\left(f_{ij} + \frac{q_{ij}(\varphi)^{1+\frac{1}{\gamma}}}{\varphi}\right) \tag{4}$$

where $A_i > 0$ is incorporated as an exogenous parameter which captures the productivity of workers in the entire country.[7] As implied by equation (4), the fixed costs component

[5] Arkolakis (2010) also notes several other studies supporting that advertising expenditures are subject to diminishing returns, cf., Simonovska and Waugh (1980), Saunders (1987), Sutton (1991), and Jones (1995).

[6] See Flach and Unger (2022), equation (5), for a similar formulation in the context of a model with quality differentiation.

[7] Countries with more productive workers (i.e., with higher A_i) require fewer workers, all else equal, to produce a given quantity of output or cover fixed costs. The special case of $i = j$ represents the demand for labor for domestic sales. As standard to this literature, for the domestic market, the fixed costs f_{ii} capture the costs of setting up a production facility, as well as advertising and domestic distribution costs. For foreign

(f_{ij}) is common across firms for a given origin-destination pair, whereas marginal costs vary across firms for two reasons.[8] First, as conventional to a Melitz model, more productive firms (i.e., with higher φ) need fewer workers to produce a given level of firm output.[9] Second, marginal costs are a function of destination output such that, all else equal, larger firms face higher marginal costs to reach more consumers in a market. The parameter γ determines the marginal cost elasticity of output. For any value of $\gamma \in (0, \infty)$, marginal costs are increasing. When γ goes to infinity, we obtain the constant marginal cost function in most workhorse trade models.[10]

As common in this literature, sales to foreign consumers are subject to iceberg trade costs. Firms in country i must ship $\tau_{ij} \geq 1$ units of output to sell one unit in destination j. As typical, we assume $\tau_{ij} > 1$ for all $i \neq j$ and $\tau_{ii} = 1$ for all i. As in Feenstra (2010), we let $p_{ij}(\varphi)$ and $q_{ij}(\varphi)$ denote the factory gate price and quantity shipped. Since a firm in country i producing for and selling to market j incurs *ad valorem* iceberg costs τ_{ij}, only $c_{ij} = q_{ij}/\tau_{ij}$ arrives at destination j. Moreover, drawing upon section 2.1, it follows that, for consumers in j, the unit price will be $p^c_{ij} = \tau_{ij} p_{ij}$.

2.3 Firm Behavior

Firms make two decisions for each potential market (including the domestic market). First, they must decide whether or not to enter the market. Second, for each market they enter, they must choose the sale price of a unit of output (or, equivalently, the quantity of output to sell). We look at each decision, beginning with the pricing one.

Firm profits in each market are given by revenues less labor costs:

$$\pi_{ij}(\varphi) = r_{ij}(\varphi) - w_i l_{ij}(\varphi) = p_{ij}(\varphi) q_{ij}(\varphi) - \frac{w_i}{A_i} \left[f_{ij} + \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \right], \tag{5}$$

markets ($i \neq j$), the fixed costs f_{ij} represent only the additional fixed costs of selling to the foreign market (such costs associated with advertising, distribution, and conforming to foreign regulations).

[8]In our model, we follow Bernard et al. (2011) in assuming that, for export fixed costs, domestic labor is employed. However, it is straightforward to consider instead the cases where labor in the foreign market is used as in Redding (2011) or labor from both countries is used as in ACR's equation (23). Naturally, this would have the associated implications for our results as discussed in ACR.

[9]We model higher productivity as producing a symmetric variety at lower marginal cost. However, higher productivity may also be thought of as producing a higher quality variety at equal cost. As noted in Melitz (2003), given the form of product differentiation, the modeling of either type of productivity difference is isomorphic.

[10]The cost function assumed here allows closed-form analytical solutions in a world with asymmetrically-sized countries and asymmetric bilateral trade costs. It is also feasible to follow instead Vannoorenberghe (2012) in a special case of symmetric country sizes and bilateral trade costs where marginal costs are simply increasing in total firm output; hence, Vannoorenberghe (2012) was the first to introduce increasing marginal costs in total firm output in a Melitz framework. We solve this case in Online Appendix C, noting that – with a large number of countries – the trade, extensive-margin, and intensive-margin elasticities are identical.

where the second equality uses cost function (4). Because each firm produces only one of a continuum of varieties, its pricing decision has no impact on the price index in the destination market (P_j). In other words, the structure of the model eliminates strategic interactions between firms. Firm profit maximization yields the following optimal (factory-gate) pricing rule:[11]

$$p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i q_{ij}(\varphi)^{\frac{1}{\gamma}}}{A_i \varphi}. \tag{6}$$

We note that all country-i firms with productivity φ will charge the same price in destination j, such that the price of varieties can be identified by an origin country and a firm productivity, $p_j(\nu) = p_{ij}(\varphi)$.

Pricing rule (6) differs from standard Melitz models in two respects. First, the markup is no longer a function of only the elasticity of substitution (σ), but also depends on the inverse marginal cost elasticity of output (γ). As a result, conditional on the distribution of firm productivities, prices will be higher by a factor of $1 + 1/\gamma$ under IMC. Second, prices are an increasing function of quantity; this provides a rationale for the upward-sloping bilateral export supply functions in F/BW. We note that, when γ goes to infinity, the first term of the pricing rule converges to 1 and quantity vanishes from the equation such that we obtain the CMC pricing rule typical to a standard Melitz model and most workhorse trade models.

Next, we consider the decision to enter a market or not. As a first step, we compute firm profits. We can use pricing rule (6) to express firm profits, defined in equation (5), as:

$$\pi_{ij}(\varphi) = \left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{r_{ij}(\varphi)}{\sigma} - \frac{w_i}{A_i} f_{ij} \tag{7}$$

where $r_{ij}(\varphi)$ is the firm's optimal revenue. This result is analogous to a standard Melitz model with the exception of the first term $(\sigma+\gamma)/(1+\gamma)$, which exceeds unity because $\sigma > 1$. Our model implies that profits are higher when marginal costs are increasing in output (i.e., $1/\gamma > 0$). Again, when γ goes to infinity, the benchmark result obtains.

We can combine the zero-cutoff-profit (ZCP) condition $\pi_{ij}(\varphi^*_{ij}) = 0$, the optimal pricing equation (6), and profits equation (7) to solve for the output and the productivity of the ZCP firm as follows:

$$q_{ij}(\varphi^*_{ij}) = \left[\frac{\gamma}{\sigma+\gamma}(\sigma-1) f_{ij} \varphi^*_{ij}\right]^{\frac{\gamma}{1+\gamma}}, \tag{8}$$

where φ^*_{ij} is the productivity level of the ZCP firm and:

$$(\varphi^*_{ij})^{-\theta} = \left[\frac{\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}\frac{w_i}{A_i}\right)^{\sigma}}{b_i^{1-\sigma} E_j P_j^{\sigma-1}}\right]^{\frac{-\theta}{\frac{\gamma}{1+\gamma}(\sigma-1)}} \left[\frac{\gamma}{\sigma+\gamma}(\sigma-1) f_{ij}\right]^{\frac{-\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}} \tau_{ij}^{-\theta\frac{1+\gamma}{\gamma}}. \tag{9}$$

[11] Detailed derivations are available in sections 1 and 2 of Online Appendix A.

Because $\gamma/(\sigma+\gamma)$ and $\gamma/(1+\gamma)$ in equation (8) are both positive and smaller than one, for a given φ_{ij}^* the level of output $q_{ij}(\varphi_{ij}^*)$ is smaller than in the CMC case. Equation (9) provides an explicit link between *ad valorem* variable trade costs (τ_{ij}) and a country-pair's export cutoff productivity (φ_{ij}^*).[12] Under CMC (i.e., $\gamma = \infty$), these two variables are proportionate. However, under IMC, a one percent change in τ_{ij} has a more-than-proportionate effect on φ_{ij}^*. We will show later that this implies the trade elasticity is larger under IMC relative to CMC. Finally, we note that when $\gamma \to \infty$, equations (8) and (9) simplify to the standard result in the benchmark CMC case.

Revenue is increasing in firm productivity, so that profits are also increasing in firm productivity. As a result, firms in country i with productivity above the productivity cutoff φ_{ij}^* will enter market j, while those with productivity below the cutoff will not. Furthermore, equation (9) implies that the ratio of export and domestic cutoff productivities is:

$$\frac{\varphi_{ij}^*}{\varphi_{ii}^*} = \left(\frac{E_i P_i^{\sigma-1} f_{ij}^{\frac{1+\gamma}{\sigma+\gamma}}}{E_j P_j^{\sigma-1} f_{ii}^{\frac{1+\gamma}{\sigma+\gamma}}} \right)^{\frac{1}{\sigma-1}\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{\frac{1+\gamma}{\gamma}} \equiv \Gamma_{ij} \quad \Rightarrow \quad \varphi_{ij}^* = \Gamma_{ij}\varphi_{ii}^*. \tag{10}$$

As in Bernard et al. (2011), we assume that $\Gamma_{ij} > 1, \forall\, i \neq j$ (see page 1284). In that case, only the most productive firms export, while intermediate productivity firms serve only the domestic market and the low productivity firms exit. The assumption that there are no "pure exporters" is consistent with the empirical literature on firms in international trade.[13]

2.4 Trade Flows

We can now characterize equilibrium aggregate trade flows.[14] Imposing the labor-market-clearing condition and assuming a Pareto distribution for firms' productivities, we can solve for the mass of incumbent firms in each country i that sell to each destination j:

$$M_{ij} = \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_i L_i}{\delta\theta f^e}(\varphi_{ij}^*)^{-\theta}. \tag{11}$$

In the case of $\gamma = \infty$, M_{ij} simplifies to the respective term in a standard Melitz model with Pareto distribution. Next, using pricing rule (6) and mass of firms equation (11), we can express bilateral trade flows as:

$$X_{ij} \equiv M_{ij}\int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = \left[\frac{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta - \frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right]\frac{w_i L_i f_{ij}}{\delta f^e}(\varphi_{ij}^*)^{-\theta}. \tag{12}$$

[12]Detailed derivations are available in section 3 of Online Appendix A.
[13]The findings in Lu (2010) to the contrary are explained in Dai et al. (2016) as processing trade.
[14]Derivation details are provided in sections 4–8 of Online Appendix A.

We use the goods-market-clearing condition, $R_i = E_i$, to express trade flows as a gravity equation. Substituting equation (12) into expenditure function $E_j = \sum_{k=1}^{N} X_{kj}$, using the definition of the productivity threshold in equation (9), and solving yields the following gravity equation:

$$X_{ij} = \left[\frac{A_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_k w_k^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}} \right] w_j L_j. \quad (13)$$

We note that when $\gamma \to \infty$ the benchmark result obtains.[15, 16]

2.5 General Equilibrium

In section 9 of Online Appendix A, we develop the dynamic aspect of the model and show that it is possible to define a set of free entry conditions that depend only on parameters and the productivity cutoffs. These conditions serve to identify equilibrium values for the productivity thresholds. As explained in section 10 of Online Appendix A, we can determine the general equilibrium using the recursive structure of the model as in Bernard et al. (2011).

3 Implications

In this section, we provide several theoretical implications from the model. In section 3.1, we derive novel *ad valorem* variable trade-cost and fixed trade-cost trade elasticities under IMC. With IMC, the variable trade-cost trade elasticity changes *relative* to the fixed trade-cost trade elasticity (relative to CMC), which has implications for estimating the relative welfare benefits of fixed trade-cost liberalizations relative to variable trade-cost liberalizations within deep trade agreements. In section 3.2, we show that under IMC the welfare effect of a change in trade costs is still measured by the change in the domestic trade share raised to the (negative of the) inverse of the (variable trade-cost) trade elasticity, as in ACR. However, the welfare effect is diminished for a given domestic trade share; we explain the source of this "welfare diminution effect."

[15] Note that the wage-rate elasticity is equivalent to that in Bernard et al. (2011) if one assumes $\gamma = \infty$, as we have followed their assumption of export fixed costs using the exporter's (i's) labor. By contrast, Redding (2011) assumes export fixed costs use the importer's (j's) labor. ACR's equation (23) allows either of those two cases; our setting is analogous to ACR in their case of $\mu = 1$. In the case of $\gamma = \infty$ and $\mu = 1$, our wage-rate elasticity is equivalent mathematically to ACR's.

[16] As shown in section 11 of Online Appendix A, equation (13) and the associated variable- and fixed-trade-cost trade elasticities are consistent also with a "structural gravity " representation that is common in the literature. As a result, the method developed in Head and Mayer (2014) to estimate the general equilibrium trade impacts (GETI) of changes in trade barriers remains applicable for us.

3.1 Trade Elasticities

As shown in section 12 of Online Appendix A, the (positively defined) *ad valorem* variable trade-cost trade elasticity (ε_τ) is:

$$\varepsilon_\tau \equiv -\frac{\partial X_{ij}}{\partial \tau_{ij}}\frac{\tau_{ij}}{X_{ij}} = -\left[\underbrace{-\theta\left(\frac{1+\gamma}{\gamma}\right)}_{\text{extensive}} + \underbrace{\frac{1+\gamma}{\sigma+\gamma}(1-\sigma)}_{\text{intensive}} + \underbrace{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}_{\text{compositional}}\right] = \theta\left(\frac{1+\gamma}{\gamma}\right). \quad (14)$$

Following Head and Mayer (2014), we decompose this trade elasticity into extensive margin, intensive margin, and compositional margin components.[17] The extensive- and intensive-margin components have the usual interpretations. The extensive margin elasticity is caused by changes in the mass of firms serving each market. The intensive margin elasticity is caused by changes in firm-level exports.[18] The compositional-margin elasticity is caused by the fact that new entrants or exitors do not have the same productivity as the existing exporters. This margin is a function of the difference between the average shipment of the incumbent firms (X_{ij}/M_{ij}) and that of the marginal firm. All three components converge to the benchmark Melitz model values as $\gamma \to \infty$.

In line with previous results for Melitz models, the trade elasticity is determined entirely by the extensive margin elasticity. At the intensive margin, lower *ad valorem* trade costs increase exports of a given firm to a given country, which raise average exports per firm. At the compositional margin, lower *ad valorem* trade costs induce low productivity firms to enter the export market, which lowers average exports per firm. With a Pareto productivity distribution, the intensive margin and compositional margin elasticities offset one another exactly.

Under IMC, the elasticity of trade with respect to *ad valorem* trade costs, ε_τ, depends on θ, as in the benchmark, but is scaled up by the additional term $\frac{1+\gamma}{\gamma}$. Whenever $\gamma < \infty$, the trade elasticity is magnified relative to the benchmark ($\gamma \to \infty$). The intuition can be traced back to equations (9) and (11). Equation (9) reveals that, with IMC, a fall in τ_{ij} has a magnified effect of $\frac{1+\gamma}{\gamma}$ on lowering the country-pair's export cutoff productivity. In light of equation (11), this lower export productivity threshold makes it profitable for more firms to export from i to j and hence M_{ij} increases, enlarging the aggregate trade flow from i

[17] We note that this decomposition nests other decompositions proposed in the literature. First, in the decomposition of Redding (2011), the intensive and compositional margins are lumped together and labeled as the "intensive margin." It also nests the decomposition proposed by Chaney (2008), which is obtained by taking the sum of the extensive and the compositional margins and calling it the "extensive margin."

[18] The intensive-margin elasticity here is consistent with that in a special case of Bergstrand (1985) with homogeneous firms. We address this in Online Appendix B.

to j. Due to diminishing marginal returns, the trade elasticity is augmented and is now a nonlinear function of the two supply-side parameters, θ and γ.

As shown in section 13 of Online Appendix A, we can also decompose the (positively defined) elasticity of trade with respect to fixed trade costs (ε_f) into three margins:

$$\varepsilon_f \equiv -\frac{\partial X_{ij}}{\partial f_{ij}}\frac{f_{ij}}{X_{ij}} = -\left[\underbrace{-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}_{\text{extensive}} + \underbrace{0}_{\text{intensive}} + \underbrace{1}_{\text{compositional}}\right] = \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1. \quad (15)$$

All components converge to the benchmark values as $\gamma \to \infty$. The fixed trade-cost trade elasticity is also scaled up compared to the CMC case where $\varepsilon_f = \theta/(\sigma-1) - 1$.[19] An explanation for the different elasticity under IMC also can be traced intuitively back to equations (9) and (11). Using equation (9), with increasing marginal costs a fall in f_{ij} has a magnified effect on lowering the country-pair's export cutoff productivity relative to the case of CMC. In the IMC case, the scaling up of the numerator by $\frac{1+\gamma}{\gamma}$ and scaling down of the denominator of this elasticity by $\frac{1+\gamma}{\sigma+\gamma}$ augments the reduction in the country-pair's export productivity cutoff. Using equation (11), this lower export productivity threshold makes it profitable for more firms to export from i to j and hence M_{ij} increases, enlarging the aggregate trade flow from i to j.[20]

So far we have shown that, for given values of the structural parameters, the elasticities of trade are magnified under IMC. As a result, any trade-policy liberalization or transport-cost reduction that lowers bilateral *ad valorem* variable trade costs or fixed trade costs will have a *larger* impact on trade flows and consequently on the domestic expenditure share than in the CMC case. Moreover, equations (14) and (15) reveal not only that IMC increases both elasticities in absolute terms, but also the fixed trade-cost trade elasticity increases *relative* to the variable trade-cost trade elasticity. To understand why fixed trade-cost reductions have a relatively larger effect on trade than variable trade-cost reductions with IMC, consider

[19] In the CMC case, the assumption that $\frac{\theta}{\sigma-1} > 1$ is necessary to solve the Melitz model. However, some empirical researchers have found evidence that estimates of θ are often below estimates of $\sigma - 1$, violating a necessary assumption of this model, cf., Feenstra (2016), page 168. Our results in equation (15) shed light on this finding. Our Melitz model under IMC requires only that $\theta > \frac{\gamma}{\sigma+\gamma}(\sigma-1)$. Hence, θ can be less than $\sigma - 1$ as long as θ exceeds $\frac{\gamma}{\sigma+\gamma}(\sigma-1)$, where $0 < \frac{\gamma}{\sigma+\gamma} < 1$.

[20] In Online Appendix C, we solved the model for the case of increasing marginal costs in *total firm output* (instead of destination-specific output). In the case of marginal costs depending on total firm output, we must assume symmetric countries and symmetric trade costs to obtain closed-form solutions. Since overall output is endogenous to the set of countries to which firms export, one cannot solve the model analytically with asymmetric country sizes and asymmetric trade costs. Yet, in the symmetric world, we can solve for analogous trade elasticities. In fact, in Online Appendix C, we show that – when the number of countries is large – the variable-trade-cost trade elasticity and the fixed-trade-cost trade elasticity are *identical* to those in equations (14) and (15), respectively.

equations (8) and (9). The variable trade-cost trade elasticity in a Melitz model is determined by extensive margin effects solely; consistent with these equations, lower τ_{ij} increases trade exclusively by increasing the mass of firms exporting from i to j (as under Pareto, the intensive margin effect is offset perfectly by the compositional margin effect). Due to IMC, the trade elasticity scales up θ by $\frac{1+\gamma}{\gamma}$ due to diminishing marginal returns, cf., equation (9). By contrast, the fixed trade-cost trade elasticity is determined by the *ratio* of the extensive margin elasticity to the intensive margin elasticity. Recall, under CMC, reductions in τ_{ij} change φ^*_{ij} proportionately; however, reductions in f_{ij} change φ^*_{ij} less than proportionately (i.e., φ^*_{ij} is proportionate to $f_{ij}^{1/(\sigma-1)}$ in equation (8)). The introduction of IMC causes both the variable trade-cost trade elasticity to increase from θ to $\theta\frac{1+\gamma}{\gamma}$, but also the intensive margin effect to decline from $\sigma-1$ to $\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)$. This is confirmed in equation (9). As we will show later, this result is important for evaluating the relative trade and welfare benefits of "shallow" trade agreements (that only lower variable trade costs) with those of "deep" trade agreements (that also reduce fixed trade costs).

3.2 Welfare

In this section, we show two results related to welfare effects under IMC relative to CMC. First, we show that under IMC the two sufficient statistics to measure the welfare effects of international trade-cost shocks remain the share of domestic expenditure on domestic output and the trade elasticity as in ACR. Second, because for a set of parameter values the trade elasticity is magnified under IMC relative to that under CMC, the predicted welfare gains from trade are smaller.

First, in section 14 of Online Appendix A, we show that the change in welfare of a given "foreign" shock (to τ_{ij} or f_{ij}) that leaves unchanged country j's labor endowment, L_j, as well as the costs to serve its own market (τ_{jj} and f_{jj}) can be expressed as:

$$\hat{W}_j = \hat{\lambda}_{jj}^{-1/[\theta(1+\frac{1}{\gamma})]} = \hat{\lambda}_{jj}^{-1/\varepsilon_\tau}, \tag{16}$$

where $\hat{\lambda}_{jj} \equiv \lambda'_{jj}/\lambda_{jj}$ is the (gross) change in the share of domestic expenditure (where $\lambda_{jj} = X_{jj}/E_j$) and $\hat{W}_j \equiv W'_j/W_j$ is the change in welfare. In the special case of a move from trade (λ_{jj}) to autarky ($\lambda'_{jj}=1$), the gains from trade (G_j) can be expressed as:

$$G_j = 1 - \lambda_{jj}^{1/[\theta(1+\frac{1}{\gamma})]} = 1 - \lambda_{jj}^{1/\varepsilon_\tau}, \tag{17}$$

which is identical to equation (12) in Costinot and Rodriguez-Clare (2014). These results imply that, conditional on the trade elasticity, the impact of trade shocks on welfare are independent of the structure of marginal costs. At the same time, note that the definition of the trade elasticity itself is different in our model. In the presence of IMC, the larger trade elasticity implies (for a given λ_{jj}) a smaller welfare effect than in the constant marginal cost case, which we will term in this paper the "welfare diminution effect."

Second, to understand intuitively this welfare diminution effect, consider the benchmark Melitz model with CMC. The change in welfare ($\hat{W}_j$) from a reduction in variable trade costs is directly proportionate to the change in average productivity ($\hat{\tilde{\varphi}}_{ij}$) and the change in the number of varieties ($\hat{M}_{ij}$), cf., Melitz (2003), equation (17). Feenstra (2010) shows also that the change in welfare can be simplified further to be proportionate to the change in output of the ZCP firm ($q_{ij}(\hat{\varphi}^*_{ij})$) (see his page 52). As seen in equation (8), under IMC the output of the cutoff productivity firm is proportional to the cutoff productivity:

$$q_{ij}(\varphi^*_{ij}) \propto \left(\varphi^*_{ij}\right)^{\frac{\gamma}{1+\gamma}} \tag{18}$$

due to diminishing marginal returns. In the limit, as γ approaches ∞, the relationship between $q_{ij}(\varphi^*_{ij})$ and φ^*_{ij} becomes linear, as in the benchmark Melitz model. As a result, a given change in φ^*_{ij} has a smaller effect on output under IMC than CMC. This is the intuition underlying the "welfare diminution effect" from increasing marginal costs.[21]

Finally, it will be useful to summarize in a table the differences between the various "trade" elasticities and welfare-change effects of our model relative to those of the main models in the trade literature. Adapting Table 3.1 in Head and Mayer (2014), Table 1 contrasts the *ad valorem* variable trade-cost intensive margin elasticities, *ad valorem* variable trade-cost trade elasticities, fixed trade-cost trade elasticities, and welfare effects from the large class of models addressed in Arkolakis et al. (2012) with those from this paper.

4 Estimation Methodology, Specifications, and Data

In order to conduct numerical analyses of the welfare gains from fixed- versus variable-trade-cost changes under increasing versus constant marginal costs in section 6, we need to estimate all three main structural parameters of the model: σ, γ, and θ.[22] To do so, in this section we introduce a two-pronged estimation method that consists of two reduced-form equations, both derived from our theoretical model. As is well known, properly specified

[21] We formalize this intuition using the constant-elasticity-of-transformation (CET) approach of Feenstra (2010) in sections 14 and 15 of Online Appendix A.

[22] In recent work, Fajgelbaum et al. (2020) use a setup similar to Feenstra (1994) to estimate both the bilateral import demand and the bilateral export supply elasticities using disaggregated trade data. Their approach is quite different from ours. First, they do not include firm heterogeneity in their theoretical framework; hence, estimating equations differ across the two studies. Second, they identify both elasticities using a single instrumental variable, tariff rates. Third, they estimate one demand parameter and one supply parameter common to all industries; by contrast, we estimate *hundreds* of industry-specific demand and supply parameters.

TABLE 1
ELASTICITIES AND WELFARE MEASURES BY MODEL

Model	Intensive margin elast.	Var. trade elast. (ε_τ)	Fixed trade elast. (ε_f)	Welfare measure
Armington differentiation (Anderson, 1979)	$\sigma - 1$	$\sigma - 1$	n.a.	$\hat{\lambda}_{jj}^{-\frac{1}{\sigma-1}}$
Armington differentiation and CET (Bergstrand, 1985)	$\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)$	$\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)$	n.a.	$\hat{\lambda}_{jj}^{-\frac{1}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}$
Monopolistic Competition (Krugman, 1980)	$\sigma - 1$	$\sigma - 1$	n.a.	$\hat{\lambda}_{jj}^{-\frac{1}{\sigma-1}}$
Heterogeneity without fixed trade costs (Eaton-Kortum, 2002)	n.a.	θ	n.a.	$\hat{\lambda}_{jj}^{-\frac{1}{\theta}}$
Heterogeneity with fixed trade costs and Pareto (Chaney, 2008)	$\sigma - 1$	θ	$\frac{\theta}{\sigma-1} - 1$	$\hat{\lambda}_{jj}^{-\frac{1}{\theta}}$
Heterogeneity with fixed trade costs, Pareto, and IMC (BCG)	$\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)$	$\theta\left(\frac{1+\gamma}{\gamma}\right)$	$\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1$	$\hat{\lambda}_{jj}^{-\frac{1}{\theta\left(\frac{1+\gamma}{\gamma}\right)}}$

Notes: This table reports the (positively-defined) *ad valorem* variable trade-cost intensive margin elasticity, the *ad valorem* variable trade-cost trade elasticity, the fixed trade-cost trade elasticity, and the measure of welfare effects, under various theoretical assumptions as indicated in the first column's papers. The trade and intensive margin elasticities reported here for Bergstrand (1985) assume the case in that paper of $\sigma = \mu$ and $\gamma = \eta$. See Online Appendix B for an explanation; CET denotes constant elasticity of transformation. n.a. denotes not applicable.

and estimated gravity equations can generate estimates of the (variable trade-cost) trade elasticity, ε_τ. In our model, this elasticity is the product of θ and $\frac{1+\gamma}{\gamma}$. Consequently, to identify θ, we generate estimates of γ and σ by estimating an extension of the F/BW reduced form equation. Because of firm heterogeneity, our reduced form equation depends upon a large number of variables not included in the standard F/BW estimating equation.[23] Using the gravity equation implied by our theoretical model to estimate the trade elasticity and an estimate of γ from our F/BW reduced form, we can recover an estimate of θ.[24]

[23] Although the coefficients of our reduced-form extension of F/BW depend only on the three parameters of the model (σ, γ, and θ), the large number of non-linear restrictions precludes identification of θ from that regression. We return to this issue later.

[24] Due to our goal here of providing a novel methodological approach to estimate all three parameters (σ, γ, and θ) under our modified Melitz model framework, we omit allowing for heterogeneous bilateral export supply elasticities – across exporter-importer pairs – as addressed recently in Soderbery (2018) and Farrokhi and Soderbery (2020). Soderbery (2018), though still relying upon the same assumed bilateral export supply function as in the F/BW models, moves the literature in a different direction from our paper by exploring how heterogeneous (by exporter-importer pair) bilateral supply elasticities can help explain importers' market power and be adapted to evaluate optimal trade policy. Farrokhi and Soderbery (2020) extend Soderbery (2018) further. Their section 3 shows that the F/BW approach is a restricted version of a more general model allowing external economies of scale and labor mobility across industries. Specifically, they argue that the F/BW approach constrains the bilateral export supply elasticities to have positive slopes and assumes demand is not "convoluted by supply when using unit values." Three other studies examine the implications of increasing returns to scale *external to the industry* in extensions of the new quantitative trade models, cf., Kucheryavyy et al. (2019), Lashkaripour and Lugovskyy (2019), and Bartelme et al. (2019). Extending our paper in these directions far exceeds the scope of our paper.

In section 4.1, we derive our extension of the F/BW reduced-form estimating equation accounting for firm heterogeneity. We begin by summarizing the key aspects of the F/BW methodology in section 4.1.1. In section 4.1.2, we derive the bilateral import demand (structural) equation from our model with firm heterogeneity. F/BW assume the existence of a positively sloped bilateral export supply curve and assume exogenous demand and supply "shocks" to derive a reduced-form equation. By contrast, in section 4.1.3 we derive the (inverse) bilateral export-supply (structural) equation from our theoretical model. In section 4.1.4, we show that F/BW corresponds to a special case of our model and we discuss specifications and data requirements. In section 4.1.5, we discuss the moment and identification conditions in the context of our theoretically-based extended model. In section 4.2, we develop the other reduced-form estimating equation based on the gravity equation implied by our model to estimate the trade elasticity, which we will then use to generate estimates of θ. While we implement the model using data for hundreds of industries (as discussed in section 4.1.4), we omit the industry subscripts in what follows to simplify notation; hence, the derivations are for any industry.

4.1 F/BW Estimation Methodology Accounting for Firm Heterogeneity

In this section, we derive a (structural) nominal bilateral import-demand-share (X_{ij}^D/E_j) equation that motivates an estimable bilateral trade-flow-share equation, and a nominal bilateral export-supply-share (X_{ij}^S/E_j) equation that motivates an estimable bilateral import-unit-value equation, akin to F/BW. We will show that two error terms surface in these equations; one error term accounts for the role of deviations from the Pareto assumption for productivities (for small exporters) addressed in Arkolakis (2010) influencing variables determining the bilateral trade-share (demand) equation and the other error term accounts for the role of deviations from the Pareto assumption for productivities influencing variables determining the bilateral import unit-value (supply) equation. We then derive the reduced-form estimating equation that controls *explicitly* for firm heterogeneity, and we demonstrate that both the moment and identification conditions addressed in F/BW are satisfied.

4.1.1 The Basic F/BW Approach

To understand our contribution, we first provide a brief summary of the F/BW methodology. The F/BW approach entails a bilateral nominal import-demand-share equation:

$$\Delta^k \ln(X_{ijt}^D/E_{jt}) = (1-\sigma)\Delta^k \ln \overline{p}_{ijt}^c + \epsilon_{ijt} \tag{19}$$

where $\overline{p}_{ijt}^c$ is the observed bilateral import unit value, $t = 1, ..., T$ indexes time periods, $\Delta^k \ln$ refers to the *double difference* of a variable with respect to both time and a "reference" exporting country k, e.g., $\Delta^k \ln \overline{p}^c_{ijt} = (\ln \overline{p^c_{ijt}} - \ln \overline{p}^c_{ij,t-1}) - (\ln \overline{p}^c_{kjt} - \ln \overline{p}^c_{kj,t-1})$, and ϵ_{ijt} is an error term that will be discussed shortly.

Rather than estimating the demand equation using instrumental variables to address simultaneity, F/BW introduce an *ad hoc* "supply" equation and rely on orthogonal supply shocks. Their method proceeds in three steps. First, F/BW assume monopolistically competitive firms face upward sloping bilateral export supply to each market, implying a (inverse supply) function in terms of a nominal bilateral export-supply share (X^S_{ijt}/E_{jt}):

$$\Delta^k \ln \overline{p}^c_{ijt} = \frac{1}{1+\gamma} \Delta^k \ln(X^S_{ijt}/E_{jt}) + \psi_{ijt} \tag{20}$$

where ψ_{ijt} is an error term that will be discussed shortly. Second, F/BW combine these demand and supply equations in a particular manner. They rewrite equations (19) and (20) with ϵ_{ijt} and ψ_{ijt}, respectively, on the LHS, take the latter two terms' product, and rearrange terms to obtain:

$$\begin{aligned}(\Delta^k \ln \overline{p}^c_{ijt})^2 = \frac{1}{(\sigma-1)(1+\gamma)} \left(\Delta^k \ln s_{ijt}\right)^2 \\ + \frac{\sigma-\gamma-2}{(\sigma-1)(1+\gamma)} \left(\Delta^k \ln s_{ijt} \Delta^k \ln \overline{p}^c_{ijt}\right) + \epsilon_{ijt}\psi_{ijt},\end{aligned} \tag{21}$$

where s_{ijt} denotes the (partial equilibrium) trade share, but is measured using the actual bilateral trade share. Third, under the assumption that the demand and supply error terms are orthogonal, F/BW use the moment condition $\mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = 0$ (where $\mathbb{E}$ denotes the expectation operator) to derive a reduced-form equation, averaging each of the variables over all T observations. Letting $\overline{Y}_{ij}$, $\overline{Z}_{1ij}$, $\overline{Z}_{2ij}$, and $\overline{\epsilon_{ij}\psi_{ij}}$ denote the time-averaged means of the respective variables in equation (21), consistent estimates of the coefficients are obtained by estimating:

$$\overline{Y}_{ij} = \beta_0 + \beta_1 \overline{Z}_{1ij} + \beta_2 \overline{Z}_{2ij} + \overline{\epsilon_{ij}\psi_{ij}} \tag{22}$$

separately for each industry. Because of the double difference, the empirical model identifies the coefficients from the second moments of the data (i.e., variances and covariances). Identification therefore relies on the presence of heteroskedasticity such that $\overline{Z}_{1ij}$ and $\overline{Z}_{2ij}$ are not perfectly collinear, cf., Feenstra (1994), page 164.

In the remaining subsections of section 4.1, we show first that our model delivers both the bilateral trade-flow-share and bilateral import-unit-value analogue equations to those in F/BW, but based upon micro-foundations from our general equilibrium model of trade. Second, we show that the moment condition requires the inclusion of additional controls suggested by our theory. Third, in the context of our general equilibrium framework, the model calls for a reinterpretation of the error terms used for identification of the coefficients.

4.1.2 Bilateral Import Demand

In this section and the next one, for brevity we omit the time subscript, t (as well as the industry subscript, as earlier); in section 4.1.4, we reintroduce the time subscript. Given product-level bilateral import demand equation (2), aggregate bilateral import demand (for each industry), Q_{ij}^D, is:

$$Q_{ij}^D = M_{ij} \int_{\varphi_{ij}^*}^{\infty} c_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = M_{ij} b_i^{1-\sigma} E_j P_j^{\sigma-1} \int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{-\sigma}\mu_{ij}(\varphi)d\varphi, \tag{23}$$

and the value of aggregate bilateral import demand, X_{ij}^D, is:

$$X_{ij}^D = M_{ij} \int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi) c_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = M_{ij} b_i^{1-\sigma} E_j P_j^{\sigma-1} \int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi. \tag{24}$$

Because the two integrals over variety-level prices in the last two equations are not observable, we cannot use equations (23) or (24) to estimate the parameters of the model. Instead, as in F/BW, we need to rely on *observed* bilateral import unit values, $\overline{p}_{ij}^c$, defined as the ratio of bilateral import value to bilateral import quantity. From equations (23) and (24), the analytical expression for this ratio is:

$$\overline{p}_{ij}^c \equiv \frac{X_{ij}^D}{Q_{ij}^D} = \frac{\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi}{\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{-\sigma}\mu_{ij}(\varphi)d\varphi} \tag{25}$$

As this last result makes clear, bilateral import unit values are non-linear functions of variety-level prices. In the remainder of this section, we use the theoretical model to obtain analytical expressions for each of the unobserved price integrals in equation (25). Using these results, we will then be able to express bilateral trade (in shares) as functions of observable bilateral import unit values (and additional variables).

We proceed in three steps. First, in Online Appendix D, we show that the numerator and denominator in (25) are both proportional to a ZCP price, $p_{ij}(\varphi_{ij}^*)$, as follows:

$$\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right] [p_{ij}^c(\varphi_{ij}^*)]^{1-\sigma} e_{ij}^{P1}, \tag{26}$$

$$\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{-\sigma}\mu_{ij}(\varphi)d\varphi = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right] [p_{ij}^c(\varphi_{ij}^*)]^{-\sigma} e_{ij}^{P2}, \tag{27}$$

where error terms e_{ij}^{P1} and e_{ij}^{P2} capture deviations between the theoretical Pareto distributions and the actual distributions of productivities. Recall in section 2, we introduced a Pareto distribution for heterogeneous productivities of firms in order to obtain *closed-form* solutions, as common to theoretical Melitz models. As shown in Arkolakis (2010) and Eaton et al. (2011), empirical evidence suggests that the Pareto distribution does not approximate firms' sales distribution very well for very small exporters, with heterogeneity in this effect across country-pairs. To account for these deviations from the Pareto assumption on the distribution of productivities in an empirically tractable manner, we introduce, for example, a multiplicative error term e_{ij}^{P1} such that the empirical counterpart to $\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi$ is $e_{ij}^{P1}\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi$. When $e_{ij}^{P1}=1$, the distribution is Pareto precisely; for $e_{ij}^{P}\neq 1$, the distribution is only approximately Pareto.[25] As we explain below, the deviations from the Pareto distribution will play a central role in the identification of the structural parameters of the model.

Second, we use equations (26) and (27) to solve for the price integral in equation (24). We begin by substituting with (26) and (27) in equation (25) and rearrange to obtain an expression for the ZCP price as a function of the corresponding observed average unit value:

$$p_{ij}^c(\varphi_{ij}^*) = \left[\frac{\theta(\sigma+\gamma)-\gamma(\sigma-1)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]\overline{p}_{ij}^c\left(\frac{e_{ij}^{P2}}{e_{ij}^{P1}}\right). \tag{28}$$

Third, we can combine the results in equations (24), (26) and (28) – after first substituting equation (9) to replace the productivity threshold φ_{ij}^* and an extended version of equation (11) to allow for deviations (e_{ij}^{P3}) from Pareto for the endogenous mass of firms M_{ij} – to express the share of aggregate nominal bilateral trade flow in total expenditures (s_{ij}) as a function of bilateral import unit value:

$$s_{ij} \equiv \frac{X_{ij}^D}{E_j} = k_3 A_i^{1+\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i w_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}$$
$$\times E_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} P_j^{(\sigma-1)+\theta\left(\frac{1+\gamma}{\gamma}\right)} (\overline{p}_{ij}^c)^{1-\sigma} e_{ij}^D \tag{29}$$

where k_3 is a constant (defined in Online Appendix D) that depends only on the structural parameters σ, γ, θ, δ and f^e and $e_{ij}^D \equiv e_{ij}^{P1}\left(e_{ij}^{P2}/e_{ij}^{P1}\right)^{1-\sigma} e_{ij}^{P3}$. Note the inclusion in e_{ij}^D of Pareto deviation component e_{ij}^{P3}. This term captures the influences of deviations from Pareto associated with the determination of the mass of firms, M_{ij}, and is determined by two components, one of which (e_{ij}^{P4}) surfaces because $1-G(\varphi_{ij}^*)=(\varphi_{ij}^*)^{-\theta}e_{ij}^{P4}$ once one allows deviations from Pareto. e_{ij}^{P4} is important for e_{ij}^{P3}, and hence e_{ij}^D, because of its particular influence on small exporters that tend to be near the cutoff productivity, consistent with the

[25] We have re-derived all the results in Online Appendix A accounting for the deviations from Pareto. In fact, we uncovered five different sources of deviations from Pareto due to the complexity of the model ($e_{ij}^{P1}, ..., e_{ij}^{P5}$). Moreover, we show in Online Appendix D that one such deviation from Pareto (e_{ij}^{P4}) contributes specifically to deviations of firms at the cutoff productivity level (φ_{ij}^*), which tend to be smaller exporters.

evidence that deviations from Pareto tend to surface for small exporters. The superscript D in e_{ij}^{D} refers to the role of deviations from Pareto on the "demand" side (s_{ij}).[26]

4.1.3 Bilateral Export Supply

We now turn our attention to the bilateral export supply equation. We can invert the optimal pricing function (6) to get an analytical expression for output as a function of the price. Using the result, we can define average bilateral export supply (in physical units) as:

$$\overline{q}_{ij}^{S} \equiv \int_{\varphi_{ij}^{*}}^{\infty} q_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = \left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_i}{w_i}\right]^{\gamma}\int_{\varphi_{ij}^{*}}^{\infty}\left[\varphi p_{ij}(\varphi)\right]^{\gamma}\mu_{ij}(\varphi)d\varphi. \tag{30}$$

Defining industry bilateral export supply (in physical units) as $Q_{ij}^{S} \equiv M_{ij}\overline{q}_{ij}^{S}$, using equation (30) yields any industry's bilateral export supply:

$$\begin{aligned} Q_{ij}^{S} &= M_{ij}\int_{\varphi_{ij}^{*}}^{\infty} q_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi \\ &= M_{ij}\left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_i}{w_i}\right]^{\gamma}\int_{\varphi_{ij}^{*}}^{\infty}\left[\varphi p_{ij}(\varphi)\right]^{\gamma}\mu_{ij}(\varphi)d\varphi. \end{aligned} \tag{31}$$

Because the integral over firm-level prices and productivity is not observable, we need to solve for it as a function of the observed bilateral import unit value.

Similar to the import demand equation, we proceed in several steps. First, we solve for the integral of firm-level prices as a function of the ZCP firm's productivity level and price. As shown in Online Appendix D (section D.1), we can use optimal demand equation (2), the optimal pricing rule (6), and equation (8) for the output of the ZCP firm to derive an optimal pricing equation that is a function of $f_{ij}, \varphi_{ij}^{*}, w_i/A_i$, and φ (equation (D.4) in Online Appendix D). Substituting this optimal price equation into the price integral in equation (31), assuming our Pareto distribution allowing for deviations e_{ij}^{P5}, and solving yields:

$$\int_{\varphi_{ij}^{*}}^{\infty}\left[\varphi p_{ij}(\varphi)\right]^{\gamma}\mu_{ij}(\varphi)d\varphi = \frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\left[\varphi_{ij}^{*}p_{ij}(\varphi_{ij})^{*}\right]^{\gamma}e_{ij}^{P5}. \tag{32}$$

[26] In reality, X_{ij}^{D} on the LHS of equation (29) is unobservable. Beginning with Feenstra (1994), empirical implementation has used actual industry-level bilateral trade flows, presumably reflecting (bilateral) partial equilibrium, i.e., $X_{ij}^{D} = X_{ij}^{S}$. This literature has not incorporated the general equilibrium considerations addressed in sections 2.4 and 2.5. However, the theoretical trade flows are determined under equilibrium conditions; in particular, the framework in sections 2.4 and 2.5 assumes goods-market clearing $R_i = E_i$, i.e., multilateral trade balances. Observed trade flows may be influenced by deviations from multilateral trade balances; in reality, multilateral trade imbalances exist at the aggregate level and at the industry level. In order to allow for the fact that actual trade flows are not likely to equal equilibrium trade flows, we *could* allow theoretical trade-flow shares to deviate from actual trade-flow shares (labeled s_{ij}) by an error term e_{ij}^{X}. However, this additional error term is unnecessary to obtain identification for estimation; hence, for simplicity, we ignore it.

Second, we can use equation (28) and the fact that $p^c_{ij} = \tau_{ij} p_{ij}$ to define $p_{ij}(\varphi^*_{ij})$. Substituting with the result in equation (32) yields:

$$\int_{\varphi^*_{ij}}^{\infty} \left[\varphi p_{ij}(\varphi)\right]^{\gamma} \mu_{ij}(\varphi) d\varphi = \frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma} \left[\frac{\theta(\sigma+\gamma)-\gamma(\sigma-1)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]^{\gamma} \left(\overline{p}_{ij}\right)^{\gamma} \left(\varphi^*_{ij}\right)^{\gamma} e^{P5}_{ij}. \quad (33)$$

Third, substituting the RHS of equation (33) for the integral in equation (31) yields:

$$Q^S_{ij} = k_4 M_{ij} \left(\frac{A_i \overline{p}_{ij} \varphi^*_{ij}}{w_i}\right)^{\gamma} e^{P5}_{ij}, \quad (34)$$

where k_4 is a constant that depends only on the structural parameters σ, γ, and θ (defined in Online Appendix D). Solving for average price, we get:

$$\overline{p}_{ij} = k_4^{-\frac{1}{\gamma}} \left(\frac{Q^S_{ij}}{M_{ij}}\right)^{\frac{1}{\gamma}} \frac{w_i}{A_i \varphi^*_{ij}} \left(e^{P5}_{ij}\right)^{-\frac{1}{\gamma}}. \quad (35)$$

Fourth, we make the industry bilateral export-supply equation (35) comparable to the industry bilateral trade-flow-share equation by eliminating Q^S_{ij}, M_{ij}, and φ^*_{ij}. The value of industry bilateral export supply (X^S_{ij}) equals the value of bilateral import demand (X^D_{ij}), such that $M_{ij} \overline{p}_{ij} \overline{q}^S_{ij} = \frac{X^D_{ij}}{E_j} E_j$ or $Q^S_{ij} = \tau_{ij} \frac{X^D_{ij}}{E_j} \frac{E_j}{\overline{p}^c_{ij}}$ (recalling that $\overline{p}^c_{ij} = \tau_{ij} \overline{p}_{ij}$). Substituting this expression for Q^S_{ij} in equation (35), substituting for φ^*_{ij} using equation (9) and for M_{ij} using an extended version of equation (11) allowing deviations from Pareto, substituting s_{ij} for X^D_{ij}/E_j as in the previous section, and solving the resulting expression for $\overline{p}^c_{ij}$ yields:

$$\begin{aligned} \overline{p}^c_{ij} = k_5 A_i^{-1-\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i^{-\frac{1}{1+\gamma}} w_i^{\frac{\gamma}{1+\gamma}+\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{\left(\frac{\theta-\gamma}{\gamma}\right)\sigma} \tau_{ij}^{\frac{\theta-\gamma}{\gamma}} f_{ij}^{\frac{\frac{\theta-\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}} \\ \times E_j^{\frac{1}{1+\gamma}+\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} P_j^{-\left(\frac{\theta-\gamma}{\gamma}\right)\sigma} s_{ij}^{\frac{1}{1+\gamma}} e^S_{ij}, \end{aligned} \quad (36)$$

where k_5 is a constant that depends only on the parameters $\sigma, \gamma, \theta, \delta$, and f^e (defined in Online Appendix D) and $e^S_{ij} \equiv \left(e^{P3}_{ij} e^{P5}_{ij}\right)^{-\frac{1}{1+\gamma}}$. The superscript S in e^S_{ij} refers to the role of deviations from Pareto on the "supply" side ($\overline{p}^c_{ij}$).

4.1.4 Reduced-Form Specifications and Data Issues

Following F/BW, we eliminate time-invariant factors by first differencing structural equations (29) and (36), and then we eliminate importer-specific variables by taking a difference with respect to a reference exporting country k. From (29), we obtain the double-differenced trade-flow-share equation:

$$\Delta^k \ln s_{ijt} = (1-\sigma)\Delta^k \ln \overline{p}^c_{ijt} + \delta_{ijt}. \tag{37}$$

where we define a new term δ_{ijt} as:

$$\delta_{ijt} = \left[1+\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta^k \ln A_{it} + \Delta^k \ln L_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta^k \ln w_{it}$$
$$-\theta\left(\frac{1+\gamma}{\gamma}\right)\Delta^k \ln b_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta^k \ln \tau_{ijt} - \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}\Delta^k \ln f_{ijt} + \epsilon_{ijt}. \tag{38}$$

where (error term) $\epsilon_{ijt} \equiv \sigma\Delta^k \ln e^{P1}_{ijt} + (1-\sigma)\Delta^k \ln e^{P2}_{ijt} + \Delta^k \ln e^{P3}_{ijt}$. The general form of equation (37) corresponds to the benchmark F/BW structural bilateral trade-flow-share equation (19) in section 4.1.1 with three notable differences: (i) δ_{ijt} includes a host of additional variables (beyond $\Delta^k \ln \overline{p}^c_{ijt}$ in equation (19)); (ii) error term ϵ_{ijt} in equation (19) now has a clear interpretation; and (iii) some of the additional variables are unobservable (e.g., A_{it} and b_{it}).

First, in the context of our general equilibrium Melitz model, numerous determinants of the mass of varieties exported from i and the cutoff productivity need also to be accounted for in the trade-flow-share equation ($A_{it}, L_{it}, w_{it}, b_{it}, \tau_{ijt}$, and f_{ijt}). In their absence, coefficient estimates in benchmark F/BW reduced forms may be substantially biased (i.e., omitted variables bias, or OVB).

Second, the literature beginning with Feenstra (1994) has assumed that the error term in the basic F/BW structural trade-flow-share equation can be interpreted simply as a "taste shock." However, in the context of our general equilibrium framework, b_{it} is a determinant of the trade-flow share and so cannot represent the error term. By contrast, in our framework ϵ_{ijt} is driven by deviations from the Pareto distribution for productivities across country pairs that influence $\left[\int_{\varphi^*_{ij}}^{\infty} p^c_{ij}(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi\right]$, $\left[\int_{\varphi^*_{ij}}^{\infty} p^c_{ij}(\varphi)^{-\sigma}\mu_{ij}(\varphi)d\varphi\right]$, and M_{ij}. Note that $\Delta^k \ln e^{P1}_{ijt}$, $\Delta^k \ln e^{P2}_{ijt}$, and $\Delta^k \ln e^{P3}_{ijt}$ all have expected values of zero.

Third, several of the additional variables – A_{it}, L_{it}, w_{it}, and b_{it} – are exporter-specific variables but unobservable (A_{it}, b_{it}) or difficult-to-measure across countries and over time at the industry level (L_{it}, w_{it}). At the same time, as discussed below, the coefficients on these variables will not be relevant to estimating σ, γ, and θ in section 5 and conducting our counterfactual exercises in section 6. Consequently, we can hold constant the influences of these four variables by employing an exporter-year fixed effect, labeled α^1_{it}, allowing us to rewrite equation (38) more efficiently as:

$$\delta_{ijt} = \alpha_{it}^{1} - \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta^{k}\ln\tau_{ijt} - \frac{\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}\Delta^{k}\ln f_{ijt} + \epsilon_{ijt}. \tag{39}$$

We obtain the analogous double-difference supply-side equation from equation (36):

$$\Delta\ln\overline{p}_{ijt}^{c} = \left(\frac{1}{1+\gamma}\right)\Delta\ln s_{ijt} + \eta_{ijt}, \tag{40}$$

where we define η_{ijt} analogously as:

$$\eta_{ijt} = \alpha_{it}^{2} + \frac{\theta-\gamma}{\gamma}\Delta\ln\tau_{ijt} + \frac{\frac{\theta-\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}\Delta\ln f_{ijt} + \psi_{ijt} \tag{41}$$

where $\psi_{ijt} \equiv -\frac{1}{1+\gamma}\Delta^{k}\ln e_{ijt}^{P3} - \frac{1}{1+\gamma}\Delta^{k}\ln e_{ijt}^{P5}$.[27] The general form of equation (40) corresponds to the benchmark F/BW structural bilateral import unit value equation (20) in section 4.1.1 with the three notable differences analogous to equation (38): (i) η_{ijt} includes the same host of additional variables (beyond $\Delta^{k}\ln s_{ijt}$ in equation (20)); (ii) error term ψ_{ijt} now has a clear interpretation; and (iii) some of the additional variables are unobservable (e.g., A_{it}). First, if we ignore the additional covariates, the estimated coefficient will suffer from OVB. Second, the literature beginning with Feenstra (1994) has assumed that the error term in the structural bilateral import unit value equation can be interpreted simply as a "technology shock." However, our general equilibrium framework shows that A_{it} is a determinant of the bilateral import unit value and so cannot represent the error term. Third, several of these additional variables are exporter-specific and unobservable or difficult to measure, suggesting inclusion of an analogous exporter fixed effect term α_{ijt}^{2}.

It will be useful at this point to note that, in the trade literature, *ad valorem* variable trade costs τ_{ijt} typically reflect the product of gross tariff rates, labeled tar_{ijt}, and gross c.i.f.-f.o.b. transport-cost factors, labeled $trans_{ijt}$ (both of which are greater than 1). Consequently, in the following empirical specifications, we account for both components of variable trade costs separately.

Having motivated equations (37), (39), (40), and (41), we are now in a position to derive our reduced-form specifications, following in the spirit of F/BW and section 4.1.1. First, we can substitute equation (39) for δ_{ijt} in equation (37), and then solve for ϵ_{ijt} on the LHS. Second, we can substitute equation (41) for η_{ijt} in equation (40), and then solve for ψ_{ijt} on

[27] For brevity, we omit here the analogue to equation (38); refer to equation (36) for guidance on this expression.

the LHS. Third, we take the product of the two expressions and rearrange terms to yield:

$$Y_{ijt} = \sum_{k=1}^{20} \beta_k Z_{ijt,k} + \xi_{ijt}, \tag{42}$$

where

Group 1: $Y_{ijt} = \left(\Delta^k \ln \overline{p}^c_{ijt}\right)^2$, $Z_{ijt,1} = \left(\Delta^k \ln s_{ijt}\right)^2$, $Z_{ijt,2} = \Delta^k \ln s_{ijt} \Delta^k \ln \overline{p}^c_{ijt}$,

Group 2: $Z_{ijt,3} = \Delta^k \ln \overline{p}^c_{ijt} \Delta^k \ln tar_{ijt}$, $Z_{ijt,4} = \Delta^k \ln s_{ijt} \Delta^k \ln tar_{ijt}$, $Z_{ijt,5} = \left(\Delta^k \ln tar_{ijt}\right)^2$,
$Z_{ijt,6} = \Delta^k \ln \overline{p}^c_{ijt} \Delta^k \ln trans_{ijt}$, $Z_{ijt,7} = \Delta^k \ln s_{ijt} \Delta^k \ln trans_{ijt}$, $Z_{ijt,8} = \left(\Delta^k \ln trans_{ijt}\right)^2$,
$Z_{ijt,9} = \Delta^k \ln tar_{ijt} \Delta^k \ln trans_{ijt}$,

Group 3: $Z_{ijt,10} = \alpha_{it}$, $Z_{ijt,11} = \alpha_{it} \Delta^k \ln \overline{p}^c_{ijt}$, $Z_{ijt,12} = \alpha_{it} \Delta^k \ln s_{ijt}$,
$Z_{ijt,13} = \alpha_{it} \Delta^k \ln tar_{ijt}$, $Z_{ijt,14} = \alpha_{it} \Delta^k \ln trans_{ijt}$,

Group 4: $Z_{ijt,15} = \Delta^k \ln s_{ijt} \Delta^k \ln f_{ijt}$, $Z_{ijt,16} = \Delta^k \ln \overline{p}^c_{ijt} \Delta^k \ln f_{ijt}$, $Z_{ijt,17} = \Delta^k \ln tar_{ijt} \Delta^k \ln f_{ijt}$,
$Z_{ijt,18} = \Delta^k \ln trans_{ijt} \Delta^k \ln f_{ijt}$, $Z_{ijt,19} = (\Delta^k \ln f_{ijt})^2$, $Z_{ijt,20} = \alpha_{it} \Delta^k \ln f_{ijt}$,

where $\xi_{ijt} \equiv \epsilon_{ijt} \psi_{ijt}$ is a residual. We will explain shortly the relevance of the "Groups" for motivating the specifications.

The 20 β_k's are functions of only *three* structural parameters: σ, γ, and θ. The first two coefficients, β_1 and β_2, are defined exactly as in F/BW:

$$\beta_1 = \frac{1}{(1+\gamma)(\sigma-1)}, \quad \text{and} \quad \beta_2 = \frac{\sigma - \gamma - 2}{(1+\gamma)(\sigma-1)}. \tag{43}$$

Hence, we can use the same methodology as F/BW to back out structural parameters σ and γ from the reduced-form estimates of β_1 and β_2. Importantly, in the context of our Melitz (2003) model with firm heterogeneity and IMC, equation (42) makes it clear that estimation of the first two RHS variables will suffer from omitted variable bias (OVB) if variables in Groups 2-4 are not accounted for in the reduced-form specification.

Following F/BW, a consistent estimator of coefficients $\beta_1 - \beta_{20}$ can be obtained by averaging each of the variables in equation (42) over all $t = 1, ..., T$. Letting $\overline{Y}_{ij}, \overline{Z}_{1,ij}, ..., \overline{Z}_{2,ij}$, and $\overline{\xi}_{ij} = \overline{\epsilon_{ij} \psi_{ij}}$ denote the means, this yields the reduced form equation for estimation:

$$\overline{Y}_{ij} = \beta_0 + \sum_{k=1}^{20} \beta_k \overline{Z}_{k,ij} + \overline{\xi}_{ij}, \tag{44}$$

where the over-bar indicates that the variables are averages over time (e.g., $\overline{Z}_{ij} \equiv T^{-1} \sum_{t=1}^{T} Z_{ijt}$). In the remainder of this subsection, we describe the three specifications we estimate along with relevant data needs.

Specification 1: F/BW

As a benchmark, the first specification we estimate includes the three variables in Group 1 only. This is exactly the same specification as in F/BW. According to our model, the (reduced-form and structural) coefficient estimates will be biased because of omitted variables.

For estimation, we need data on trade flows in values and in quantities at the industry level. Data for trade flows come from the United Nations' Comtrade Database. This database collects bilateral f.o.b. export values that correspond to the transaction value of the goods, as well as bilateral c.i.f. import values which include the value of services performed to deliver goods to the border of the importing country. This database also contains information on the quantities exported and imported.[28] We combine the measures of trade flows and expenditures by industry to construct bilateral trade-flow shares and we combine measures of bilateral import values and quantities to construct bilateral import unit values. For our analysis, we define industries as four-digit Standard Industrial Trade Classification (SITC4) categories. Our sample covers the period from 1999 through 2010; after taking the time differences, we end up with years 2000-2010.

Specification 2: IMC-Partial

The second specification we estimate includes the variables in Groups 1 and 2. To generate a sense of the importance of the variables in Group 2 for correcting for OVB, we provide a stand-alone specification including the nine RHS variables. For illustrative purposes, we provide in Online Appendix D (section D.3) the derivations for the theoretical coefficients associated with this specification, labeled IMC-Partial. As explained earlier, the coefficients depend on only three structural parameters. However, the nine non-linear restrictions implied by the model prevent identification of all three parameters from this single reduced-form equation; specifically, the large number of restrictions preclude identification of θ. Nevertheless, the seven additional RHS variables are included to control for OVB, and estimates of σ and γ can *still* be determined from the estimates of β_1 and β_2.

Using the United Nations' Comtrade data discussed above, we construct *ad valorem* measures of gross transport costs factors $trans_{ijt}$ from the ratios of the c.i.f. to the f.o.b. unit values. Feenstra and Romalis (2014) provide a database of *ad valorem* tariff rates based upon Most-Favored-nation (MFN) status or any preferential status available, from which we construct tar_{ijt}.[29] The tariff rates are reported at the four-digit SITC level.

[28] When possible, we convert physical units of measurement to a common denominator (e.g., "Thousands of items" to "Items"). For industries with multiple units of measurement, we keep only the observations which report physical quantity in the unit of measurement that account for the largest value of import over the entire sample.

[29] This database combines information from the TRAINS data, the World Trade Organization's (WTO) Integrated Data Base, the International Customs Journal, and the texts of preferential trade agreements obtained from the WTO's website.

Specification 3: IMC-Full

The third specification will account fully for the variables in all four groups, and will be our preferred specification to address OVB. As with the IMC-Partial specification, the numerous non-linear restrictions preclude estimation of θ; however, we can still determine σ and γ from the estimates of β_1 and β_2. Later, with reduced-form gravity estimates of the trade elasticity, we will be able to determine θ. We discuss the motivation for this specification in two parts.

First, we address the Group 3 variables. A major benefit of specification IMC-Full is that the inclusion of the exporter fixed effects (α_i) and their interactions in Group 3 (alongside the Groups 1 and 2 variables) in the reduced-form equations of the time-averaged variables *eliminates* having to include measures of A_i, L_i, w_i, and b_i; recall that A_i and b_i are unobservable. Moreover, the inclusion of the exporter fixed effects and their interaction terms precludes including proxies for L_{it} and w_{it}, as the latter terms would create perfect multicollinearity. Nevertheless, for a robustness analysis, we will discuss later a specification including the variables in Groups 1 and 2 and crude proxies for L_i and w_i, but without exporter fixed effects and their interactions (which implies omitting controls for A_i and b_i).[30]

Second, we address the variables in Group 4, f_{ijt} and its interaction terms. While quality data exists on *ad valorem* tariff rates and transport costs, the international trade literature has so far struggled to construct and implement persuasive measures of bilateral fixed trade costs that affect only the decisions to export to a foreign market. To date, the most comprehensive effort to measure these fixed costs is the World Bank's *Doing Business* (DB) indicators. Covering a comprehensive swath of countries over multiple years, the DB indicators provide a widely respected quantification of the "ease of doing business" along numerous dimensions. However, unlike the theoretical variable, f_{ijt}, which is country-pair specific, the DB indicators are *country specific*.[31]

The World Bank also provides the *Deep Trade Agreements* database described in Hofmann et al. (2017) and Mattoo et al. (2020). This database is the first comprehensive source of information using dummy variables to indicate the presence or absence of each of 937

[30] In the robustness specification we will report later, we use per capita GDPs of countries as a proxy for w_{it}. Information on employment is not available at this level of detail. Instead, as a proxy for L_{it}, we obtain an estimate of employment. We follow Feenstra and Romalis (2014) and distribute employment across industries in proportion to export production. For each industry-country-year category, we measure employment as total employment multiplied by industry export value divided by GDP. Information on employment and GDP for each country-year is from the Penn World Tables (version 9.1).

[31] WorldBank (2020), Table 1.1 identifies 12 major country-specific categories of fixed costs that cover policy (artificial) and non-policy (natural) fixed costs associates with an importing country, ranging across base of "starting a new business, getting a location, accessing finance, dealing with day-to-day operations, and operating in a secure business environment." All such element influence the decision of potential exporter to enter a foreign market.

"deep" provisions within 219 preferential trade agreements (PTAs) between pairings of 189 countries annually from 1958-2017. Fortunately, Hofmann et al. (2017) identify so-called "core" provisions that dominate the DTAs. These core provisions are grouped in 16 "policy areas" (excluding industrial and agricultural tariff rates, common to 98 percent or of PTAs).[32] As noted in Hofmann et al. (2017), liberalization in all 16 policy areas can be categorized into multilateral (or non-discriminatory) and bilateral (or preferential) liberalization. Of the 16 areas, 12 are considered multilateral in nature and only four are considered bilateral. Furthermore, the four "core-WTO-X" policy-areas (competition policy, intellectual property rights, investment, and movement of capital) that are considered "important features of DTAs" are multilateral in nature, with almost 90 percent of PTAs including at least one of them. Finally, Hofmann et al. (2017) note that liberalizations of multilateral provisions are much more common, compared to liberalizations of bilateral provisions. From years 1980-84 to 2010-15, the average number of multilateral provisions has more than doubled from 4 to 9. By contrast, for the same 30 year period, the average number of bilateral provisions has increased by only 1, from 4 to 5. This all suggests that reductions in fixed trade costs due to deep trade agreements is largely captured by exporter-specific and importer-specific components of fixed trade costs.[33]

Consequently, our third specification, IMC-Full – including variables in Groups 1, 2, and 3 – can capture the influences of f_{ij} owing to the inclusion of an exporter fixed effect and exporter fixed effects interacted with $\overline{\Delta^k \ln \overline{p}^c_{ij}}, \overline{\Delta^k \ln s_{ij}}, \overline{\Delta^k \ln tar_{ij}}$, and $\overline{\Delta^k \ln trans_{ij}}$. The rationale is the following. Since the discussion above suggests that most of the variation in $\ln f_{ijt}$ can be explained across country-pairs and over time by variation in $\ln f_{it}$ and $\ln f_{jt}$ – treating the remaining variation as a residual, $\ln f^R_{ijt}$ – the introduction of an exporter fixed effect, alongside the differencing with respect to reference exporting country k (which removes importer effects, such as the influences of E_{jt} and P_{jt}), can account for most of the variation in $\ln f_{ijt}$, as long as the exporter-fixed-effect interactions are present. To see this, consider the variable $(\Delta^k \ln s_{ijt})(\Delta^k \ln f_{ijt})$ from equation (42). The differencing with respect to exporting reference country k removes the variation and influence of $\ln f_{jt}$, leaving variation in $\ln f_{it}$ and $\ln f^R_{ijt}$. Assume $\Delta^k \ln f^R_{ijt}$ is randomly distributed with mean 0 and variance $\sigma^2_{\Delta \ln f^R_{ijt}}$; we address this later. Suppose $\Delta^k \ln f_{it}$ follows a random walk with a drift

[32] The 16 policy areas are: competition policy, investment, movement of capital, intellectual property rights, customs (facilitation), technical barriers to trade, sanitary and photo-sanitary, state aid, GATS (services), TRIPS (intellectual property), state trading enterprises, TRIMS (investment measure state), export taxes, anti-dumping provisions, countervailing measures, and public procurement.

[33] Furthermore, studies such as Baier et al. (2014) show empirically the influence of time-invariant bilateral variables that affect fixed trade costs, such as bilateral distance. However, the time-differencing of our explanatory variables eliminates these variables' influences.

(Φ_i); hence, $\Delta^k \ln f_{it} = \Phi_i + u_{it}$. Substituting $\Phi_i + u_{it}$ for $\Delta \ln f_{it}$ yields:

$$(\Delta^k \ln s_{ijt})(\Delta^k \ln f_{ijt}) = (\Delta^k \ln s_{ijt})(\Phi_i + u_{it}) + (\Delta^k \ln s_{ijt})(\Delta^k \ln f^R_{ijt}). \tag{45}$$

Summing both sides of equation (45) over $t = 1, ..., T$ and dividing both sides by T yields:

$$\begin{aligned}\overline{(\Delta^k \ln s_{ij})(\Delta^k \ln f_{ij})} &= \frac{1}{T}\sum_{t=1}^{T}\Phi_i \Delta^k \ln s_{ijt} + \frac{1}{T}\sum_{t=1}^{T}(\Delta^k \ln s_{ijt})u_{it} + \frac{1}{T}\sum_{t=1}^{T}(\Delta^k \ln s_{ijt})(\Delta^k \ln f^R_{ijt}) \\ &= \Phi_i \overline{\Delta^k \ln s_{ij}} \end{aligned} \tag{46}$$

because the terms $\sum_{t=1}^{T}(\Delta^k \ln s_{ijt})u_{it}$ and $\sum_{t=1}^{T}(\Delta^k \ln s_{ijt})(\Delta^k \ln f^R_{ijt})$ are covariances and both covariances are zero.[34]

Finally, it is important to draw attention to the fact that ϵ_{ijt} is a linear function of $\Delta^k \ln e^{P1}_{ijt}, \Delta^k \ln e^{P2}_{ijt}$, and $\Delta^k \ln e^{P3}_{ijt}$, but ψ_{ijt} is a *different* linear function of $\Delta^k \ln e^{P3}_{ijt}$ and $\Delta^k \ln e^{P5}_{ijt}$. The necessary moment condition in this framework to employ the method-of-moments estimator is that $\mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = 0$. However, in the presence of an intercept in the estimating reduced form, $\mathbb{E}(\epsilon_{ijt}\psi_{ijt})$ need only be a constant; we will demonstrate this in the next subsection. Furthermore, identification requires that the relative variances (over time) of ϵ_{ijt} and ψ_{ijt} differ; we will demonstrate this in the next subsection as well.

4.1.5 Moment and Identification Conditions

Estimation of equation (44) produces consistent coefficient estimates under two conditions. The first is the "moment" condition, $\mathbb{E}(\xi_{ijt}) \equiv \mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = 0$. Recalling $\epsilon_{ijt} \equiv \sigma\Delta^k \ln e^{P1}_{ijt} + (1-\sigma)\Delta^k \ln e^{P2}_{ijt} + \Delta^k \ln e^{P3}_{ijt}$ and $\psi_{ijt} \equiv -\frac{1}{1+\gamma}\Delta^k \ln e^{P3}_{ijt} - \frac{1}{1+\gamma}\Delta^k \ln e^{P5}_{ijt}$, we show in Online Appendix D that:

$$\mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = -\left(\frac{1}{1+\gamma}\right)\mathrm{var}(\Delta^k \ln e^{P3}_{ijt}) = -4\left(\frac{1}{1+\gamma}\right)\mathrm{var}(\ln e^{P3}_{ijt}) \equiv -4\left(\frac{1}{1+\gamma}\right)\sigma^2_{\ln e^{P3}_{ij}} \tag{47}$$

is a constant, where $\sigma^2_{\ln e^{P3}_{ij}}$ denotes the variance over time of $\ln e^{P3}_{ijt}$.[35] Hence, the moment condition $\mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = 0$ is met as long as equation (44) includes an intercept, β_0 (as in Feenstra (1994)).

[34] Note that $\Delta \ln f^R_{ijt}$ needs to be accounted for also in the moment condition and in the identification condition to be discussed shortly below. Online Appendix D addresses how each of these conditions discussed in section 4.1.5 is altered in an inconsequential manner.

[35] Note that σ still denotes the elasticity of substitution in consumption whereas σ^2_z denotes the variance of the variable in the subscript (e.g., z).

The second is the "identification" condition, which requires that the relative variance of $\Delta^k \ln e_{ij}^D$ across country-pairs differs from the relative variance of $\Delta^k \ln e_{ij}^S$ across country-pairs. Modifying equation (12) in Feenstra (1994), identification in our context requires:

$$\frac{\sigma^2_{\epsilon_{ij}} + \sigma^2_{\epsilon_{kl}}}{\sigma^2_{\epsilon_{mn}} + \sigma^2_{\epsilon_{kl}}} \neq \frac{\sigma^2_{\psi_{ij}} + \sigma^2_{\psi_{kl}}}{\sigma^2_{\psi_{mn}} + \sigma^2_{\psi_{kl}}} \tag{48}$$

where σ_z^2, as above, denotes the variance over time of variable z. In terms of our model, the equivalent expression for identification is:

$$\frac{\sigma^2\left(\sigma^2_{\Delta^k \ln e^{P1}_{ij}}\right) + (1-\sigma)^2\left(\sigma^2_{\Delta^k \ln e^{P2}_{ij}}\right) + \left(\sigma^2_{\Delta^k \ln e^{P3}_{ij}}\right) + \sigma^2\left(\sigma^2_{\Delta^k \ln e^{P1}_{kl}}\right) + (1-\sigma)^2\left(\sigma^2_{\Delta^k \ln e^{P2}kl}\right) + \left(\sigma^2_{\Delta^k \ln e^{P3}_{kl}}\right)}{\sigma^2\left(\sigma^2_{\Delta^k \ln e^{P1}_{mn}}\right) + (1-\sigma)^2\left(\sigma^2_{\Delta^k \ln e^{P2}_{mn}}\right) + \left(\sigma^2_{\Delta^k \ln e^{P3}_{mn}}\right) + \sigma^2\left(\sigma^2_{\Delta^k \ln e^{P1}_{kl}}\right) + (1-\sigma)^2\left(\sigma^2_{\Delta^k \ln e^{P2}kl}\right) + \left(\sigma^2_{\Delta^k \ln e^{P3}_{kl}}\right)} \tag{49}$$

$$\neq \frac{\sigma^2_{\Delta^k \ln e^{P3}_{ij}} + \sigma^2_{\Delta^k \ln e^{P5}_{ij}} + \sigma^2_{\Delta^k \ln e^{P3}_{kl}} + \sigma^2_{\Delta^k \ln e^{P5}_{kl}}}{\sigma^2_{\Delta^k \ln e^{P3}_{mn}} + \sigma^2_{\Delta^k \ln e^{P5}_{mn}} + \sigma^2_{\Delta^k \ln e^{P3}_{kl}} + \sigma^2_{\Delta^k \ln e^{P5}_{kl}}}$$

where ij, kl, and mn denote different country-pairs.

The condition above requires that there must be *some differences* in the relative variances of the "demand" ($\Delta^k \ln e_{ij}^D$) and supply ($\Delta^k \ln e_{ij}^S$) disturbances. Although many factors can explain such differences, the key consideration is that the LHS and RHS of equation (49) include variances of time-differenced (as well as reference-exporting-country differenced) Pareto deviations of integrals over different variables. For instance, on the LHS $\Delta^k \ln e_{ij}^{P1}$ refers to the double-differenced deviations associated with the (demand-side) integral $\int_{\varphi^*_{ij}}^{\infty} [\tau_{ij} p_{ij}(\varphi)]^{1-\sigma} \mu_{ij}(\varphi) d\varphi$. By contrast, on the RHS $\Delta^k \ln e_{ij}^{P5}$ refers to the double-differenced deviations associated with the (supply-side) integral $\int_{\varphi^*_{ij}}^{\infty} [\varphi p_{ij}(\varphi)]^{\gamma} \mu_{ij}(\varphi) d\varphi$. Hence, the inequality condition (49) is likely to hold. It can be shown numerically that if the relative variances are different, condition (49) does hold.

4.2 Gravity Equation with Firm Heterogeneity

As discussed throughout section 4 so far, our extension of the F/BW framework precludes estimation of θ using the F/BW reduced-form equation, due to the non-linear restrictions issue raised above. The previous sections motivate estimation of σ and γ, but only under explicit controls for exporter masses and export productivity cutoffs to avoid omitted variables bias as shown by equations (42) and (44). However, our general equilibrium model suggests gravity-equation (13), as developed in sections 2.4 and 2.5. Consistent with the gravity-equation literature as summarized in Arkolakis et al. (2012), estimates of the "trade elasticity" – the elasticity of bilateral trade flows with respect to *ad valorem* variable trade costs – in the context of our model provide reduced-form estimates of $-\theta\left(\frac{1+\gamma}{\gamma}\right)$. Using the trade-elasticity estimates along with estimates of γ discussed above (by sector), industry-specific estimates of θ are readily determined. With all three structural parameters, numerical counterfactuals then can be performed in section 6.

We follow the econometric literature for estimating trade elasticities in the presence of panel data, where in our case we use industry-level nominal bilateral trade flows. Much of the recent literature on estimation of trade-policy effects using panel data in gravity equations follows Baier and Bergstrand (2007) and Baier et al. (2008, 2014, 2018). Consistent with these papers, the trade elasticity can be identified using a log-linear regression equation of bilateral trade flows (X_{ijt}) on exporter-year fixed effects, importer-year fixed effects, and a measure of *ad valorem* bilateral trade costs τ_{ijt}. Using equation (13), such a specification is:

$$X_{ijt} = \beta_0 + \phi_{it} + \Psi_{jt} - \theta\left(\frac{1+\gamma}{\gamma}\right)\ln\tau_{ijt} + \left[1 - \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}\right]\ln f_{ijt} + \vartheta_{ijt} \tag{50}$$

where ϕ_{it} are exporter-year fixed effects capturing the influences of A_{it}, L_{it}, w_{it}, and b_{it} in equation (13), Ψ_{jt} are importer-year fixed effects capturing the roles of L_{jt}, w_{jt}, and the (large) multilateral price/resistance term of the importer in the denominator of the first RHS term (in parentheses) in equation (13), and ϑ_{ijt} is an error term.

In estimating equation (50), three issues surface. The first is that τ_{ijt} is associated with both (gross) bilateral tariff rates, tar_{ijt}, and c.i.f.-f.o.b. transport-cost factors, $trans_{ijt}$. We introduce these variables separately. However, due to more confidence in the observed measures of tariff rates, we use estimates of the trade elasticity from the tariff-rate variable.

The second issue concerns the potential endogeneity of the tariff-rate variable. There is an extensive literature noting the potential endogeneity of tariff rates and/or dummy variables for economic integration agreements, cf., Trefler (1993) and Baier and Bergstrand (2007), respectively. Consequently, to anticipate this potential endogeneity of tariff rates, we estimate equation (50) using a two-stage instrumental variables approach. In the first stage, we regress the (gross) bilateral tariff rate (by industry), tar_{ijt}, on the mean over country j's bilateral tariffs with *all other non-i countries*; this variable is likely to be insensitive to X_{ijt}. We then use the instrument constructed from the first stage, $\hat{tar}_{ijt}$, in the second stage regression, the gravity equation in (50), alongside our measure of the c.i.f.-f.o.b. transport-cost factor.[36]

The third issue concerns accounting for variation in fixed trade costs, f_{ijt}. As we addressed above for the F/BW specifications, variation in f_{ijt} can be decomposed into three terms:

[36] Note that endogeneity of tariff rates is not a concern in our F/BW specifications as those are reduced-form regressions using time-averaged variances and covariances of double-differences of the underlying variables.

an exporter component f_{it}, an importer component f_{jt}, and a residual bilateral term f_{ijt}^R. As summarized above, much of the observed policy-based and non-policy-based factors that influence fixed trade costs tend to be *multilateral* – or country-specific – in nature. Consequently, regarding equation (50) above, the exporter-year and importer-year fixed effects will capture the vast bulk of variation in $\ln f_{ijt}$ via $\ln f_{it}$ and $\ln f_{jt}$, respectively, leaving residual variation in $\ln f_{ijt}^R$ to be accounted for by the error term ϑ_{ijt}.

5 Estimation Results

Section 5.1 presents the results from using our gravity equation to obtain estimates of the *ad valorem* variable trade-cost (positively defined) "trade elasticity," $\varepsilon_\tau = \theta\left(\frac{1+\gamma}{\gamma}\right)$. Section 5.2 provides the estimates of (structural) parameters σ and γ using our three specifications applying the F/BW methodology. Using these estimates, section 5.3 provides the implied estimates of θ and of the fixed trade-cost trade elasticity, $\varepsilon_f = \frac{\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1$. These estimates will then be used in section 6 for two numerical counterfactual analyses.

5.1 Estimation of ε_τ

Because we obtain hundreds of estimates across industries, it would not be practical to report them all; instead, we present only the distributions of the estimated coefficients. Table 2 provides the distribution of estimates of the (*ad valorem* variable trade-cost) "trade elasticity" in columns 2 and 3 using our gravity equation (50). The only difference between the two specifications is that column 2 uses *observed* gross tariff rates (tar_{ijt}) whereas column 3 uses our (two-stage least squares) instrument for gross tariff rates. As evident, the distributions of the two sets of estimates are very similar. On econometric grounds, our preferred specification is that in column 3. Recalling that the data is at the 4-digit SITC industry level, the median estimated *ad valorem* trade elasticity is 10.08, which is in line with previous estimates; the 10th-90th percentile range is 3.82-17.35.[37]

5.2 Estimation of σ and γ

Estimation results are reported in Table 3.[38] As addressed earlier, the first specification, labeled "F/BW," is the specification F/BW that includes only Group 1 variables and assumes

[37]The 10th-90th percentile range of values found is consistent with other estimates using disaggregated bilateral cross-sectional/time-series trade data , cf., Hillberry and Hummels (2013). Aggregate trade data generates lower trade elasticity estimates in the range of 2-10, cf., Anderson and van Wincoop (2004).

[38]We keep only industries for which the parameters of the model conform to the theoretical restrictions (i.e., $\sigma > 1$ and $\gamma > 0$) for all three specifications. About 25 percent of industries for which we have data are excluded from our sample. By comparison, Broda and Weinstein (2006) exclude about 35 percent of their industries.

TABLE 2
ESTIMATED VARIABLE TRADE COST ELASTICITIES

Percentile	Trade elasticity	
	OLS	IV
1	−0.29	0.36
5	2.42	2.24
10	3.97	3.82
25	6.55	7.16
50	9.41	10.08
75	12.37	13.59
90	15.92	17.35
95	18.52	19.67
99	28.13	27.44

Notes: This table presents the distributions of the estimated structural parameters of the model obtained from estimating equation (50) separately for each of 568 industries in our sample.

that the coefficients on the remaining variables in Groups 2-4 are zero. At the median, the estimate for σ is 4.70, which is in the middle of the range of σ estimates from Feenstra (1994), Table 2 for the six manufactured goods of 2.96 to 8.38; the median estimate in that group in Feenstra (1994) is 5.0. Furthermore, our range of σ estimates for percentiles 1-99 of 2.65-27.82 is similar to the range of 2.96-42.9 for all eight goods in Feenstra (1994). Broda and Weinstein (2006) provide four-digit SITC estimates for two different (averaged) time periods, 1972-1988 and 1990-2001. Using their mean estimates for differentiated products, the σ estimates for 1972-1988 and 1990-2001 are 5.2 and 4.7, respectively. Hence, our benchmark F/BW σ estimate of 4.70 is in line with both of those sets of estimates.

Our median estimate of γ using our benchmark two-RHS-variable F/BW specification is 4.03, which is also in the middle of the range of (positive) γ estimates from Feenstra (1994), Table 2 for four manufactured goods of 1.94 to 6.58; the median estimate in that group in Feenstra (1994) is 2.43. Furthermore, our range of γ estimates for percentiles 1-99 of 0.46-58.19 is similar to the (positive) range of 1.94-27.8 for the six goods in Feenstra (1994). Unfortunately, Broda and Weinstein (2006) do not report their estimates of γ.

The second specification, labeled "IMC-Partial," employs Group 1 and Group 2 variables, and assumes the coefficients on the remaining variables in Groups 3-4 are zero. Of course, the *ad valorem* (variable) trade-cost variable τ_{ijt} reflects both *ad valorem* (gross) tariff rates ($tar_{ijt} > 1$) as well as (gross) transport-cost factors ($trans_{ijt} > 1$), as discussed above. Hence, in the context of equation (42), adding Group 2 variables adds seven more RHS variables.[39] Turning to this specification's median estimates, the estimate of σ of 6.08 is *29 percent* larger in value than that in specification 1, implying OVB in the F/BW specification. Similarly,

[39] Recall that the explicit IMC-Partial specification is provided in section D.3 of Online Appendix D.

TABLE 3
ESTIMATED IMPORT DEMAND AND SUPPLY ELASTICITIES

Percentile	F/BW		IMC-Partial		IMC-Full	
	σ	γ	σ	γ	σ	γ
1	2.65	0.46	2.61	0.64	2.73	0.79
5	3.01	1.18	3.27	1.44	3.29	1.61
10	3.34	1.61	3.77	1.93	3.80	2.29
25	3.96	2.36	4.62	3.34	4.85	3.57
50	4.70	4.03	6.08	5.99	6.45	6.00
75	6.02	7.00	9.23	11.31	9.26	10.96
90	8.68	14.09	15.12	22.20	14.90	21.40
95	11.52	21.83	22.79	35.14	20.73	30.71
99	27.82	58.19	87.25	76.73	55.92	69.51

Notes: This table presents the distributions of the estimated structural parameters of the model obtained from estimating equation (44) separately for each of 568 industries in our sample using three different specifications (see main text for details). The parameter σ is the elasticity of substitution and the parameter γ is the inverse marginal cost elasticity of output.

the median estimate of γ in specification 2 is 5.99, which is *49 percent* larger than that in specification 1. These are notable differences.

The third specification, labeled "IMC-Full," includes the variables described earlier in Groups 1 and 2, but also includes an exporter fixed effect and exporter fixed effects *interacted* with $\overline{\Delta^k \ln \overline{p}_{ij}^c}$, $\overline{\Delta^k \ln s_{ij}}$, $\overline{\Delta^k \ln tar_{ij}}$, and $\overline{\Delta^k \ln trans_{ij}}$. Turning to this specification's median estimates, the estimate of σ of 6.45 is *37 percent* larger in value than that of σ in F/BW specification 1. Similarly, the median estimate of γ in specification 3 is 6.00, which is *49 percent* larger than that in specification 1. These estimates are similar, but with slightly larger values than the respective estimates in specification 2 and are notably different from those in benchmark specification 1. In the context of our theoretical model, previous estimates reveal OVB.[40]

The results presented in Table 3 have three important implications. First, our IMC estimates are quite different from the benchmark estimates. Our fuller specifications increase the estimated values of the elasticity of substitution and the bilateral export supply elasticity. Second, our IMC estimates are robust to changes in specifications. In addition to the median, the distributions of the estimates are quite similar across our both IMC specifications. Third, the estimated parameters are distributed densely around the medians.

Although we are able to compare our specifications' estimates for the elasticities of substitution with those in Feenstra (1994), Broda and Weinstein (2006), and potentially

[40]In a robustness check, using data and proxies discussed earlier for w_{it} and L_{it} alongside variables in Groups 1 and 2, but ignoring unobservable variables A_i and b_i and omitting exporter fixed effects, our median σ estimate is 6.50 and our median γ estimate is 6.34, both similar to the respective estimates for IMC-Partial and IMC-Full.

other studies' previous empirical results using similar data, the literature provides few comparisons for our estimates of γ; only Feenstra (1994) provides estimates for comparison. As noted, Broda and Weinstein (2006) do not report estimates for γ. However, Hottman et al. (2016) report an (implied) median estimate of γ of 6.25 using U.S. barcode firm-level data. Interestingly, this median estimate lies near our median IMC estimates using industry-level international trade data, but controlling for firm heterogeneity. In the next subsection, we address the importance of precise and unbiased estimates of γ for estimating θ and estimating the fixed-trade-cost elasticity, ε_f. In section 6, we demonstrate the importance of γ estimates for relevant policy-oriented quantitative comparative statics.

5.3 Estimation of θ and ε_f

Armed with estimates of ε_τ and γ, we can use our model to compute values for θ and ε_f. From gravity equation (50), $\varepsilon_\tau = \theta(1+\gamma)/\gamma$. This implies that we can recover estimates for θ as follows:

$$\theta = \left(\frac{\gamma}{1+\gamma}\right)\varepsilon_\tau. \tag{51}$$

The fixed cost elasticity can then be recovered from our estimates of σ, γ and θ and equation (15):

$$\varepsilon_f = \left[\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1\right]. \tag{52}$$

The distribution of estimates are reported in Table 4. Using the estimated values of the trade elasticity using IV (reported in Table 2) and the estimated values of γ (reported in the last column of Table 3), the second column of Table 4 provides at various percentiles the estimated values of θ, as implied by equation (51). As reported in the table, the median estimate of θ is 8.50. The third column of Table 4 reports the estimated values of the fixed trade-cost trade elasticities, ε_f, at various percentiles using equation (52) and our estimated values of σ, γ, and θ. These elasticities, along with our estimates of the three structural parameters, will be useful for our numerical comparative statics in the next section. As just one clue to the importance of IMC in those analyses, note that the fixed trade-cost trade elasticity at the median is 2.39. However, under the case of CMC, the theoretical fixed trade-cost trade elasticity (defined positively) is $\frac{\theta}{\sigma-1} - 1$, which is (under CMC) the (*ad valorem* variable trade-cost) "trade elasticity" relative to $\sigma - 1$ (minus 1). Using the median trade elasticity of 10.08 (from our gravity estimation) and our IMC-Full estimate of σ of 6.00, the implied CMC fixed trade-cost trade elasticity is only 1.02. Hence, under IMC, fixed trade-cost reductions have a larger impact of 2.3 times.

TABLE 4
ESTIMATED PARETO PARAMETERS AND FIXED TRADE COSTS ELASTICITIES

Percentile	θ	ε_f
1	0.27	-0.91
5	1.54	-0.40
10	3.03	0.09
25	5.73	1.08
50	8.50	2.39
75	11.34	4.04
90	15.13	5.88
95	17.32	7.89
99	24.39	11.95

Notes: This table presents the distributions of the Pareto parameters and the elasticities of trade estimated separately for each of the 568 industries in our dataset.

In the next section, we use our estimates to provide two different quantitative counterfactual exercises, with the purpose of showing the quantitative importance of accounting for empirically-justified increasing marginal costs in the evaluation of: (i) the "gains from trade," and (ii) the trade and welfare impacts of fixed trade-cost reductions relative to variable trade-cost reductions, the two main elements of deep trade agreements.

6 Numerical Analyses

Having established in the previous section strong empirical evidence of increasing marginal costs using international data, we provide in this section two numerical analyses to illustrate the importance of allowing for IMC. First, for a given set of parameters, we quantify the impact of allowing for increasing marginal costs in welfare calculations. Second, we use our estimates to show that the necessary changes to fixed trade costs, to obtain the welfare-equivalent of (small) changes to variable trade costs, are *much smaller* in the case of empirically-justified increasing marginal costs than in the case of constant marginal costs, helping to explain the increasing prominence of deep trade agreements in the world economy.

6.1 Counterfactual 1: Welfare Gains from Trade

We provide in this section a numerical analysis in the spirit of Feenstra (2010) and Costinot and Rodriguez-Clare (2014) to illustrate the importance of allowing for IMC in welfare calculations. We show using representative values of the (inverse) index of the heterogeneity of firms' productivities (θ) and of the inverse marginal cost elasticity of output (γ) that the welfare gains from trade are reduced by about one percentage point in the case of IMC relative to the case of CMC.

From equation (17), the percentage change in real income associated with moving from the initial equilibrium (with trade) to autarky for country j is given by (100 times):

$$G_j = 1 - \lambda_{jj}^{1/\varepsilon_\tau}, \tag{53}$$

where λ_{jj} is the domestic absorption share of GDP and $\varepsilon_\tau = \theta\left(\frac{1+\gamma}{\gamma}\right)$.[41] Consequently, the only additional data needed for this numerical exercise is trade shares. As in Feenstra (2010), we use information on nominal exports and nominal GDPs from the Penn World Tables to calculate export shares.[42] A key consideration here is comparing the gains from trade with CMC versus the gains from trade with IMC. Consequently, we also calculate the gains from trade assuming a value of $\gamma = \infty$ to obtain a benchmark value.

As explained earlier, welfare gains from trade depend on two sufficient statistics: the trade share and the trade elasticity. We explore the impact of variation in each separately, beginning with changes in the trade elasticity. In our sample, the mean trade share is 39.1 percent, so we set $\lambda_{jj} = 60.9$. Conditional on that trade share, Table 5 presents the distribution across industries of the gains from trade (relative to autarky) under the assumption of CMC ($\gamma = \infty$) and IMC as indicated at the top of each column. Our median estimate under IMC is 4.78 percent, which is a reduction of 15.4 percent from the welfare gain of 5.65 percent in the benchmark case of constant marginal costs ($\gamma = \infty$). These values are consistent with the "welfare-diminution" effect discussed in section 3.

[41] In Feenstra (2010), p. 53, G_j is defined as $[(1 - ExportShare_j)^{-1/\theta} - 1]/[(1 - ExportShare_j)^{-1/\theta}]$. However, using ACR notation and some algebra, this simplifies to $G_j = 1 - \lambda_{jj}^{1/\theta}$, which is identical to the measure of G_j in Costinot and Rodriguez-Clare (2014), p. 204.

[42] We could just as easily used the World Input-Output Database (WIOD) used in Costinot and Rodriguez-Clare (2014), but chose the set of countries in Feenstra (2010) largely due to the broader sample and wider variation in the levels of countries' per capita real GDPs.

TABLE 5
WELFARE GAINS FROM TRADE, 2010

Percentile	CMC	IMC
1	2.01	1.79
5	2.81	2.48
10	3.21	2.81
25	4.26	3.57
50	5.65	4.78
75	8.24	6.66
90	15.02	12.09
95	27.47	19.08
99	84.28	74.83

Notes: This table presents the absolute value of the percentage change in real income associated with moving from the initial equilibrium to autarky given by $1 - \lambda_{jj}^{1/\varepsilon_\tau}$, where λ_{jj} is domestic absorption. In our sample, the mean trade share is 39.1, so we set $\lambda_{jj} = 60.9$. We compute gains from trade separately for each of the 568 industries in our sample. In this section, we use s_{jj} to measure λ_{jj}.

Table 6 presents the results for the impact of changes in the trade share, holding the elasticity constant at the median values. It reports calculations of the gains from trade for 20 countries of various levels of per capita real GDP, similar to Table 3.1 in Feenstra (2010). As expected, countries with larger export shares have larger gains from opening up from autarky. For instance, the United States has a small export share; consequently, the gains from trade are smaller. However, the presence of IMC still has a substantive effect for the United States; the reduction of welfare of 0.24 from 1.53 to 1.29 owing to increasing marginal costs is 15.6 percent. Overall, the results presented in this section suggest that increasing marginal costs have substantive effects on welfare calculations.

6.2 Counterfactual 2: Welfare-Equivalent Changes and Deep Trade Agreements

As discussed in the introduction, the "new millennium" has also introduced "new types of trade agreements." The stark contrast between shallow versus deep trade agreements is essentially the difference between reducing *ad valorem* tariff rates on international trade versus reducing "regulatory heterogeneity":

> *Accordingly, the emphasis of trade liberalization has shifted from reducing protectionist barriers (i.e., tariff rates) to harmonizing – to the extent possible – rules and regulations. Noting the shift in emphasis, former WTO Director General Pascal Lamy put it this way: "TTIP isn't about trade trade-offs, but a process of regulatory convergence, which is a totally different ball game."* Norberg (2015), p.1.

TABLE 6
WELFARE GAINS FROM TRADE FOR SELECTED COUNTRIES, 2010

Name	GDPPC	Export Share	CMC	IMC
Guinea	1,677	30.34	4.16	3.52
Mali	1,736	22.84	3.00	2.54
Nepal	1,807	9.58	1.18	0.99
Kyrgyzstan	2,863	51.55	8.17	6.93
Republic of Moldova	3,737	39.23	5.69	4.81
Congo	4,709	65.81	11.86	10.09
Guatemala	6,293	25.81	3.45	2.91
China	9,423	26.27	3.52	2.97
Thailand	13,109	66.49	12.07	10.26
Gabon	13,151	57.66	9.62	8.16
Brazil	13,623	10.74	1.33	1.12
Malaysia	20,192	86.93	21.29	18.26
Israel	30,538	35.02	4.94	4.18
Bahamas	31,413	34.95	4.93	4.17
Italy	35,936	25.19	3.36	2.83
Germany	40,481	42.25	6.26	5.29
Saudi Arabia	41,482	49.57	7.74	6.56
United States	49,907	12.32	1.53	1.29
Norway	57,900	39.73	5.78	4.89
Bermuda	62,290	49.69	7.76	6.58

Notes: This table presents the absolute value of the percentage change in real income associated with moving from the initial equilibrium to autarky given by $1 - \lambda_{jj}^{1/\varepsilon_\tau}$, where λ_{jj} is domestic absorption, computed for selected countries for year 2010. To the extent possible, we choose the same countries as in Table 3.1 of Feenstra (2010) to facilitate comparison.

As illustrated recently in the United States-Mexico-Canada Agreement, the successor to NAFTA, deep trade agreements embody a large increase in the number of chapters and the scope of the agreement. In reality, these developments essentially span three (partially overlapping) areas:

1. Modern trade agreements have been deepened to cover services trade flows, capital flows, migration flows, and idea flows;

2. Modern trade agreements aim to reduce barriers at the border and behind the border in terms of regulatory convergence, such as trade facilitation (customs administration), technical barriers to trade, sanitary and phytosanitary measures, and competition policy;

3. Modern trade agreements extend to addressing environmental policy and labor rights.

For our purposes, we are addressing the second category, where regulatory divergences create costs of trade unrelated to the level of output, i.e., fixed trade costs.

One of the earliest studies to document and categorize the degree to which European Union and United States' preferential trade agreements (PTAs) incorporated liberalizations beyond tariff-rate reductions that would reduce fixed trade costs is Horn et al. (2010), documenting such liberalizations beyond that established by the World Trade Organization (WTO). Evidence from the World Bank's DTA database suggests that several (inverse) indexes of fixed trade costs – legally enforceable provisions provided in DTAs such as trade facilitation, technical barriers to trade, sanitary and phytosanitary measures, and competition policy – have increased over time. For instance, Hofmann et al. (2017) note that the simple count of legally enforceable provisions included in PTAs increased from 8 to 17 from the 1990s to 2015. More rigorously, Hofmann et al. (2017) created a measure of depth of PTAs using Principal Components Analysis (PCA); this PCA measure indicated that PTAs' depth has increased, on average, *150 percent* from the 1980 to 2015, as seen in Figure 2.

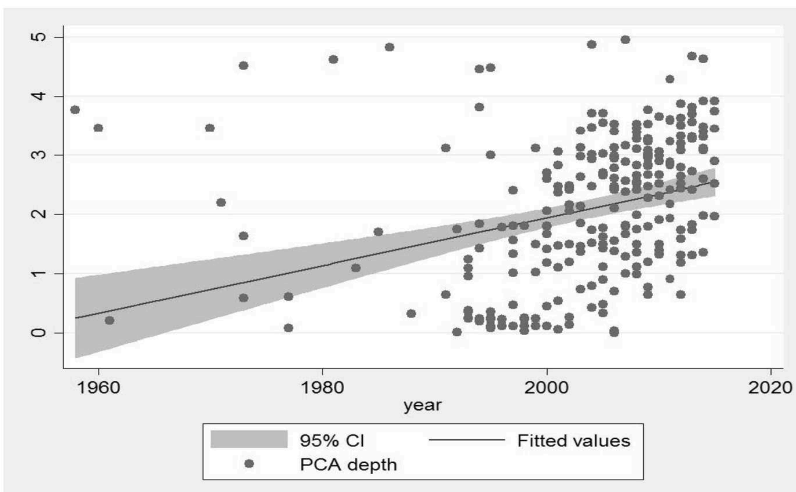

Figure 2: Hofmann et al (2017) PCA Measure of DTA "Depth" Over Time

Moreover, recent empirical studies using gravity equations have demonstrated evidence of substantive effects of reductions in policy-related fixed trade costs on bilateral trade flows. First, Baier et al. (2014) documented using state-of-the-art panel techniques that deeper economic integration agreements (embodying fixed-trade-cost reductions) had larger partial effects on bilateral trade flows. Baier and Regmi (2020) was the first study using machine-learning techniques to show that deeper PTAs have larger trade-creation effects, noting substantive effects from provisions on anti-dumping, competition policy, customs harmonization, e-commere, export and import restrictions, sanitary and phytosanitary measures, and technical barriers to trade. Breinlich et al. (2021) similarly used machine-learning techniques to examine the effects of deep trade agreement provisions on trade flows, finding that provisions associated with technical barriers to trade, anti-dumping, and competition policy had significant effects. Other recent studies finding significant effects of DTAs on trade flows are Crowley et al. (2020) and Fontagne et al. (2022).

While empirical studies are now starting to flourish given the World Bank's new data base on the DTAs, the theoretical and quantitative welfare effects of deep trade agreements have been scarcely examined, especially in the context of the new trade theory with heterogeneous firms. Specifically, to the authors' knowledge only four papers address systematically quantifying the trade and welfare effects of bilateral (*ad valorem*) variable trade-cost liberalizations *relative to* fixed trade-cost changes. As mentioned in the introduction, Zhai (2008) is among the earliest of these rare studies that have introduced a Melitz model into a CGE model to calculate the trade and welfare effects of three types of policy simulations: a 50 percent tariff-rate cut, a 5 percent reduction in variable trade costs, and a 50 percent reduction in fixed trade costs. The CGE model's implementation of the Melitz framework (under CMC) is consistent with the discussion in this paper. For purposes of this paper, we discuss the implications of the latter two simulations; the reason is that Zhai (2008) allows tariffs to generate income, whereas variable trade costs are "iceberg" trade costs, as in this paper. A 50 percent tariff-rate reduction in Zhai (2008) reduces disposable income, which has an offsetting effect on expenditures and trade; the model in this paper ignores this aspect (which is left for future research).

Using a multi-country framework, a value of σ of 5, and a value of θ of 6.2, Zhai (2008) found for the United States, for example, that a 5 percent reduction in variable trade costs increased welfare by 32.8 billion (US) dollars. In the context of his model, a 50 percent reduction in fixed trade costs increased welfare by 44.8 billion (US) dollars. Hence, the welfare-equivalent reduction in fixed trade costs would be 36 (29) percent, to match a 5 (4) percent reduction in variable trade costs (or a ratio of approximately 7.25:1). This accords quantitatively to the notion that, for the same percent reduction in the cutoff productivity φ_{ij}^*, the fall in f_{ij} would need to be about 7 times, since φ_{ij}^* adjusts in proportion to $f_{ij}^{1/(\sigma-1)}$ in the case of CMC. In CGE analyses of the TTIP, a reduction of 36 percent in non-tariff measures was considered "very ambitious," and such a differential suggests *against* the proliferation of deep trade agreements.

To the authors' knowledge, only three other papers have considered CGE analyses using a Melitz framework, Balisteri et al. (2011), Dixon et al. (2016), and Arkolakis et al. (2021). The structure of Balisteri et al. (2011) is similar in many respects to Zhai (2008), but differs in several other respects. Balisteri et al. (2011) actually estimate values for σ and even θ, and use exporter and importer fixed effects to estimate exporter- and importer-specific fixed trade costs (assuming CMC). The residuals in their approach are bilateral fixed trade costs, which adjust to match the simulated bilateral trade flows to actual trade flows. This method yields some difficult-to-rationalize bilateral fixed trade costs. For instance, the bilateral fixed trade cost of exports from the United States to Japan is twice as high as that from Canada to Japan; moreover, the fixed trade costs of intra-national Japanese trade are the same as fixed trade costs from Canada to Japan. Nevertheless, Balisteri et al. (2011) only compare a 50 percent reduction in tariff rates against a 50 percent reduction in fixed trade costs, which provides a non-comparable comparison to Zhai (2008) and our model, since tariff cuts in Balisteri et al. (2011) involve reductions in disposable income and cannot be compared to a 50 percent reduction in iceberg variable trade costs, as we know from Zhai (2008). Another CGE model with a Melitz framework is Dixon et al. (2016). However, this study only examined relative impacts of reductions in (*ad valorem*) variable trade costs across Melitz and Krugman versions of their model. Finally, as noted in the introduction, Arkolakis et al. (2021) extends the Melitz model of trade to show that – for multiproduct firms facing constant marginal costs in producing their core product (though increasing marginal market-penetration costs) – additional products that are farther from the firm's core competency face increasing marginal production costs (despite economies of scope in market-access costs). Of particular relevance to this paper, the last substantive section of Arkolakis et al. (2021) conducts counterfactual experiments of reductions in market-access costs and, for comparison, tariff rates. In their baseline simulation, the elimination of the recently observed average 4 percent tariff rates in the world generates a welfare gain of 1.8 percent. In contrast, using their Table 6, Counterfactual 1 experiment of reducing total market-access costs, a 15 percent reduction in such fixed trade costs improves welfare by 2.0 percent. Hence, for comparison of the results in this model relative to Zhai (2008) (and later to our counterfactual), it would take a 13 percent reduction in fixed trade costs to generate the same welfare as a reduction in tariff rates of 4 percent, a ratio of 3.25:1.

In our second counterfactual, we are interested in measuring fixed trade-cost changes, $\hat{f}_{ij}$, that are equivalent in welfare to changing a given (*ad valorem*) variable trade cost, $\hat{\tau}_{ij}$. In our model, as seen in equation (13), we can write:

$$\phi_{ij} = \tau_{ij}^{-\varepsilon_\tau} f_{ij}^{-\varepsilon_f}, \tag{54}$$

such that for a given value of (the gross tariff rate) $\hat{\tau}_{ij}$, we define the welfare-equivalent fixed trade-cost change as $\hat{f}_{ij} = \hat{\tau}_{ij}^{\frac{\varepsilon_\tau}{\varepsilon_f}}$. This gives the increase in fixed trade costs that is equivalent to an increase in variable trade costs in terms of its impact on trade flows and welfare.[43]

[43] We express the term as shown for expositional convenience. Mathematically, for values $\phi_{ij} = \tau_{ij}^{-\varepsilon_\tau} f_{ij}^{-\varepsilon_f}$ and $\tilde{\phi}_{ij} = \tilde{\tau}_{ij}^{-\varepsilon_\tau} \tilde{f}_{ij}^{-\varepsilon_f}$, if $\phi_{ij} = \tilde{\phi}_{ij}$ then $\hat{f}_{ij} = \hat{\tau}_{ij}^{-\frac{\varepsilon_\tau}{\varepsilon_f}}$ where $\hat{f}_{ij} \equiv \tilde{f}_{ij}/f_{ij}$ and $\hat{\tau}_{ij} \equiv \tilde{\tau}_{ij}/\tau_{ij}$.

Using results from section 3, the ratio of elasticities plays a critical role in defining welfare-equivalent trade-cost changes. From the theoretical model with IMC, we know that:

$$\frac{\varepsilon_\tau}{\varepsilon_f} \equiv \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1}. \tag{55}$$

For any value of $\gamma < \infty$, this ratio is smaller than in the benchmark CMC case. In the limit, as $\gamma \to \infty$, the ratio converges to the benchmark. This implies that under IMC the welfare-equivalent change $\hat{f}_{ij}$ for a given $\hat{\tau}_{ij}$ is smaller than under CMC. The economic intuition was discussed in section 3.

Consider the median values of our estimated parameters using the IMC-Full specification from section 5, $\sigma = 6.45$, $\gamma = 6.00$, and $\theta = 8.50$. Substituting in these values yields:

$$CMC : \frac{\varepsilon_\tau}{\varepsilon_f} = \frac{\theta}{\frac{\theta}{\sigma-1} - 1} = 15.18 \tag{56}$$

$$IMC : \frac{\varepsilon_\tau}{\varepsilon_f} = \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1} = 4.44 \tag{57}$$

Armed only with *observable* estimates of variable trade costs (for which we use average MFN tariff rates), we can then obtain a fixed trade-cost change that is *equivalent in welfare* to introducing a country's – or an average of countries' – MFN tariff rates. In our sample, the average tariff rates applied is about 4 percent. This implies that the welfare-equivalent fixed costs changes are:

$$CMC : \hat{f} = (1.04)^{15.18} = 1.81 \tag{58}$$

$$IMC : \hat{f} = (1.04)^{4.44} = 1.19. \tag{59}$$

These results make clear that the equivalent change is much larger under CMC.

The welfare-equivalent fixed trade-cost change depends on two sufficient statistics: the level of trade barriers and the ratio of elasticities. As we did for welfare, we explore the impact of each in turn. Table 7 reports the distribution of the ratio of welfare-equivalent fixed trade-cost changes across industries. To discipline the quantitative exercise, we use the mean tariff of 4 percent in each industry and consider the introduction of the tariff rate as our shock, $\hat{\tau}$.[44] We compute the welfare-equivalent change for two separate cases, the benchmark CMC case of $\gamma \to \infty$ and the IMC case, as indicated at the top of each column. The table

[44] Due to our model's construct, we are ignoring any change of tariff revenue, leaving this for future research.

shows that, for the median industry, the equivalent fixed trade-cost change under CMC is 38 percent, whereas under IMC it is only 15 percent. Both distributions of welfare-equivalent fixed trade-cost changes start at 0 percent, but the CMC distribution has a much thicker right tail. At the 90th percentile of the distribution, the equivalent fixed trade-cost change is an increase of *368 percent*. But under IMC, it is a much more reasonable 32 percent.

Most importantly, we note the following implications from Table 7. For the median industry, under CMC in our model, it would take a 28 [=100(1-1/1.38)] percent reduction in fixed trade costs to be welfare equivalent to a 4 percent reduction in variable trade costs. This result is very similar to the findings mentioned earlier for Zhai (2008). By contrast, in our model with IMC, it would take only a 13 percent reduction in fixed trade costs to be welfare equivalent to a 4 percent reduction in variable trade costs, which accords more with the study of Brazilian exporters in Arkolakis et al. (2021), which allowed increasing marginal market-penetration costs and increasing marginal costs associated with non-core additional products.

TABLE 7
EQUIVALENT FIXED COSTS CHANGE

	CMC		IMC	
Percentile	$\varepsilon_\tau/\varepsilon_f$	$\hat{f}$	$\varepsilon_\tau/\varepsilon_f$	$\hat{f}$
1	2.08	1.08	1.23	1.05
5	2.68	1.11	1.88	1.07
10	3.57	1.15	2.14	1.09
25	5.08	1.21	2.79	1.11
50	8.35	1.38	3.69	1.15
75	16.82	1.91	4.87	1.21
90	40.27	4.68	7.23	1.32
95	219.75	21.30	9.20	1.42
99	546.47	4550.28	16.90	1.91

Notes: This table presents the distribution of the average industry-level welfare-equivalent fixed-trade-cost changes ($\hat{f}$). We set $\tau = 1.04$ (the sample import-weighted mean) and let the elasticity of substitution (σ), the inverse elasticity of marginal costs (γ), and the Pareto parameter (θ) vary across industries. The equivalent fixed-trade-cost changes are obtained from $\hat{f}_{ij} = \hat{\tau}_{ij}^{\varepsilon_\tau/\varepsilon_f}$. We keep the 406 industries in the sample for which the fixed-trade-cost elasticities are positive.

Table 8 reports the distribution of the ratio of welfare-equivalent fixed trade-cost changes for selected countries. This exercise aims to illustrate the impact of differences in trade barriers, so we set the ratio of elasticities at their median values. For each country, we compute the import-weighted average tariff. Again, we set the shock to introducing the country's average tariff rate. As in Table 7, we compute the welfare-equivalent fixed trade-cost changes for the CMC and IMC cases. Here, the main point is that – even if the parameters are the same across countries – changes in the compositions of trade flows have an impact on equivalent changes.

TABLE 8
Average Equivalent Fixed Costs Changes for Selected Countries, 2010

			CMC		IMC	
Name	GDPPC	Mean tariff	$\varepsilon_\tau/\varepsilon_f$	$\hat{f}$	$\varepsilon_\tau/\varepsilon_f$	$\hat{f}$
Guinea	1,677	1.08	13.75	3.04	4.41	1.43
Mali	1,736	1.09	15.80	4.18	5.09	1.59
Nepal	1,807	1.12	16.03	5.86	5.11	1.76
Kyrgyzstan	2,863	1.01	20.91	1.25	5.94	1.07
Moldova	3,737	1.03	21.58	1.75	5.08	1.14
Congo	4,709	1.15	14.60	8.16	4.13	1.81
Guatemala	6,293	1.05	15.94	2.05	4.25	1.21
China	9,423	1.08	21.46	5.50	4.82	1.47
Thailand	13,109	1.08	18.11	3.94	4.46	1.40
Gabon	13,151	1.15	15.21	8.86	3.89	1.75
Brazil	13,623	1.11	28.60	20.73	4.57	1.62
Malaysia	20,192	1.08	20.19	4.92	4.35	1.41
Israel	30,538	1.06	20.53	3.08	4.44	1.28
Bahamas	31,413	1.29	16.70	66.00	4.48	3.08
Italy	35,936	1.01	25.42	1.33	5.10	1.06
Germany	40,481	1.01	27.05	1.43	4.49	1.06
Saudi Arabia	41,482	1.09	19.79	5.58	5.38	1.60
United States	49,907	1.03	26.01	2.31	4.52	1.16
Norway	57,900	1.01	20.70	1.14	4.35	1.03
Bermuda	62,290	1.19	20.25	31.79	3.93	1.96

Notes: This table presents the distribution of the average country-level welfare-equivalent fixed-trade-cost changes ($\hat{f}$) for selected countries for year 2010. We set all parameters equal to the country's import-weighted averages. The equivalent fixed-trade-cost changes are obtained from $\hat{f}_{ij} = \hat{\tau}_{ij}^{\varepsilon_\tau/\varepsilon_f}$. To the extent possible, we choose the same countries as in Table 3.1 of Feenstra (2010) to facilitate comparison.

We conclude by addressing a result for each of the United States and Germany. For the United States (Germany), the MFN tariff rate is only about 3 (1) percent, which conforms to most observers knowledge of it. While the initial value of bilateral fixed trade costs is unknown, the lack of that knowledge is immaterial for our calculations. All that is needed here is values of average tariff rates (or variable trade costs), the well-known (*ad valorem* variable-trade-cost) "trade elasticity," and a value for the fixed trade-cost trade elasticity. With little empirical knowledge of the *levels* of fixed trade costs, our estimates of σ, γ, and θ allow us to construct an estimate of $\frac{\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)} - 1$. Given the framework above and assuming IMC, we find that eliminating remaining U.S. tariffs of 3 percent are welfare-equivalent to a reduction in fixed trade costs of *only 14 percent* [= 100(1 − 1/1.16)]. For Germany, we find that eliminating their remaining tariffs of 1 percent are welfare-equivalent to a reduction in fixed trade costs of 6 percent. These results make deep trade agreements much more attractive to pursue, with a 14 (6) percent reduction well below the reductions of 25 percent used in earlier analyses of the Transatlantic Trade and Investment Partnership (TTIP) in Berden et al. (2010).

7 Conclusions

This paper has offered three contributions to the international trade literature: theoretical, empirical, and numerical. First, extending theoretically a standard (one-sector) Melitz model of international trade to the case of increasing marginal costs, we generated a gravity equation where the trade elasticity (θ) was magnified by one plus the marginal cost elasticity of output, implying that the welfare gains from trade are reduced as diminishing marginal returns interact with the Pareto shape parameter to lower the average productivity gains from trade liberalizations.

Second, introducing a novel econometric extension of the Feenstra/Broda-Weinstein method that controlled explicitly for firm heterogeneity, we find that increasing marginal costs exist, with an across-industry median bilateral export supply elasticity estimate of 6.00 in our preferred specification – *far below* ∞, which is assumed in the benchmark models in the trade literature assuming constant marginal costs.

Third, we provided two numerical analyses to illustrate quantitatively the relative importance of our study. In the first counterfactual, we examined the relative quantitative importance of increasing marginal costs for estimating the welfare gains from trade. Our second – and more novel – counterfactual provided insight into the increasing prominence of deep trade agreements in the world economy. Under constant marginal costs for the median industry, the needed reduction in fixed trade costs to be equivalent in welfare-improvement to a 4 percent reduction in *ad valorem* variable trade costs was 28 percent, the latter considered "ambitious" in most CGE analyses of deep trade liberalizations. By contrast, under increasing marginal costs, the welfare-equivalent reduction in fixed trade costs is only *13 percent* for the median industry.

We offer two suggestions for future research in this area. First, to reduce theoretical complexity, we have omitted disposable income associated with tariff revenues; future work could incorporate tariff revenue for computing the welfare effects of reducing tariff rates. Second, our framework could be extended in the future to incorporate the role of tastes for regulatory divergences addressed in Grossman et al. (2021) to better understand and potentially quantify the welfare-equivalent effects of fixed versus variable trade-cost reductions.

References

Alvarez, F. and R. Lucas (2007). General equilibrium analysis of the eaton-kortum model of international trade. *Journal of Monetary Economics 54*(6), 1726–1768.

Anderson, J. (1979). A theoretical foundation for the gravity equation. *American Economic Review 69*(1), 106–116.

Anderson, J. (2011). The gravity model. *Annual Review of Economics 3*, 133–160.

Anderson, J. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review 93*(1), 170–192.

Anderson, J. and E. van Wincoop (2004). Trade costs. *Journal of Economic Literature 42*(3), 691–751.

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

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

Arkolakis, C., S. Ganapati, and M.-A. Muendler (2021). The extensive margin of exporting products: A firm-level analysis. *American Economic Journal: Macroeconomics 13*(4), 182–245.

Bagwell, K. (2007). The economic analysis of advertising. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*, pp. 1701–1844. Amsterdam, Netherlands: North Holland.

Baier, S. and J. Bergstrand (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics 71*(1), 72–95.

Baier, S., J. Bergstrand, and M. Clance (2018). Heterogeneous effects of economic integration agreements. *Journal of Development Economics 135*, 587–608.

Baier, S., J. Bergstrand, P. Egger, and P. McLaughlin (2008). Do economic integration agreements actually work? issues in understanding the causes and consequences of the growth of regionalism. *The World Economy 31*(4), 461–497.

Baier, S., J. Bergstrand, and M. Feng (2014). Economic integration agreements and the margins of international trade. *Journal of International Economics 93*(2), 339–350.

Baier, S., A. Kerr, and Y. Yotov (2017). Gravity, distance, and international trade. In W. Wesley and B. Blonigen (Eds.), *Handbook of International Trade and Transportation*. Edward Elgar Publishing.

Baier, S. and N. Regmi (2020). Using machine learning to capture heterogeneity in trade agreements and its effect on trade flows. *Working Paper*.

Balisteri, E., R. Hillberry, and T. Rutherford (2011). Structural estimation and solution of international trade models with heterogeneous firms. *Journal of International Economics 83*, 95–108.

Bartelme, D., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2019). The textbook case for industrial policy: Theory meets data. *Working Paper*.

Berden, K., J. Francois, S. Tamminen, M. Thelle, and P. Wymenga (2010). *Non-Tariff Measures in EU-US Trade and Investment*. The Netherlands: ECORYS.

Bergstrand, J. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics 67*(3).

Bernard, A., S. Redding, and P. Schott (2011). Multiproduct firms and trade liberalization. *Quarterly Journal of Economics 126*, 1271–1318.

Breinlich, H., V. Corradi, N. Rocha, M. Ruta, J. Silva, and T. Zylkin (2021). Machine learning in international trade research: Evaluating the impact of trade agreements. *Working Paper*.

Broda, C., N. Limao, and D. Weinstein (2008). Optinmal tariffs and market power: The evidence. *American Economic Review 98*(5), 2032–2065.

Broda, C. and D. Weinstein (2006). Globalization and the gains from variety. *Quarterly Journal of Economics 121*(2), 541–585.

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

Costinot, A. and A. Rodriguez-Clare (2014). Trade theory with numbers. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of Interantional Economics, Volume 4*. Amsterdam: Elsevier.

Crowley, M., L. Han, and T. Prayer (2020). The value of deep trade agreements in the presence of pricing-to-market. *Working Paper*.

Dai, M., M. Muitra, and M. Yu (2016). Unexceptional exporter performance in China? The role of processing trade. *Journal of Development Economics 121*, 177–189.

Dixon, P., M. Jerie, and M. Rimmer (2016). Modern trade theory for CGE modelling: The Armington, Krugman and Melitz models. *Journal of Global Economic Analysis 1*(1), 1–110.

Eaton, J. and S. Kortum (2002). Technology, geography and trade. *Econometrica 70*(5).

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

Fajgelbaum, P. D., O. K. Goldberg, P. J. Kennedy, and A. K. Khandelwal (2020). The return of protectionism. *Quarterly Journal of Economics 135*(1), 1–55.

Farrokhi, F. and A. Soderbery (2020). Trade elasticities in general equilibrium. *Working Paper*.

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

Feenstra, R. (2010). *Product Variety and the Gains from International Trade*. Cambridge, MA: MIT Press.

Feenstra, R. (2016). *Advanced International Trade: Theory and Evidence, Second Edition*. Princeton, New Jersey: Princeton University Press.

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

Feenstra, R. C., , P. Luck, M. Obstfeld, and K. N. Russ (2018). In search of the Armington elasticity. *Review of Economics and Statistics 100*(1), 135–150.

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

Flach, L. and F. Unger (2022). Quality and gravity in international trade. *Journal of International Economics 137*.

Fontagne, L., N. Rocha, M. Ruta, and G. Santoni (2022). The economic impact of deep trade agreements. *CESifo Working Paper* (9529).

Goldberg, P. K. and N. Pavcnik (2016). The effects of trade policy. In K. Bagwell and R. Staiger (Eds.), *The Handbook of Commercial Policy. Elsevier, Amsterdam*, Volume 1, pp. 161–206. Elsevier.

Grossman, G., P. McCalman, and R. Staiger (2021). The new economics of trade agreements: From trade liberalization to regulatory convergence? *Econometrica 89*(1).

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of Interantional Economics, Volume 4*. Amsterdam: Elsevier.

Hillberry, R. and D. Hummels (2013). Trade elasticity parameters for a computable general equilibrium model. In P. B. Dixon and D. W. Jorgenson (Eds.), *Handbook of Computable General Equiilibrium Modeling*, pp. 1213–1269. Amsterdam: Elsevier.

Hofmann, C., A. Osnago, and M. Ruta (2017). Horizontal depth: A new database on the content of preferential trade agreements. *World Bank Working Paper 7981*.

Horn, H., P. Mavroidis, and A. Sapir (2010). Beyond the wto? an anatomy of eu and us preferential trade agreements. *The World Economy 33*(11), 1565–1588.

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

Jones, J. P. (1995). *New Proof Tat Advertising Triggers Sales.* New York, NY: Lexington.

Kohl, T., S. Brakman, and H. Garretsen (2016). Do trade agreements stimulate international trade differently? evidence from 296 trade agreements. *The World Economy 39*(1), 97–131.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review 70*(5), 950–959.

Kucheryavyy, K., G. Lyn, and A. Rodriguez-Clare (2019). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. unpublished manuscript.

Lashkaripour, A. and V. Lugovskyy (2019). Scale economies and the structure of trade and industrial policy. unpublished manuscript.

Lu, D. (2010). Exceptional exporter performance? Evidence from Chinese manufacturing firms. *Job Market Paper*.

Mas-Colell, A., M. Whinston, and J. Green (1995). *Microeconomic Theory.* Oxford University Press.

Mattoo, A., N. Rocha, and M. Ruta (2020). The evolution of deep trade agreements. In A. Mattoo and N. Rocha (Eds.), *Handbook of Deep Trade Agreements.* World Bank.

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

Norberg, H. (2015). TTIP is not your father's free trade agreement: It has potential to benefit all. *CATO Online Forum*.

Ossa, R. (2016). Quantitative models of commerical policy. In K. Bagwell and R. Staiger (Eds.), *Handbook of Commercial Policy*, Volume 1A, pp. 207–259. Amsterdam, The Netherlands: Elsevier.

Redding, S. (2011). Theories of heterogeneous firms. *Annual Review of Economics 3*.

Saunders, J. (1987). The specification of aggregate market models. *European Journal of Marketing 21*, 5–47.

Simonovska, J. and J. Waugh (1980). The shape of the advertising response function. *Journal of Advertising Research 20*, 767–784.

Soderbery, A. (2015). Estimating import supply and demand elasticities: Analysis and implications. *Journal of International Economics 96*, 1–17.

Soderbery, A. (2018). Trade elasticities, heterogeneity, and optimal tariffs. *Journal of International Economics 114*, 44–62.

Sutton, J. (1991). *Sunk Costs and Market Structure*. Cambridge, MA: MIT Press.

Trefler, D. (1993). Trade liberalization and the theory of endogeneous protection: An econometric study of u.s. import policy. *Journal of Political Economy 101*(1), 138–160.

U.S. Intl. Trade Commission (2019). *U.S.-Mexico-Canada Trade Agreement: Likely Impact on the U.S. Economy and on Specific Industry*. Washington, DC: USITC.

Vannoorenberghe, G. (2012). Firm-level volatility and exports. *Journal of International Economics 86*(1), 57–67.

WorldBank (2020). *Doing Business*. Washington, DC: World Bank.

Zhai, F. (2008). Armington meets Melitz: Introducing firm heterogeneity in a global CGE model of trade. *Journal of Economic Integration 23*(3), 575–604.

Online Appendices

to

"Increasing Marginal Costs, Firm Heterogeneity, and the Gains from "Deep" International Trade Agreements"

March 17, 2023

A Appendix A

A.1 Pricing Rule and Firm Revenue

As in Feenstra (2010), we let $p_{ij}(\varphi)$ and $q_{ij}(\varphi)$ denote the (free-on-board or fob) price received and the quantity shipped by the firm at the factory gate, respectively. A firm with productivity φ in country i serving country j maximizes profits by choosing the factory-gate price p_{ij}:

$$\max_{p_{ij}} \pi_{ij}(\varphi) = p_{ij}(\varphi)q_{ij}(\varphi) - \tilde{w}_i \left[f_{ij} + \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \right]. \tag{A.1}$$

where $\tilde{w}_i \equiv w_i/A_i$. By the definition of iceberg trade costs, we have that the quantity produced after the "iceberg melt" is equal to the quantity consumed: $q_{ij}(\varphi)/\tau_{ij} = c_{ij}(\varphi)$. Furthermore, because firms charge $p_{ij}(\varphi)$ per unit *produced*, consumers pay $p^c_{ij}(\varphi) \equiv \tau_{ij}p_{ij}(\varphi)$ per unit *consumed*. Combining these results and making use of the demand function in equation (2) in the paper, we can express output as:

$$q_{ij}(\varphi) = \tau_{ij}c_{ij}(\varphi) = \tau_{ij}E_jP_j^{\sigma-1}b_i^{1-\sigma}p^c_{ij}(\varphi)^{-\sigma} = E_jP_j^{\sigma-1}\tau_{ij}^{1-\sigma}b_i^{1-\sigma}p_{ij}(\varphi)^{-\sigma}. \tag{A.2}$$

Substituting this last result into equation (A.1) yields

$$\max_{p_{ij}} \pi_{ij}(\varphi) = E_jP_j^{\sigma-1}\tau_{ij}^{1-\sigma}b_i^{1-\sigma}p_{ij}(\varphi)^{1-\sigma} - \tilde{w}_if_{ij} - \frac{\tilde{w}_i}{\varphi}\left[E_jP_j^{\sigma-1}\tau_{ij}^{1-\sigma}b_i^{1-\sigma}p_{ij}(\varphi)^{-\sigma}\right]^{\frac{1+\gamma}{\gamma}}.$$

Because each firm produces only one of a continuum of varieties, a change in p_{ij} has a negligible effect on the price index P_j. As a result, the first order condition for the profit-maximization problem is:

$$\frac{\partial \pi_{ij}}{\partial p_{ij}} = (1-\sigma)E_jP_j^{\sigma-1}b_i^{1-\sigma}\tau_{ij}^{1-\sigma}p_{ij}(\varphi)^{-\sigma} + \sigma\left(\frac{1+\gamma}{\gamma}\right)\frac{\tilde{w}_i}{\varphi}\left(E_jP_j^{\sigma-1}b_i^{1-\sigma}\tau_{ij}^{1-\sigma}\right)^{\frac{1+\gamma}{\gamma}} p_{ij}(\varphi)^{-\sigma\left(\frac{1+\gamma}{\gamma}\right)-1} = 0,$$

Simplifying the equation above yields:

$$p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i}{\varphi}\left[E_j P_j^{\sigma-1} b_i^{1-\sigma} \tau_{ij}^{1-\sigma} p_{ij}(\varphi)^{-\sigma}\right]^{\frac{1}{\gamma}}.$$

From equation (A.2) we can replace with $q_{ij}(\varphi)$ the last term in the squared brackets in the equation above to obtain the optimal factory-gate price:

$$p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i}{\varphi} q_{ij}(\varphi)^{\frac{1}{\gamma}}. \tag{A.3}$$

We can use this result to derive optimal firm-destination revenue as follows:

$$r_{ij}(\varphi) = p_{ij}(\varphi) q_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}. \tag{A.4}$$

As explained earlier, firms charge $p_{ij}(\varphi)$ per unit *produced* such that consumers pay $p^c_{ij}(\varphi) \equiv \tau_{ij} p_{ij}(\varphi)$ per unit *consumed*. From equation (A.3), consumers pay a price per unit consumed of:

$$p^c_{ij}(\varphi) \equiv \tau_{ij} p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i \tau_{ij}}{\varphi} q_{ij}(\varphi)^{\frac{1}{\gamma}}. \tag{A.5}$$

Finally, we note that our solution for optimal consumer price converges to the benchmark result as $\gamma \to \infty$:

$$\lim_{\gamma \to \infty} p^c_{ij}(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i \tau_{ij}}{\varphi}.$$

A.2 Firm Profits

From equation (A.1), we have:

$$\begin{aligned}
\pi_{ij}(\varphi) &= p_{ij}(\varphi) q_{ij}(\varphi) - \tilde{w}_i \left[f_{ij} + \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}\right] \\
&= r_{ij}(\varphi) - \tilde{w}_i f_{ij} - \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}\right] \\
&= r_{ij}(\varphi) - \tilde{w}_i f_{ij} - \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right) r_{ij}(\varphi) \\
&= \left[1 - \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\right] r_{ij}(\varphi) - \tilde{w}_i f_{ij} \\
&= \left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{r_{ij}(\varphi)}{\sigma} - \tilde{w}_i f_{ij} \qquad \text{(A.6)}
\end{aligned}$$

where the third line uses the definition of optimal revenue in equation (A.4). We note that our solution for profits converges to the benchmark result as $\gamma \to \infty$:

$$\lim_{\gamma\to\infty} \pi_{ij}(\varphi) = \frac{r_{ij}(\varphi)}{\sigma} - \tilde{w}_i f_{ij}.$$

A.3 Cutoff Productivity

Together, the profit function defined in equation (A.1) and the zero-profit condition $\pi_{ij}(\varphi^*_{ij}) = 0$ imply that:

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{r_{ij}(\varphi^*_{ij})}{\sigma} = \tilde{w}_i f_{ij}. \tag{A.7}$$

Substituting into this last equation optimal revenue, as defined in equation (A.4), yields:

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right)\left(\frac{1}{\sigma}\right)\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i q_{ij}(\varphi^*_{ij})^{\frac{1+\gamma}{\gamma}}}{\varphi^*_{ij}} = \tilde{w}_i f_{ij}, \tag{A.8}$$

which, after rearranging, yields an expression for the optimal output of the cutoff firm:

$$q_{ij}(\varphi^*_{ij}) = \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\varphi^*_{ij}\right]^{\frac{\gamma}{1+\gamma}}. \tag{A.9}$$

We can substitute this last result into equation (A.3) to obtain an expression for the optimal factory-gate price for the cutoff firm:

$$\begin{aligned} p_{ij}(\varphi^*_{ij}) &= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i}{\varphi^*_{ij}}\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\varphi^*_{ij}\right]^{\frac{1}{1+\gamma}} \\ &= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{1}{1+\gamma}} \tilde{w}_i (\varphi^*_{ij})^{\frac{-\gamma}{1+\gamma}}. \end{aligned} \tag{A.10}$$

From equation (A.2), we can express firm revenue as:

$$r_{ij}(\varphi) = p_{ij}(\varphi) q_{ij}(\varphi) = E_j P_j^{\sigma-1} b_i^{1-\sigma}\tau_{ij}^{1-\sigma} p_{ij}(\varphi)^{1-\sigma}.$$

Using this last result, we can express the zero-profit condition in equation (A.7) as:

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{E_j P_j^{\sigma-1} b_i^{1-\sigma}\tau_{ij}^{1-\sigma} p_{ij}(\varphi^*_{ij})^{1-\sigma}}{\sigma} = \tilde{w}_i f_{ij}. \tag{A.11}$$

Substituting for the factory-gate price in equation (A.11) using equation (A.10), we can solve for the zero-cutoff-profit productivity:

$$\tilde{w}_i f_{ij} = \left(\frac{\sigma+\gamma}{1+\gamma}\right) \frac{E_j P_j^{\sigma-1} b_i^{1-\sigma} \tau_{ij}^{1-\sigma} \left\{ \left(\frac{1+\gamma}{\gamma}\right) \left(\frac{\sigma}{\sigma-1}\right) \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{1}{1+\gamma}} \tilde{w}_i (\varphi_{ij}^*)^{\frac{-\gamma}{1+\gamma}} \right\}^{1-\sigma}}{\sigma}$$

$$\Rightarrow (\varphi_{ij}^*)^{(\sigma-1)\left(\frac{\gamma}{1+\gamma}\right)} = \left(\frac{1+\gamma}{\sigma+\gamma}\right) \left(\frac{\sigma \tilde{w}_i f_{ij}}{E_j P_j^{\sigma-1} b_i^{1-\sigma} \tau_{ij}^{1-\sigma}}\right) \left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right) \tilde{w}_i\right]^{\sigma-1} \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{\sigma-1}{1+\gamma}}$$

$$\Rightarrow \varphi_{ij}^* = \left[\frac{\left(\frac{1+\gamma}{\gamma} \frac{\sigma}{\sigma-1} \tilde{w}_i\right)^{\sigma}}{E_j P_j^{\sigma-1} b_i^{1-\sigma}}\right]^{\frac{1}{\frac{\gamma}{1+\gamma}(\sigma-1)}} \left[\frac{\gamma}{\sigma+\gamma}(\sigma-1) f_{ij}\right]^{\frac{\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}} \tau_{ij}^{\frac{1+\gamma}{\gamma}}. \qquad \text{(A.12)}$$

Again, when $\gamma \to \infty$ we obtain the benchmark result:

$$\lim_{\gamma\to\infty} \varphi_{ij}^* = \left[\left(\frac{\sigma}{\sigma-1}\right)^{\sigma} (\sigma-1) \frac{f_{ij} \tilde{w}_i^{\sigma}}{E_j P_j^{\sigma-1} b_i^{1-\sigma}}\right]^{\frac{1}{\sigma-1}} \tau_{ij} = \frac{\sigma^{1+\frac{1}{\sigma-1}} \tilde{w}_i^{1+\frac{1}{\sigma-1}} f_{ij}^{\frac{1}{\sigma-1}} b_i \tau_{ij}}{(\sigma-1) E_i^{\frac{1}{\sigma-1}} P_j}$$

$$= \left(\frac{\sigma}{\sigma-1}\right) \frac{\tilde{w}_i b_i \tau_{ij}}{P_j} \left(\frac{\sigma w_i f_{ij}}{E_j}\right)^{\frac{1}{\sigma-1}}.$$

A.4 Average Profits

In our model, the relationship between the relative revenues of two firms in country i serving the domestic market and their relative productivities is similar to – but nontrivially different from – the constant marginal cost case. From equation (A.2) and the pricing rule (A.5), we can express the ratio of output between any firm and the cutoff firm as follows

$$\frac{q_{ij}(\varphi)}{q_{ij}(\varphi_{ij}^*)} = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{\sigma\left(\frac{\gamma}{\sigma+\gamma}\right)}, \qquad \text{(A.13)}$$

which differs from the constant marginal cost case because of the extra term in the exponent (i.e., $\gamma/(\sigma+\gamma)$). However, when $\gamma \to \infty$ the result is the same as in Melitz (2003). Using equation (A.3) to define the ratio of prices and multiplying by the ratio of quantities to obtain relative revenues yields:

$$\frac{r_{ij}(\varphi)}{r_{ij}(\varphi_{ij}^*)} = \frac{p_{ij}(\varphi)}{p_{ij}(\varphi_{ij}^*)} \times \frac{q_{ij}(\varphi)}{q_{ij}(\varphi_{ij}^*)} = \left[\frac{q_{ij}(\varphi)^{\frac{1}{\gamma}}/\varphi}{q_{ij}(\varphi_{ij}^*)^{\frac{1}{\gamma}}/\varphi_{ij}^*}\right] \left[\frac{q_{ij}(\varphi)}{q_{ij}(\varphi_{ij}^*)}\right] = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} \qquad \text{(A.14)}$$

where the last equality follows from equation (A.13). Note that when $\gamma \to \infty$, the relationship is identical to the constant marginal cost case. The sufficient condition here for a positive relationship between productivity and revenue is $\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right) > 1$, instead of the typical assumption $\sigma > 1$.

From the zero-profit condition $\pi_{ij}(\varphi_{ij}^*) = 0$ and the definition of profits in equation (A.6), we have:

$$\pi_{ij}(\varphi_{ij}^*) = 0 \quad \Leftrightarrow \quad r_{ij}(\varphi_{ij}^*) = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\sigma\tilde{w}_i f_{ij}. \tag{A.15}$$

Using this result and equation (A.14), we obtain:

$$r_{ij}(\varphi) = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} r_{ij}(\varphi_{ij}^*) = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\left(\frac{\varphi}{\varphi_{ij}^*}\right)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} \sigma\tilde{w}_i f_{ij}, \tag{A.16}$$

which shows clearly that firm revenue is increasing in firm productivity. Using this last result, we can express average revenue for a country i firm selling to country j as:

$$\begin{aligned}
\overline{r}_{ij}(\varphi_{ij}^*) &= \int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi \\
&= \left(\frac{1+\gamma}{\sigma+\gamma}\right)\left(\frac{1}{\varphi_{ij}^*}\right)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} \sigma\tilde{w}_i f_{ij} \int_{\varphi_{ij}^*}^{\infty} \varphi^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\mu_{ij}(\varphi)d\varphi \\
&= \left(\frac{1+\gamma}{\sigma+\gamma}\right)\left[\frac{\tilde{\varphi}_{ij}(\varphi_{ij}^*)}{\varphi_{ij}^*}\right]^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} \sigma\tilde{w}_i f_{ij}
\end{aligned} \tag{A.17}$$

where

$$\mu_{ij}(\varphi) = \begin{cases} \frac{g(\varphi)}{1-G\left(\varphi_{ij}^*\right)} = \theta(\varphi_{ij}^*)^{\theta}\varphi^{-\theta-1}, & \text{if } \varphi \geq \varphi_{ij}^*, \\ 0 & \text{otherwise} \end{cases} \tag{A.18}$$

is the Pareto distribution of firm productivity, and

$$\tilde{\varphi}_{ij}(\varphi_{ij}^*) = \left[\int_{\varphi_{ij}^*}^{\infty} \varphi^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\mu_{ij}(\varphi)d\varphi\right]^{\left(\frac{1}{\sigma-1}\right)\frac{\sigma+\gamma}{\gamma}}. \tag{A.19}$$

defines an aggregate productivity level as a function of the cutoff level φ_{ij}^*.

Using equation (A.19), we can define average profit for each destination market as follows:

$$
\begin{aligned}
\overline{\pi}_{ij}(\varphi_{ij}^*) &= \int_{\varphi_{ij}^*}^{\infty} \pi_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = \int_{\varphi_{ij}^*}^{\infty} \left[\left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{r_{ij}(\varphi)}{\sigma} - \tilde{w}_i f_{ij}\right]\mu_{ij}(\varphi)d\varphi \\
&= \left(\frac{\sigma+\gamma}{1+\gamma}\right)\int_{\varphi_{ij}^*}^{\infty} \frac{r_{ij}(\varphi)}{\sigma}\mu_{ij}(\varphi)d\varphi - \tilde{w}_i f_{ij} = \left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{\overline{r}_{ij}(\varphi_{ij}^*)}{\sigma} - \tilde{w}_i f_{ij} \\
&= \left\{\left[\frac{\tilde{\varphi}_{ij}(\varphi_{ij}^*)}{\varphi_{ij}^*}\right]^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} - 1\right\}\tilde{w}_i f_{ij}. \qquad \text{(A.20)}
\end{aligned}
$$

This result is analogous to the zero-cutoff-profit condition in Melitz (2003), with $\overline{\pi}_{ij}$ a negative function of φ_{ij}^*. The nontrivial difference is the necessary condition that $\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right) > 1$.

By definition, the average profit of an incumbent firm is the sum of the average profits from sales to all markets:

$$
\overline{\pi}_i = \sum_{j=1}^{N}\left[\frac{1-G(\varphi_{ij}^*)}{1-G(\varphi_{ii}^*)}\right]\overline{\pi}_{ij}(\varphi_{ij}^*) = \sum_{j=1}^{N}\left(\frac{\varphi_{ij}^*}{\varphi_{ii}^*}\right)^{-\theta}\overline{\pi}_{ij}(\varphi_{ij}^*), \qquad \text{(A.21)}
$$

where the last equality follows from the Pareto distribution assumption. This expression includes domestic profits (i.e., when $i = j$). Using equation (A.20) in (A.21), we can express average total firm profit (under the Pareto distribution assumption) as:

$$
\overline{\pi}_i = \sum_{j=1}^{N}\left(\frac{\varphi_{ij}^*}{\varphi_{ii}^*}\right)^{-\theta}\left\{\left[\frac{\tilde{\varphi}_{ij}(\varphi_{ij}^*)}{\varphi_{ij}^*}\right]^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} - 1\right\}\tilde{w}_i f_{ij}. \qquad \text{(A.22)}
$$

We can further simplify this expression using the definition of average productivity in equation (A.19), which implies that:

$$
\begin{aligned}
\left[\tilde{\varphi}_{ij}(\varphi_{ij}^*)\right]^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} &= \int_{\varphi_{ij}^*}^{\infty}\varphi^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\mu_{ij}(\varphi)d\varphi = \int_{\varphi_{ij}^*}^{\infty}\varphi^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\frac{\theta\varphi^{-\theta-1}}{(\varphi_{ij}^*)^{-\theta}}d\varphi \\
&= \theta(\varphi_{ij}^*)^{\theta}\int_{\varphi_{ij}^*}^{\infty}\varphi^{\sigma\left(\frac{\gamma+1}{\sigma+\gamma}\right)-\theta-2}d\varphi = \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right](\varphi_{ij}^*)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}. \qquad \text{(A.23)}
\end{aligned}
$$

Using this last result in equation (A.22) yields:

$$\overline{\pi}_i = \sum_{j=1}^{N}\left(\frac{\varphi_{ij}^*}{\varphi_{ii}^*}\right)^{-\theta}\left\{\left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]-1\right\}\tilde{w}_i f_{ij} = \frac{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\sum_{j=1}^{N}\left(\frac{\varphi_{ii}^*}{\varphi_{ij}^*}\right)^{\theta}\tilde{w}_i f_{ij}. \tag{A.24}$$

A.5 Masses of Firms

Consumers have no taste for leisure, so the supply of labor is fixed at L_i. There are three sources of demand for labor: labor for entry costs (f^e), labor for fixed trade costs (f_{ij}), and labor for production. Therefore, the labor-market-clearing condition is given by:

$$L_i = \frac{M_i^e f^e}{A_i} + \sum_{j=1}^{N} M_{ij}\int_{\varphi_{ij}^*}^{\infty}\frac{1}{A_i}\left[f_{ij}+\frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}\right]\mu_{ij}(\varphi)d\varphi, \tag{A.25}$$

where M_i^e is the mass of firms attempting to enter the industry in country i, M_{ij} is the mass of firms based in i that serve market j, and

$$\mu_{ij}(\varphi) = \begin{cases} \frac{g(\varphi)}{1-G(\varphi_{ij}^*)} = \theta(\varphi_{ij}^*)^{\theta}\varphi^{-\theta-1}, & \text{if } \varphi \geq \varphi_{ij}^*, \\ 0 & \text{otherwise} \end{cases} \tag{A.26}$$

is the Pareto distribution of firms' productivities.

Multiplying both sides of equation (A.25) by w_i, yields:

$$w_i L_i = \tilde{w}_i M_i^e f^e + \tilde{w}_i\sum_{j=1}^{N} M_{ij} f_{ij} + \tilde{w}_i\sum_{j=1}^{N} M_{ij}\int_{\varphi_{ij}^*}^{\infty}\frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}\mu_{ij}(\varphi)d\varphi. \tag{A.27}$$

From the optimal revenue equation (A.4), we can show that:

$$\frac{\tilde{w}_i q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} = \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right) r_{ij}(\varphi).$$

Using this result in equation (A.27) yields:

$$w_i L_i = \tilde{w}_i M_i^e f^e + \tilde{w}_i\sum_{j=1}^{N} M_{ij} f_{ij} + \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\sum_{j=1}^{N} M_{ij}\int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi. \tag{A.28}$$

As in Feenstra (2010) and Redding (2011), zero expected profits imply that aggregate revenue is equal to expenditure such that:

$$w_i L_i = \sum_{j=1}^{N} M_{ij} \int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi) \mu_{ij}(\varphi) d\varphi. \tag{A.29}$$

Substituting with this result for the last term on the right-hand-side of equation (A.28) yields:

$$\begin{aligned} w_i L_i &= \tilde{w}_i M_i^e f^e + \tilde{w}_i \sum_{j=1}^{N} M_{ij} f_{ij} + \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right) w_i L_i \\ \Leftrightarrow \quad \left[1 - \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\right] w_i L_i &= \tilde{w}_i M_i^e f^e + \tilde{w}_i \sum_{j=1}^{N} M_{ij} f_{ij}. \end{aligned} \tag{A.30}$$

Substituting the left-hand-side of equation (A.30) for the first two terms on the right-hand-side of equation (A.27) yields:

$$\begin{aligned} w_i L_i &= \left[1 - \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\right] w_i L_i + \tilde{w}_i \sum_{j=1}^{N} M_{ij} \int_{\varphi_{ij}^*}^{\infty} \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \mu_{ij}(\varphi) d\varphi \\ \Leftrightarrow \quad \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right) w_i L_i &= \tilde{w}_i \sum_{j=1}^{N} M_{ij} \int_{\varphi_{ij}^*}^{\infty} \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \mu_{ij}(\varphi) d\varphi \end{aligned} \tag{A.31}$$

We can now solve for $\tilde{w}_i \sum_{j=1}^{N} M_{ij} f_{ij}$. From equation (A.13), we can express the output for any firm as a function of the output of the cutoff firm as follows:

$$q_{ij}(\varphi) = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{\frac{\sigma\gamma}{\sigma+\gamma}} q_{ij}(\varphi_{ij}^*). \tag{A.32}$$

Using this result and the Pareto distribution, we can solve the integral on the right-hand-side of equation (A.31):

$$\begin{aligned}
\int_{\varphi_{ij}^*}^{\infty} \frac{q(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \mu_{ij}(\varphi) d\varphi &= \int_{\varphi_{ij}^*}^{\infty} \frac{\left[q(\varphi_{ij}^*)\left(\frac{\varphi}{\varphi_{ij}^*}\right)^{\frac{\sigma\gamma}{\sigma+\gamma}}\right]^{\frac{1+\gamma}{\gamma}}}{\varphi} \mu_{ij}(\varphi) d\varphi \\
&= \int_{\varphi_{ij}^*}^{\infty} \frac{q(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}} \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right)}}{\varphi} \left[\frac{\theta\varphi^{-\theta-1}}{(\varphi_{ij}^*)^{-\theta}}\right] d\varphi \\
&= q(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}} \left(\frac{1}{\varphi_{ij}^*}\right)^{\sigma\frac{1+\gamma}{\sigma+\gamma}-\theta} \theta \int_{\varphi_{ij}^*}^{\infty} \varphi^{\frac{\gamma}{\sigma+\gamma}(\sigma-1)-(\theta+1)} d\varphi \\
&= q(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}} \left(\frac{1}{\varphi_{ij}^*}\right)^{\sigma\frac{1+\gamma}{\sigma+\gamma}-\theta} \left[\frac{\theta}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right] \left[\left(\frac{1}{\infty}\right)^{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)} - (\varphi_{ij}^*)^{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right] \\
&= \left[\frac{\theta}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right] q(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}} (\varphi_{ij}^*)^{\frac{\gamma}{\sigma+\gamma}(\sigma-1)-\sigma\frac{1+\gamma}{\sigma+\gamma}} \\
&= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right] \frac{q(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}}}{\varphi_{ij}^*}. \qquad \text{(A.33)}
\end{aligned}$$

Importantly, note that, for a positive integral, we require only that $\theta > \frac{\gamma}{\sigma+\gamma}(\sigma-1)$ and not $\theta > \sigma - 1$, as in the standard constant marginal cost Melitz models.

Rearranging equation (A.9), we can show that:

$$\frac{q_{ij}(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}}}{\varphi_{ij}^*} = \left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)f_{ij}.$$

Using this result in the equation just above it yields:

$$\int_{\varphi_{ij}^*}^{\infty} \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \mu_{ij}(\varphi) d\varphi = \left[\frac{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right] f_{ij}, \qquad \text{(A.34)}$$

which implies that:

$$\tilde{w}_i \sum_{j=1}^{N} M_{ij} \int_{\varphi_{ij}^*}^{\infty} \frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi} \mu_{ij}(\varphi) d\varphi = \left[\frac{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right] \tilde{w}_i \sum_{j=1}^{N} M_{ij} f_{ij}. \qquad \text{(A.35)}$$

Substituting with this last result into equation (A.31) yields:

$$\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)w_iL_i = \left[\frac{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right]\tilde{w}_i\sum_{j=1}^{N}M_{ij}f_{ij}$$

$$\Rightarrow \quad \tilde{w}_i\sum_{j=1}^{N}M_{ij}f_{ij} = \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\left[\frac{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}\right]w_iL_i. \tag{A.36}$$

Substituting this result into equation (A.35) yields:

$$\tilde{w}_i\sum_{j=1}^{N}M_{ij}\int_{\varphi_{ij}^*}^{\infty}\frac{q_{ij}(\varphi)}{\varphi}^{\frac{1+\gamma}{\gamma}}\mu_{ij}(\varphi)d\varphi = \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)w_iL_i. \tag{A.37}$$

We can now solve for M_i^e. Substituting equations (A.36) and (A.37) into equation (A.27), and eliminating out the w_i, yields:

$$\begin{aligned}
L_i &= \frac{1}{A_i}M_i^ef^e + \frac{1}{A_i}\sum_{j=1}^{N}M_{ij}f_{ij} + \frac{1}{A_i}\sum_{j=1}^{N}M_{ij}\int_{\varphi_{ij}^*}^{\infty}\frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}\mu_{ij}(\varphi)d\varphi \\
&= \frac{M_i^ef^e}{A_i} + \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\left[\frac{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}\right]L_i + \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)L_i \\
&= \frac{M_i^ef^e}{A_i} + \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\left[1+\frac{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}\right]L_i \\
&= \frac{M_i^ef^e}{A_i} + \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\left[\frac{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)+\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}\right]L_i \\
&= \frac{M_i^ef^e}{A_i} + \left[\frac{(\theta-1)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)+\theta}{\theta\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right)}\right]L_i
\end{aligned}$$

which implies that:

$$
\begin{aligned}
M_i^e &= \left[1 - \frac{(\theta-1)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)+\theta}{\theta\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right)}\right]\frac{A_iL_i}{f^e} \\
&= \left[\frac{\theta\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right) - (\theta-1)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)-\theta}{\theta\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right)}\right]\frac{A_iL_i}{f^e} \\
&= \left(\frac{1}{\theta\sigma}\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)\left(\frac{\theta\sigma+\theta\sigma\gamma-\theta\sigma\gamma+\theta\gamma+\gamma\sigma-\gamma-\theta\gamma-\theta\sigma}{\sigma+\gamma}\right)\frac{A_iL_i}{f^e} \\
&= \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_iL_i}{\theta f^e}. \qquad \text{(A.38)}
\end{aligned}
$$

We now solve for M_{ii}. As standard, we assume a fraction δ of existing firms M_{ii} exit the industry. In a steady state equilibrium, the mass of new entrant (M_i^e) must replace firms hit by the exogenous shock and forced to exit the industry. Hence, in a steady state:

$$[1 - G(\varphi_{ii}^*)]M_i^e = \delta M_{ii} \qquad \text{(A.39)}$$

where $[1 - G(\varphi_{ii}^*)] = (\varphi_{ii}^*)^{-\theta}$ is the probability of successful entry. It follows that:

$$M_{ii} = \frac{[1 - G(\varphi_{ii}^*)]M_i^e}{\delta} = \frac{M_i^e}{\delta(\varphi_{ii}^*)^\theta} = \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_iL_i}{\theta\delta f^e(\varphi_{ii}^*)^\theta} \qquad \text{(A.40)}$$

Finally, we can solve for the mass of exporting firms M_{ij}. A successful entrant in country-i will export to country j if it is is productive enough to be profitable in the foreign country. This implies that:

$$M_{ij} = \left[\frac{1 - G(\varphi_{ij}^*)}{1 - G(\varphi_{ii}^*)}\right]M_{ii} = \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_iL_i}{\theta\delta f^e(\varphi_{ij}^*)^\theta}. \qquad \text{(A.41)}$$

A.6 Price Index

In this section, we solve for the price index. Substituting equation (A.2) into optimal pricing rule (A.5) we obtain:

$$
\begin{aligned}
p^c_{ij}(\varphi) &= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i\tau_{ij}}{\varphi}q_{ij}(\varphi)^{\frac{1}{\gamma}} \\
&= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i\tau_{ij}}{\varphi}\left[E_j P_j^{\sigma-1} b_i^{1-\sigma} p^c_{ij}(\varphi)^{-\sigma}\right]^{\frac{1}{\gamma}} \\
&= \left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i\tau_{ij}}{\varphi}\right]^{\frac{\gamma}{\sigma+\gamma}} E_j^{\frac{1}{\sigma+\gamma}} P_j^{\frac{\sigma-1}{\sigma+\gamma}} b_i^{\frac{1-\sigma}{\sigma+\gamma}}
\end{aligned} \tag{A.42}
$$

Substituting this result into the definition of the price index

$$
P_j = \left[\int_{\nu\in\Omega_j} b_i^{1-\sigma} p^c_j(\nu)^{1-\sigma} d\nu\right]^{\frac{1}{1-\sigma}}, \tag{A.43}
$$

and rearranging, we obtain:

$$
\begin{aligned}
P_j^{1-\sigma} &= \int_{\nu\in\Omega_j} b_i^{1-\sigma} p^c_j(\nu)^{1-\sigma} d\nu = \sum_i M_{ij}\int_{\varphi^*_{ij}}^{\infty} b_i^{1-\sigma} p^c_{ij}(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi \\
&= \sum_i M_{ij}\int_{\varphi^*_{ij}}^{\infty}\left\{\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i\tau_{ij}}{\varphi}\right]^{\frac{\gamma}{\sigma+\gamma}} E_j^{\frac{1}{\sigma+\gamma}} P_j^{\frac{\sigma-1}{\sigma+\gamma}} b_i^{\frac{1-\sigma}{\sigma+\gamma}}\right\}^{1-\sigma}\mu_{ij}(\varphi)d\varphi \\
&= \sum_i M_{ij}\left\{\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\tilde{w}_i\tau_{ij}\right]^{\frac{\gamma}{\sigma+\gamma}} E_j^{\frac{1}{\sigma+\gamma}} P_j^{\frac{\sigma-1}{\sigma+\gamma}} b_i^{\frac{1-\sigma}{\sigma+\gamma}}\right\}^{1-\sigma}\int_{\varphi^*_{ij}}^{\infty}\varphi^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\mu_{ij}(\varphi)d\varphi \\
&= \sum_i M_{ij}\left\{\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i\tau_{ij}}{\varphi^*_{ij}}\right]^{\frac{\gamma}{\sigma+\gamma}} E_j^{\frac{1}{\sigma+\gamma}} P_j^{\frac{\sigma-1}{\sigma+\gamma}} b_i^{\frac{1-\sigma}{\sigma+\gamma}}\right\}^{1-\sigma}\left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right] \\
&= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\sum_i M_{ij} b_i^{1-\sigma}[p^c_{ij}(\varphi^*_{ij})]^{1-\sigma} \\
&= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\sum_i M_{ij}\tau_{ij}^{1-\sigma} b_i^{1-\sigma}[p_{ij}(\varphi^*_{ij})]^{1-\sigma}.
\end{aligned} \tag{A.44}
$$

We can use the productivity cutoff in equation (A.12) and the mass of firms in equation (A.41) to obtain an expression also for the price index P_j as a function of the endogenous wages and parameters of the model, given in equation (A.87) below.

A.7 Trade Flows

Using the pricing rule (A.3), the result in equation (A.34), and equation (A.41) for the mass of firms, we can express trade flows as:

$$
\begin{aligned}
X_{ij} &\equiv M_{ij}\int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = M_{ij}\int_{\varphi_{ij}^*}^{\infty} p_{ij}(\varphi)q_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi \\
&= M_{ij}\int_{\varphi_{ij}^*}^{\infty}\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{\tilde{w}_i}{\varphi}q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}\mu_{ij}(\varphi)d\varphi \\
&= M_{ij}\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\tilde{w}_i\int_{\varphi_{ij}^*}^{\infty}\frac{q_{ij}(\varphi)^{\frac{1+\gamma}{\gamma}}}{\varphi}\mu_{ij}(\varphi)d\varphi \\
&= M_{ij}\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\frac{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right]\tilde{w}_i f_{ij} \\
&= \left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_iL_i}{\theta\delta f^e(\varphi_{ij}^*)^{\theta}}\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\frac{\theta\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right]\frac{w_i f_{ij}}{A_i} \\
&= \left[\frac{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\frac{w_iL_if_{ij}}{\delta f^e(\varphi_{ij}^*)^{\theta}}. \qquad \text{(A.45)}
\end{aligned}
$$

By definition, aggregate expenditure in country j is given by:

$$
E_j = \sum_k X_{kj} = \left[\frac{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\frac{1}{\delta f^e}\sum_k w_kL_kf_{kj}(\varphi_{kj}^*)^{-\theta}. \qquad \text{(A.46)}
$$

Therefore, the share of country j's expenditure on goods supplied by country i is given by:

$$
\lambda_{ij} \equiv \frac{X_{ij}}{E_j} = \frac{w_iL_if_{ij}(\varphi_{ij}^*)^{-\theta}}{\sum_{k=1}^N w_kL_kf_{kj}(\varphi_{kj}^*)^{-\theta}}. \qquad \text{(A.47)}
$$

Adapting equation (A.12), we know:

$$
\varphi_{kj}^* = \left[\frac{\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}\tilde{w}_k\right)^{\sigma}}{E_jP_j^{\sigma-1}b_i^{1-\sigma}}\right]^{\frac{1}{\frac{\gamma}{1+\gamma}(\sigma-1)}}\left[\frac{\gamma}{\sigma+\gamma}(\sigma-1)f_{kj}\right]^{\frac{\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}\tau_{kj}^{\frac{1+\gamma}{\gamma}}.
$$

Substituting this equation for φ_{kj}^* and equation (A.12) for φ_{ij}^* into equation (A.47), we obtain bilateral trade from i to j as a share of j's expenditures (λ_{ij}):

$$
\begin{aligned}
\lambda_{ij} &= \frac{w_i L_i f_{ij} \left[\tilde{w}_i^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{\frac{1+\gamma}{\gamma}} \tau_{ij}^{\frac{1+\gamma}{\gamma}} f_{ij}^{\left(\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right)} \right]^{-\theta}}{\sum_{k=1}^{N} w_k L_k f_{kj} \left[\tilde{w}_k^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_k^{\frac{1+\gamma}{\gamma}} \tau_{kj}^{\frac{1+\gamma}{\gamma}} f_{kj}^{\left(\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right)} \right]^{-\theta}} \\
&= \frac{A_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_k w_k^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}.
\end{aligned}
\tag{A.48}
$$

A.8 Wage Rates

We now determine aggregate revenue in equilibrium. First, total payments to production workers, which we denote L_i^p, must be equal to the difference between aggregate revenue and aggregate profit such that $w_i L_i^p = R_i - \Pi_i$, where $\Pi_i \equiv M_{ii}\overline{\pi}_i$. Second, in equilibrium, the mass of successful entrants must be equal to the mass of firms forced to exit the industry. This aggregate stability condition requires that $[1 - G(\varphi_{ij}^*)]M_i^e = \delta M_{ii}$. Combining this last result with the free entry condition (A.53) (provided later) implies that total payments to labor used in entry equal total profits: $w_i L_i^e = w_i M_i^e f^e = \Pi_i$. It follows that aggregate revenue, which is the sum of total payments to labor and profits, is equal to payroll $R_i = w_i L_i^p + \Pi_i = w_i L_i$.

The equilibrium wage rate (w_i) in each country can be determined from the requirement that total revenue equals total expenditure on goods produced there:

$$
w_i L_i = \sum_{j=1}^{N} \lambda_{ij} w_j L_j.
$$

Substituting in equation (A.48) yields the following system of N equations (one for each of N countries):

$$
w_i L_i = \sum_{j=1}^{N} \left[\frac{A_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_k w_k^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_k^{-\theta\frac{1+\gamma}{\gamma}} \tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}} \right] w_j L_j
\tag{A.49}
$$

Equation (A.49) implies a system of N equations in the N unknown wage rates in each country, w_i. Note that this equation takes the same form as equation (3.14) on p. 1734 of

Alvarez and Lucas (2007). Using equation (A.49), we can define the following excess demand system:

$$\Xi(\boldsymbol{w}) = \frac{1}{w_i}\left[\sum_{j=1}^{N} \frac{A_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\frac{1+\gamma}{\gamma}} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_k w_k^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_k^{-\theta\frac{1+\gamma}{\gamma}} \tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}} - w_i L_i\right] \tag{A.50}$$

where $\boldsymbol{w}$ denotes the vector of wage rates across countries.

Proposition 1. *There exists a unique wage-rate vector* $\boldsymbol{w} \in \mathbb{R}^N_{++}$ *such that* $\Xi(\boldsymbol{w}) = 0$.

Proof. Note that $\Xi(\boldsymbol{w})$ has the following properties:

1. $\Xi(\boldsymbol{w})$ is continuous (by assumption on the parameters).

2. $\Xi(\boldsymbol{w})$ is homogenous of degree zero.

3. $\boldsymbol{w} \cdot \Xi(\boldsymbol{w}) = 0$ for all $\boldsymbol{w} \in \mathbb{R}^N_{++}$ (Walras Law).

4. There exists a constant $s > 0$ such that $\Xi_i(\boldsymbol{w}) > -s$ for each country i and all $\boldsymbol{w} \in \mathbb{R}^N_{++}$.

5. If $\boldsymbol{w}^m \to \boldsymbol{w}^0$ where $\boldsymbol{w}^0 \neq 0$ and $w_i^0 = 0$ for some country i, then $\max_j\{\Xi_j(\boldsymbol{w})\} \to \infty$.

6. $\Xi(\boldsymbol{w})$ satisfies the gross substitutes property

$$\frac{\partial \Xi_i(\boldsymbol{w})}{\partial w_j} > 0,\ i \neq j, \quad \text{and} \quad \frac{\partial \Xi_i(\boldsymbol{w})}{\partial w_i} < 0, \forall\, \boldsymbol{w} \in \mathbb{R}^N_{++}.$$

Under these conditions, Propositions 17.C.1 and 17.F.3 of Mas-Colell et al. (1995) or Theorems 1-3 of Alvarez and Lucas (2007) hold, such that there exists a unique vector of wage rates $\boldsymbol{w} \in \mathbb{R}^N_{++}$ that satisfies the clearing conditions $\Xi(\boldsymbol{w}) = 0$. □

A.9 Free Entry

There is an unbounded set of potential entrants in the industry. To enter the industry, firms must incur a fixed entry cost of f^e units of labor. That sunk entry cost provides the firm with a blue print for a unique variety and also reveals the firm's productivity, φ, a random draw from a common distribution $G(\varphi)$. Once the fixed entry cost is paid, firms can begin production.

The value of a successful entrant with productivity φ is equal to the discounted sum of lifetime profits. Following Melitz (2003), we assume that each period there is a probability

$\delta \in (0,1)$ that an incumbent firm will be hit by an adverse shock and be forced to exit the industry. In that case, the value of a successful entrant in the industry can be expressed as:

$$V_i(\varphi) = \sum_{t=1}^{\infty}(1-\delta)^t \pi_{it}(\varphi) = \frac{\pi_i(\varphi)}{\delta}, \tag{A.51}$$

where the second equality follows from the fact that profits are constant throughout the lifetime of the firm, i.e., $\pi_{it}(\varphi) = \pi_i(\varphi)$. Therefore, the value of entry as a function of productivity is given by:

$$V_i(\varphi) = \max\left\{0, \frac{\pi_i(\varphi)}{\delta}\right\}. \tag{A.52}$$

Firms with productivity above the domestic cutoff, φ^*_{ii}, will generate enough variable profits to cover the fixed costs. As a result, they stay in the industry and earn a lifetime profit proportional to their per-period profits. Firms with productivity lower than the domestic cutoff would earn negative profits if they remain in the industry. Hence, they prefer to exit the industry and get a null return.

In a free entry equilibrium, the expected value of entry, V^e_i, must be equal to the cost of entry such that:

$$V^e_i = [1 - G(\varphi^*_{ii})]\frac{\overline{\pi}_i}{\delta} = \tilde{w}_i f^e. \tag{A.53}$$

The expected value of entry is defined as the product of the probability of successful entry, $1 - G(\varphi^*_{ii})$, and the lifetime profits of the average incumbent firm, $\overline{\pi}_i/\delta$. The cost of entry is defined as the product of $\tilde{w}_i$ and the fixed entry cost, f^e, defined in units of labor.

By definition, the average profit of an incumbent firm is the sum of the average profits from sales to each market (including the domestic market) multiplied by the probability of entering each market conditional on producing for the domestic market:

$$\overline{\pi}_i = \sum_{j=1}^{N}\left[\frac{1 - G(\varphi^*_{ij})}{1 - G(\varphi^*_{ii})}\right]\overline{\pi}_{ij}(\varphi^*_{ij}). \tag{A.54}$$

To obtain an analytical solution, we follow the literature and assume that the productivity distribution is Pareto, such that $G(\varphi) = 1 - \varphi^{-\theta}$. We can combine the zero-cutoff-profit condition $\pi_{ij}(\varphi^*_{ij}) = 0$, the optimal pricing function in equation (6), and the definition of profits in equation (5), to express average total firm profit as:

$$\overline{\pi}_i = \frac{(\sigma-1)\frac{\gamma}{\sigma+\gamma}}{\theta - (\sigma-1)\frac{\gamma}{\sigma+\gamma}} \sum_{j=1}^{N}\left(\frac{\varphi^*_{ii}}{\varphi^*_{ij}}\right)^{\theta} \tilde{w}_i f_{ij}. \tag{A.55}$$

Substituting this last result for average profits into equation (A.53), we obtain an expression for the free-entry condition that depends only on the productivity cutoffs and parameters of the model:

$$V_i^e = \frac{(\sigma - 1)\frac{\gamma}{\sigma+\gamma}}{\theta - (\sigma - 1)\frac{\gamma}{\sigma+\gamma}} \sum_{j=1}^{N} \frac{f_{ij}}{(\varphi_{ij}^*)^\theta} = \delta f^e, \tag{A.56}$$

where the wage rates have canceled in the expression above. This result shows that the value of entry is proportionate to fixed entry costs (f^e).

A.10 General Equilibrium

As in Bernard et al. (2011), we determine general equilibrium using the recursive structure of the model. The system of equations (A.50) determines a unique equilibrium wage in each country (w_i). Furthermore, the mass of entrants M_i^e is determined as a function of parameters in equation (A.38). With these two equilibrium components, we can solve for all the other endogenous variables as follows. The price index P_j follows from the wage rate as explained in section A.6. The productivity cutoffs then follow from equation (9), the wage rates, the price indexes, and that $E_i = R_i = w_i L_i$ in equilibrium. The mass of firms in each country i serving each destination country j, M_{ij}, follows from equation (11) and the productivity cutoffs. Finally, the trade shares λ_{ij} follow directly from equation (A.47), the wage rates, and the productivity cutoffs. This completes the characterization of the general equilibrium.

A.11 Structural Gravity

In this section, we show how to derive the structural gravity equation from our theoretical model. Substituting equation (A.12) for the ZCP productivity threshold in the solution for bilateral trade flows in equation (A.45), we can solve for:

$$X_{ij} = BA_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i \left(E_j P_j^{\sigma-1}\right)^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta}{\sigma-1}\right)} w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}. \tag{A.57}$$

where B is a constant and a function of parameters $\sigma, \gamma, \theta, \delta$, and f^e. By the definition of revenue, it follows that:

$$R_i = \sum_{j=1}^{N} X_{ij} = BA_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tilde{\Pi}_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tag{A.58}$$

where

$$\tilde{\Pi}_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \equiv \tilde{\Pi}_i^{-\varepsilon_\tau} = \sum_{j=1}^{N} \left(E_j^{\frac{1}{\sigma-1}} P_j\right)^{\theta\left(\frac{1+\gamma}{\gamma}\right)} \phi_{ij} \equiv \sum_{j=1}^{N} \left(E_j^{\frac{1}{\sigma-1}} P_j\right)^{\varepsilon_\tau} \phi_{ij} \tag{A.59}$$

and

$$\phi_{ij} \equiv \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}.$$

Rearranging equation (A.58) to solve for $R_i\tilde{\Pi}_i^{\varepsilon_\tau}$ (where $\varepsilon_\tau \equiv \theta\frac{1+\gamma}{\gamma}$) yields:

$$R_i\tilde{\Pi}_i^{\varepsilon_\tau} = BA_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}. \tag{A.60}$$

Substituting the LHS term from above for the RHS term into equation (A.57) yields:

$$X_{ij} = \frac{R_i}{\tilde{\Pi}_i^{-\varepsilon_\tau}} \left(E_j^{\frac{1}{\sigma-1}} P_j\right)^{\varepsilon_\tau} \phi_{ij}. \tag{A.61}$$

We can define $\tilde{\Phi}_j$ such that:

$$E_j\tilde{\Phi}_j^{\varepsilon_\tau} \equiv \left(E_j^{\frac{1}{\sigma-1}} P_j\right)^{\varepsilon_\tau}. \tag{A.62}$$

Substituting the LHS term for the RHS term in equation (A.61) yields:

$$X_{ij} = \frac{R_i}{\tilde{\Pi}_i^{-\varepsilon_\tau}} \frac{E_j}{\tilde{\Phi}_j^{-\varepsilon_\tau}} \phi_{ij}. \tag{A.63}$$

Using the definition of $\tilde{\Phi}_j$ in equation (A.62), we can rewrite the multilateral resistance term $\tilde{\Pi}_i$ using equation (A.59) as follows:

$$\tilde{\Pi}_i^{-\varepsilon_\tau} = \sum_{j=1}^{N} \frac{E_j}{\tilde{\Phi}_j^{-\varepsilon_\tau}} \phi_{ij}. \tag{A.64}$$

Finally, by definition of expenditure and equation (A.63) it follows that:

$$E_j = \sum_{i=1}^{N} X_{ij} = \sum_{i=1}^{N} \frac{R_i}{\tilde{\Pi}_i^{-\varepsilon_\tau}} \frac{E_j}{\tilde{\Phi}_j^{-\varepsilon_\tau}} \phi_{ij}. \tag{A.65}$$

This result implies that:

$$\tilde{\Phi}_j^{-\varepsilon_\tau} = \sum_{i=1}^{N} \frac{R_i}{\tilde{\Pi}_i^{-\varepsilon_\tau}} \phi_{ij}. \tag{A.66}$$

If we define $\Pi_i \equiv \tilde{\Pi}_i^{-\varepsilon_\tau}$ and $\Phi_j \equiv \tilde{\Phi}_j^{-\varepsilon_\tau}$, the system of equations (A.63), (A.64), and (A.66) forms a structural gravity-equation equivalent to equation (2) in Head and Mayer (2014).

A.12 Elasticity of Trade with respect to *Ad Valorem* Variable Trade Costs

First, we determine the elasticity of trade with respect to *ad valorem* variable trade costs. By definition, aggregate bilateral trade flows are given by:

$$X_{ij} \equiv M_{ij} \int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = M_{ij}[1 - G(\varphi_{ij}^*)]^{-1} \int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)g(\varphi)d\varphi. \tag{A.67}$$

It follows that:

$$\begin{aligned}\frac{\partial X_{ij}}{\partial \tau_{ij}} = \frac{\partial M_{ij}}{\partial \tau_{ij}}\frac{X_{ij}}{M_{ij}} &+ M_{ij}[1 - G(\varphi_{ij}^*)]^{-2}\frac{\partial G(\varphi_{ij}^*)}{\partial \varphi}\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}}[1 - G(\varphi_{ij}^*)]\frac{X_{ij}}{M_{ij}} \\ &- M_{ij}[1 - G(\varphi_{ij}^*)]^{-1} r_{ij}(\varphi_{ij}^*)g(\varphi_{ij}^*)\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}} \\ &+ M_{ij}[1 - G(\varphi_{ij}^*)]^{-1}\int_{\varphi_{ij}^*}^{\infty}\frac{\partial r_{ij}(\varphi)}{\partial \tau_{ij}}g(\varphi)d\varphi. \end{aligned} \tag{A.68}$$

From this last result, it is straightforward to define the elasticity as follows:

$$\begin{aligned}\varepsilon_\tau \equiv -\frac{\partial X_{ij}}{\partial \tau_{ij}}\frac{\tau_{ij}}{X_{ij}} &= \frac{\partial M_{ij}}{\partial \tau_{ij}}\frac{X_{ij}}{M_{ij}}\frac{\tau_{ij}}{X_{ij}} + M_{ij}[1 - G(\varphi_{ij}^*)]^{-2}\frac{\partial G(\varphi_{ij}^*)}{\partial \varphi}\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}}[1 - G(\varphi_{ij}^*)]\frac{X_{ij}}{M_{ij}}\frac{\tau_{ij}}{X_{ij}} \\ &\quad - M_{ij}\frac{\tau_{ij}}{X_{ij}}[1 - G(\varphi_{ij}^*)]^{-1} r_{ij}(\varphi_{ij}^*)g(\varphi_{ij}^*)\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}} \\ &\quad + M_{ij}\frac{\tau_{ij}}{X_{ij}}[1 - G(\varphi_{ij}^*)]^{-1}\int_{\varphi_{ij}^*}^{\infty}\frac{\partial r_{ij}(\varphi)}{\partial \tau_{ij}}g(\varphi)d\varphi \\ &= -\Bigg\{\underbrace{\frac{\partial M_{ij}}{\partial \tau_{ij}}\frac{\tau_{ij}}{M_{ij}}}_{\text{extensive}} + \underbrace{\frac{g(\varphi_{ij}^*)\varphi_{ij}^*}{1 - G(\varphi_{ij}^*)}\left[1 - \frac{r_{ij}(\varphi_{ij}^*)}{X_{ij}/M_{ij}}\right]\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}}\frac{\tau_{ij}}{\varphi_{ij}^*}}_{\text{compositional}} \\ &\quad + \underbrace{\int_{\varphi_{ij}^*}^{\infty}\frac{\partial r_{ij}(\varphi)}{\partial \tau_{ij}}\frac{\tau_{ij}}{X_{ij}/M_{ij}}\mu_{ij}(\varphi)d\varphi}_{\text{intensive}}\Bigg\}, \end{aligned} \tag{A.69}$$

where the last equality follows from simplifying and rearranging terms.

We now calculate each component of equation (A.69) separately. From equation (9), we have:

$$\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}} = \left(\frac{1+\gamma}{\gamma}\right)\frac{\varphi_{ij}^*}{\tau_{ij}}, \tag{A.70}$$

which implies that:

$$\frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}} \frac{\tau_{ij}}{\varphi_{ij}^*} = \frac{1+\gamma}{\gamma}. \tag{A.71}$$

Using equations (11) and (A.70), we have:

$$\frac{\partial M_{ij}}{\partial \tau_{ij}} = -\theta \left(\frac{M_{ij}}{\varphi_{ij}^*}\right) \frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}} = -\theta \left(\frac{M_{ij}}{\varphi_{ij}^*}\right) \left(\frac{1+\gamma}{\gamma}\right) \frac{\varphi_{ij}^*}{\tau_{ij}} = -\theta \left(\frac{1+\gamma}{\gamma}\right) \frac{M_{ij}}{\tau_{ij}}.$$

This last result implies that:

$$\frac{\partial M_{ij}}{\partial \tau_{ij}} \frac{\tau_{ij}}{M_{ij}} = -\theta \left(\frac{1+\gamma}{\gamma}\right). \tag{A.72}$$

Under the Pareto distribution assumption it follows that:

$$\frac{g(\varphi_{ij}^*)\varphi_{ij}^*}{1 - G(\varphi_{ij}^*)} = \frac{\theta(\varphi_{ij}^*)^{-\theta-1}\varphi_{ij}^*}{(\varphi_{ij}^*)^{-\theta}} = \theta, \tag{A.73}$$

where the last equality uses equation (A.70).

Next, using the solution for the equilibrium mass of firms in equation (11) and cutoff-firm revenue:

$$r_{ij}(\varphi_{ij}^*) = \left(\frac{1+\gamma}{\sigma+\gamma}\right) \sigma \tilde{w}_i f_{ij}, \tag{A.74}$$

which, as shown in section 4 of Appendix A, is obtained from the zero profit condition, we can show that:

$$1 - \frac{r_{ij}(\varphi_{ij}^*)}{X_{ij}/M_{ij}} = 1 - \frac{1}{\theta}\left[\theta - \left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)\right] = \frac{1}{\theta}\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1). \tag{A.75}$$

Finally, as shown in section A.4, it is possible to express firm revenue as a function of the cutoff productivity as follows:

$$r_{ij}(\varphi) = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{(\sigma-1)\frac{\gamma}{\sigma+\gamma}} r_{ij}(\varphi_{ij}^*) = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{(\sigma-1)\frac{\gamma}{\sigma+\gamma}} \sigma \tilde{w}_i f_{ij}.$$

Using this result, we get:

$$\frac{\partial r_{ij}(\varphi_{ij})}{\partial \tau_{ij}} = -\left[\sigma\left(\frac{1+\gamma}{\sigma+\gamma}\right) - 1\right] \frac{r_{ij}(\varphi_{ij})}{\varphi_{ij}^*} \frac{\partial \varphi_{ij}^*}{\partial \tau_{ij}} = -(\sigma-1)\left(\frac{1+\gamma}{\sigma+\gamma}\right) \frac{r_{ij}(\varphi_{ij})}{\tau_{ij}}. \tag{A.76}$$

It then follows that:

$$\int_{\varphi_{ij}^*}^{\infty} \frac{\partial r_{ij}(\varphi)}{\partial \tau_{ij}} \frac{\tau_{ij}}{X_{ij}/M_{ij}} \mu_{ij}(\varphi) d\varphi = -\int_{\varphi_{ij}^*}^{\infty} (\sigma - 1) \left(\frac{1+\gamma}{\sigma+\gamma}\right) \frac{r_{ij}(\varphi_{ij})}{\tau_{ij}} \frac{\tau_{ij}}{X_{ij}/M_{ij}} \mu_{ij}(\varphi) d\varphi$$
$$= -(\sigma - 1) \left(\frac{1+\gamma}{\sigma+\gamma}\right) \left(\frac{1}{X_{ij}}\right) M_{ij} \int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi_{ij}) \mu_{ij}(\varphi) d\varphi$$
$$= -(\sigma - 1) \left(\frac{1+\gamma}{\sigma+\gamma}\right) \frac{X_{ij}}{X_{ij}} = -(\sigma - 1) \left(\frac{1+\gamma}{\sigma+\gamma}\right). \tag{A.77}$$

Substituting results (A.71), (A.72), (A.73), (A.75) and (A.77) into equation (A.69), we get:

$$\varepsilon_\tau = -\left[\underbrace{-\theta\left(\frac{1+\gamma}{\gamma}\right)}_{\text{extensive}} + \underbrace{(1-\sigma)\left(\frac{1+\gamma}{\sigma+\gamma}\right)}_{\text{intensive}} + \underbrace{(\sigma-1)\left(\frac{1+\gamma}{\sigma+\gamma}\right)}_{\text{compositional}}\right]$$
$$= \theta\left(\frac{1+\gamma}{\gamma}\right) = \theta\left(1 + \frac{1}{\gamma}\right),$$

which is the result in the paper.

A.13 Elasticity of Trade with respect to Fixed Trade Costs

The computations for the fixed-trade-cost trade elasticity are similar to those for the *ad valorem* variable-trade-cost trade elasticity. From equation (A.67), we get:

$$\frac{\partial X_{ij}}{\partial f_{ij}} = \frac{\partial M_{ij}}{\partial f_{ij}} \frac{X_{ij}}{M_{ij}} + M_{ij}[1 - G(\varphi_{ij}^*)]^{-2} \frac{\partial G(\varphi_{ij}^*)}{\partial \varphi} \frac{\partial \varphi_{ij}^*}{\partial f_{ij}} [1 - G(\varphi_{ij}^*)] \frac{X_{ij}}{M_{ij}}$$
$$- M_{ij}[1 - G(\varphi_{ij}^*)]^{-1} r_{ij}(\varphi_{ij}^*) g(\varphi_{ij}^*) \frac{\partial \varphi_{ij}^*}{\partial f_{ij}} \tag{A.78}$$
$$+ M_{ij}[1 - G(\varphi_{ij}^*)]^{-1} \int_{\varphi_{ij}^*}^{\infty} \frac{\partial r_{ij}(\varphi)}{\partial f_{ij}} g(\varphi) d\varphi,$$

such that

$$\varepsilon_f \equiv -\frac{\partial X_{ij}}{\partial f_{ij}} \frac{f_{ij}}{X_{ij}} = -\left\{ \underbrace{\frac{\partial M_{ij}}{\partial f_{ij}} \frac{f_{ij}}{M_{ij}}}_{\text{extensive}} + \underbrace{\frac{g(\varphi_{ij}^*)\varphi_{ij}^*}{1 - G(\varphi_{ij}^*)} \left[1 - \frac{r_{ij}(\varphi_{ij}^*)}{X_{ij}/M_{ij}}\right] \frac{\partial \varphi_{ij}^*}{\partial f_{ij}} \frac{f_{ij}}{\varphi_{ij}^*}}_{\text{compositional}} \right.$$
$$\left. + \underbrace{\frac{f_{ij}}{X_{ij}/M_{ij}} \int_{\varphi_{ij}^*}^{\infty} \frac{\partial r_{ij}(\varphi)}{\partial f_{ij}} \mu_{ij}(\varphi) d\varphi}_{\text{intensive}} \right\}. \tag{A.79}$$

Some of the "components" of this last result are the same as those in equation (A.69). So, we calculate only the new components of equation (A.79). First, from equation (9), we have:

$$\frac{\partial \varphi_{ij}^*}{\partial f_{ij}} \frac{f_{ij}}{\varphi_{ij}^*} = \left(\frac{\sigma + \gamma}{\gamma}\right) \left(\frac{1}{\sigma - 1}\right). \tag{A.80}$$

Using this result and equations (A.73) and (A.75), it follows that the compositional margin defined in (A.79) simplifies to 1:

$$\begin{aligned} &\frac{g(\varphi_{ij}^*)\varphi_{ij}^*}{1 - G(\varphi_{ij}^*)} \left[1 - \frac{r_{ij}(\varphi_{ij}^*)}{X_{ij}/M_{ij}}\right] \frac{\partial \varphi_{ij}^*}{\partial f_{ij}} \frac{f_{ij}}{\varphi_{ij}^*} \\ &\qquad = \theta \left[1 - 1 + \frac{1}{\theta}\left(\frac{\gamma}{\sigma + \gamma}\right)(\sigma - 1)\right] \left(\frac{\sigma + \gamma}{\gamma}\right) \left(\frac{1}{\sigma - 1}\right) = 1. \end{aligned} \tag{A.81}$$

Next, using the definition of firm-level revenue in equation (A.76), we can show that:

$$\frac{\partial r_{ij}(\varphi_{ij})}{\partial f_{ij}} = 0. \tag{A.82}$$

This result implies that the intensive-margin component of the elasticity in (A.79) is equal to 0. Finally, from the equilibrium mass of firms in equation (11), we have:

$$\begin{aligned} \frac{\partial M_{ij}}{\partial f_{ij}} &= -\theta \left(\frac{M_{ij}}{\varphi_{ij}^*}\right) \frac{\partial \varphi_{ij}^*}{\partial f_{ij}} = -\theta \left(\frac{\sigma + \gamma}{\gamma}\right) \left(\frac{1}{\sigma - 1}\right) \left(\frac{M_{ij}}{\varphi_{ij}^*}\right) \frac{\varphi_{ij}^*}{f_{ij}} \\ &= -\theta \left(\frac{\sigma + \gamma}{\gamma}\right) \left(\frac{1}{\sigma - 1}\right) \frac{M_{ij}}{f_{ij}}. \end{aligned}$$

This last result implies that:

$$\frac{\partial M_{ij}}{\partial f_{ij}} \frac{f_{ij}}{M_{ij}} = -\theta \left(\frac{\sigma + \gamma}{\gamma}\right) \left(\frac{1}{\sigma - 1}\right). \tag{A.83}$$

Substituting equations (A.81), (A.82), and (A.83) into equation (A.69), we get:

$$\varepsilon_f = -\left[\underbrace{-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma - 1)}}_{\text{extensive}} + \underbrace{0}_{\text{intensive}} + \underbrace{1}_{\text{compositional}}\right] = \frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma - 1)} - 1 = \frac{\theta \frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma - 1)} - 1,$$

which is the result in the paper.

A.14 Welfare

In the model, welfare (W_j) is equal to purchasing power. Letting the consumption aggregate $C_j \equiv U_j$, then by definition of the price index it follows that:

$$P_j C_j = w_j \quad \Leftrightarrow \quad W_j = \frac{w_j}{P_j}, \tag{A.84}$$

where P is the ideal price index. To compute welfare, we need to define each term of W_j. We begin with the price index.

From the zero-profit condition $\pi_{ij}(\varphi_{ij}^*) = 0$ and the definition of profits in equation (A.6), we have:

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{r_{ij}(\varphi_{ij}^*)}{\sigma} = \tilde{w}_i f_{ij}.$$

Substituting demand function (A.2) into the equation above for $r_{ij}(\varphi_{ij}^*)$ yields:

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right)\frac{E_j P_j^{\sigma-1} b_i^{1-\sigma} p_{ij}(\varphi_{ij}^*)^{1-\sigma}}{\sigma} = \tilde{w}_i f_{ij} \quad \Rightarrow \quad b_i^{1-\sigma} p_{ij}(\varphi_{ij}^*)^{1-\sigma} = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\frac{\sigma \tilde{w}_i f_{ij}}{E_j P_j^{\sigma-1}}. \tag{A.85}$$

Substituting this result into equation (A.44), we obtain:

$$\begin{aligned} P_j^{1-\sigma} &= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\left(\frac{1+\gamma}{\sigma+\gamma}\right)\frac{\sigma}{E_j P_j^{\sigma-1}}\sum_i M_{ij}\tilde{w}_i f_{ij} \\ \Leftrightarrow \quad 1 &= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\left(\frac{1+\gamma}{\sigma+\gamma}\right)\frac{\sigma}{E_j}\sum_i M_{ij}\tilde{w}_i f_{ij}. \end{aligned} \tag{A.86}$$

Substituting in the equation above with the mass of firms from equation (A.41) yields:

$$1 = \left[\frac{\theta}{\theta-(\sigma-1)\frac{\gamma}{\sigma+\gamma}}\right]\left(\frac{1+\gamma}{\sigma+\gamma}\right)\frac{\sigma}{E_j}\sum_i\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{A_i L_i \frac{w_i}{A_i} f_{ij}}{\theta\delta f^e(\varphi_{ij}^*)^{\theta}}$$

which simplifies to:

$$1 = \left[\frac{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right] E_j^{-1}\sum_i \frac{w_i L_i f_{ij}}{\delta f^e}(\varphi_{ij}^*)^{-\theta}.$$

Substituting in equation (A.12) for φ_{ij}^* in the equation above yields:

$$1=\left[\frac{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right]\frac{E_j^{-1}}{\delta f^e}\sum_i w_i L_i f_{ij}\left\{\left[\frac{\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}\tilde{w}_i\right)^{\sigma}}{E_j P_j^{\sigma-1} b_i^{1-\sigma}}\right]^{\frac{\frac{1+\gamma}{\gamma}}{\sigma-1}}\left[\frac{\gamma}{\sigma+\gamma}(\sigma-1)f_{ij}\right]^{\frac{\frac{1+\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}\tau_{ij}^{\frac{1+\gamma}{\gamma}}\right\}^{-\theta}.$$

Solving the equation above for $P_j^{-\theta\frac{1+\gamma}{\gamma}}$ on the LHS yields:

$$P_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}=DE_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)-1}\sum_{i=1}^{N}A_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}w_i L_i w_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}\tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}f_{ij}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}$$

where

$$D=\left[\frac{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta-\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right](\delta f^e)^{-1}\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)\right]^{\frac{-\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}$$

is a constant that depends on parameters σ, γ, θ, δ and f^e. It will be convenient to rewrite the equation above as:

$$P_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}=DE_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)-1}\sum_{k=1}^{N}A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}w_k L_k w_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}b_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}\tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}. \tag{A.87}$$

Having defined the first component of welfare (P_j), we turn to the second component: wage rates. From equation (A.48), we have:

$$\lambda_{jj}=\frac{w_j L_j w_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}A_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}b_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}f_{jj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N}w_k L_k w_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}b_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}\tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)}f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}$$

where $\tau_{jj}=1$, as standard in the literature. Dividing both sides by λ_{jj} and multiplying both sides by $w_j^{\theta\frac{1+\gamma}{\gamma}}$ yields:

$$w_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)} = \left(\frac{1}{\lambda_{jj}}\right) \frac{w_j L_j w_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} A_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{jj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} w_k L_k w_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} A_k^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_k^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}}, \tag{A.88}$$

Multiplying equations (A.87) and (A.88) yields:

$$W_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)} = DE_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)-1} \left(\frac{1}{\lambda_{jj}}\right) w_j L_j w_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} A_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{jj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}.$$

Since $E_j = w_j L_j$, then:

$$W_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)} = D\left(\frac{1}{\lambda_{jj}}\right) L_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} A_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{jj}^{1-\frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}$$

or

$$W_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)} = D^{\frac{1}{\theta\left(\frac{1+\gamma}{\gamma}\right)}} \lambda_{jj}^{-\frac{1}{\theta\left(\frac{1+\gamma}{\gamma}\right)}} L_j^{\frac{1}{\sigma-1}} A_j^{\frac{\sigma}{\sigma-1}} b_j^{-1} f_{jj}^{\left(\frac{1}{1+\gamma}\right)\left(\frac{\gamma}{\theta}-\frac{\sigma+\gamma}{\sigma-1}\right)}.$$

Hence, for any *foreign* shock (i.e., holding constant L_j, A_j, b_j and f_{jj}), then:

$$\hat{W}_j = \hat{\lambda}_{jj}^{-\frac{1}{\theta\left(\frac{1+\gamma}{\gamma}\right)}} \tag{A.89}$$

where the hat denotes the gross change, i.e., W_j'/W_j and $\lambda_{jj}'/\lambda_{jj}$, where W_j' and λ_{jj}' denote the post-shock values of W_j and λ_{jj}, respectively.

Feenstra (2010) insightfully shows that one can interpret the gains from trade in a Melitz model as a gain due to increase in "export variety" or "average productivity." Importantly, the gain reflects the increase in real wage rates due to the productivity improvement as new exporting firms drive out less productive domestic firms, *raising average productivity*.[45]

To make this point, Feenstra (2010) derives a transformation curve between masses of varieties for sale to different markets, M_{ij}, and shows that trade increases real income by allowing the economy to reach more productive output combinations. As shown below in section 15 of Appendix A, we can solve for the concave transformation frontier between the

[45] As Feenstra (2010) notes, because the gains from new imported varieties exactly offset the losses from fewer domestic varieties (under the Pareto distribution assumption), there are no further gains from trade on the consumption side.

(output-adjusted) masses of varieties, $\tilde{M}_{ij}$, as follows:

$$L_i = k_1 (f^e)^{\frac{1}{1+\eta}} \left(\sum_j f_{ij}^{\frac{\eta-\theta}{\eta}} \tilde{M}_{ij}^{\frac{1+\eta}{\eta}} \right)^{\frac{\eta}{1+\eta}}, \tag{A.90}$$

where $k_1 > 0$ is a constant that depends only on parameters of the model (with the exact definition of k_1 provided later in section A.15). The economically important difference between our result under IMC and that in Feenstra (2010) under CMC is that the constant-elasticity-of-transformation (CET) in our model is $\eta = \theta \left(\frac{\sigma}{\sigma-1} \right) \left(\frac{1+\gamma}{\gamma} \right) - 1 > 0$, whereas Feenstra's CET is $\omega = \theta \left(\frac{\sigma}{\sigma-1} \right) - 1 > 0$. All else equal, $\eta \geq \omega$ because $(1+\gamma)/\gamma \geq 1$, with strict inequality when $\gamma < \infty$. Thus, with IMC, the CET curve will be flatter than under CMC as long as $\gamma < \infty$. In fact, we can show:

$$\eta = \omega + (\omega + 1)/\gamma,$$

which reveals the degree to which the CET under IMC is larger. As γ declines from ∞, η increases relative to ω. As γ approaches ∞, $\eta = \omega$, as in Feenstra (2010).

In section 15 of Appendix A below, we show that aggregate income in our model is a linear function of the (output-adjusted) masses of varieties:

$$R_i = \sum_{j=1}^{N} A_{ij} \tilde{M}_{ij}, \tag{A.91}$$

where, to be consistent (and tractable) with Feenstra (2010), the A_{ij}s now denote demand-shift parameters that depend only on parameters of the model; in the remainder of this section and in the next, we omit any TFP shocks (labeled previously A) and preference shocks (labeled previously b). As explained in Feenstra (2010), the welfare maximizing combination of (output-adjusted) masses of varieties can be obtained by maximizing income in equation (A.91) subject to the transformation curve in equation (A.90).

We can now evaluate the impact of trade liberalization on welfare. For simplicity, consider the two-country case illustrated in Figure A.1 (an extended version of Figure 5 in Feenstra (2010)). As shown in Figure A.1, our transformation curve (the dashed bowed-out line from point A to point B) is flatter compared to that of Feenstra (2010) under CMC (the solid bowed-out line from point A to point B). Point A represents the equilibrium under autarky for both cases. At that point, the mass of (output-adjusted) varieties for sale in the domestic market is positive, $\tilde{M} > 0$, and the mass of (output-adjusted) varieties for sale in the foreign market is null, $\tilde{M}_x = 0$. Autarky income is represented by the straight line closest to the origin, starting at point A. By opening up to trade, the economy can increase

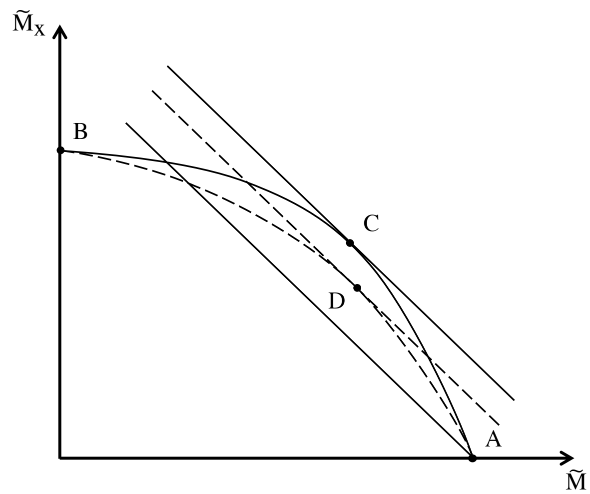

Figure A.1: CET Frontier with Increasing Marginal Costs and Constant Marginal Costs

its mass of (output-adjusted) varieties for sale in the foreign market and reduce its mass of (output-adjusted) varieties for sale in the domestic market. Under CMC, the gain in income is shown by the shift outward of the straight line through point A to the straight line tangent to the (solid-line) transformation curve at point C. Under IMC, the transformation curve is flatter which leads to smaller gains in income, as shown by the shift outward of the straight line through point A to the straight line tangent to the (dashed-line) transformation curve at point D. The difference between the income line tangent to point C and the income line tangent to point D represents the *welfare diminution effect* associated with IMC.

The diminished welfare gains due to IMC can also be interpreted mathematically in the context of Feenstra (2010). In a Melitz model with constant marginal costs, the change in welfare ($\hat{W}_j$) from a reduction in variable trade costs is proportionate to the change in average productivity ($\hat{\tilde{\varphi}}_{ij}$) and the change in the number of varieties ($\hat{M}_{ij}$), cf., Melitz (2003), equation (17). Feenstra (2010) shows also that the change in welfare ($\hat{W}_j$) can be simplified further to be proportionate to the change in output of the zero-cutoff-profit firm ($q_{ij}(\hat{\varphi}^*_{ij})$), cf. Feenstra (2010). As seen in equation (8) in the paper, under IMC the output of the cutoff productivity firm is proportional to the cutoff productivity according to:

$$q_{ij}(\varphi^*_{ij}) = \left[\left(\frac{\gamma}{\sigma + \gamma} \right) (\sigma - 1) f_{ij} \varphi^*_{ij} \right]^{\frac{\gamma}{1+\gamma}}.$$

Because a property of the Pareto distribution is that the average productivity, $\tilde{\varphi}_{ij}$, is proportionate to cutoff productivity, φ^*_{ij}, changes in welfare will be proportional to $(\hat{\varphi}^*_{ij})^{\frac{\gamma}{1+\gamma}}$. Under CMC, there is a linear relationship between the productivity cutoff and the output, i.e., as γ approaches ∞, $\frac{\gamma}{1+\gamma}$ approaches 1. However, when we introduce IMC, this relationship becomes concave. As a result, a given change in φ_{ij}^* has a *smaller effect* on output, $q_{ij}(\varphi_{ij}^*)$, under IMC than under CMC. This is the intuition underlying the "welfare diminution effect" from increasing marginal costs.

A.15 Constant Elasticity of Transformation

In this section, we derive the constant-elasticity-of-transformation (CET) function for our model. As a first step, we define aggregate revenue in our model. Using equations (A.45), (A.46), and (A.47):

$$R_i = \sum_{j=1}^{N} X_{ij} = \sum_{j=1}^{N} M_{ij} \int_{\varphi_{ij}^*}^{\infty} r_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi. \tag{A.92}$$

In our model, we can solve for $p_{ij}(\varphi) = q_{ij}(\varphi)^{-\frac{1}{\sigma}} \tau_{ij}^{\frac{1-\sigma}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} (w_j L_j)^{\frac{1}{\sigma}}$. Since $r_{ij}(\varphi) = p_{ij}(\varphi)q_{ij}(\varphi)$ and assuming aggregate revenue (R_i) equals aggregate income ($w_i L_i$), we can write:

$$R_i = w_i L_i = \sum_{j=1}^{N} A_{ij} M_{ij} \int_{\varphi_{ij}^*}^{\infty} q_{ij}(\varphi)^{\frac{\sigma-1}{\sigma}} \mu_{ij}(\varphi)d\varphi = \sum_{j=1}^{N} A_{ij}\tilde{M}_{ij} \tag{A.93}$$

where, analogous to Feenstra (2010):

$$A_{ij} = \tau_{ij}^{\frac{1-\sigma}{\sigma}} P_j \left(\frac{w_j L_j}{P_j}\right)^{\frac{1}{\sigma}} \tag{A.94}$$

and we denote $\tilde{M}_{ij}$ as the "output-adjusted" mass of varieties produced in country i and sold in market j:

$$\tilde{M}_{ij} = M_{ij} \int_{\varphi_{ij}^*}^{\infty} q_{ij}(\varphi)^{\frac{\sigma-1}{\sigma}} \mu_{ij}(\varphi)d\varphi. \tag{A.95}$$

In the context of our model, we know from in section 4 of Appendix A that:

$$\tilde{\varphi}_{ij} = \left[\int_{\varphi_{ij}^*}^{\infty} \varphi^{\frac{\gamma}{\gamma+\sigma}(\sigma-1)} \mu_{ij}(\varphi)d\varphi\right]^{\frac{1}{\frac{\gamma}{\gamma+\sigma}(\sigma-1)}} \tag{A.96}$$

is a measure of average productivity ($\tilde{\varphi}_{ij}$). Using equation (A.13) from section 4 of Appendix A, we can write:

$$q_{ij}(\varphi) = \left(\frac{\varphi}{\tilde{\varphi}_{ij}}\right)^{\sigma\frac{\gamma}{\gamma+\sigma}} q_{ij}(\tilde{\varphi}_{ij}). \tag{A.97}$$

Using equation (A.97) in the middle equality in equation (A.93) yields:

$$
\begin{aligned}
w_i L_i &= \sum_{j=1}^{N} A_{ij} M_{ij} \int_{\varphi_{ij}^*}^{\infty} \left[\left(\frac{\varphi}{\tilde{\varphi}_{ij}} \right)^{\sigma \frac{\gamma}{\gamma+\sigma}} q_{ij}(\tilde{\varphi}_{ij}) \right]^{\frac{\sigma-1}{\sigma}} \mu_{ij}(\varphi) d\varphi \\
&= \sum_{j=1}^{N} A_{ij} M_{ij} [q_{ij}(\tilde{\varphi}_{ij})]^{\frac{\sigma-1}{\sigma}} \tilde{\varphi}_{ij}^{(1-\sigma)\frac{\gamma}{\gamma+\sigma}} \int_{\varphi_{ij}^*}^{\infty} \varphi^{(\sigma-1)\frac{\gamma}{\gamma+\sigma}} \mu_{ij}(\varphi) d\varphi. \qquad \text{(A.98)}
\end{aligned}
$$

Since the integral term in the equation above simplifies to $\tilde{\varphi}_{ij}^{(\sigma-1)\frac{\gamma}{\gamma+\sigma}}$, then:

$$
w_i L_i = \sum_{j=1}^{N} A_{ij} M_{ij} [q_{ij}(\widetilde{\varphi}_{ij})]^{\frac{\sigma-1}{\sigma}} = \sum_{j=1}^{N} A_{ij} \tilde{M}_{ij} \qquad \text{(A.99)}
$$

where

$$
\tilde{M}_{ij} = M_{ij} [q_{ij}(\tilde{\varphi}_{ij})]^{\frac{\sigma-1}{\sigma}}.
$$

Using the equations for output and average productivity (A.9) and (A.23), respectively, and inverting equation (A.41) to solve for φ_{ij}^* as a function of M_{ij}, we find:

$$
\tilde{M}_{ij} = k_0 f_{ij}^{\frac{\gamma}{\gamma+1} \frac{\sigma-1}{\sigma}} \left(\frac{f^e}{L_i} \right)^{-\frac{\gamma}{1+\gamma} \frac{\sigma-1}{\theta\sigma}} M_{ij}^{1-\frac{\gamma}{\gamma+1} \frac{\sigma-1}{\theta\sigma}}, \qquad \text{(A.100)}
$$

where k_0 is a constant that depends only on parameters σ, γ, θ, and δ:

$$
k_0 = \left[\frac{\theta}{\theta - (\sigma-1)\frac{\gamma}{\gamma+\sigma}} \right] \left[\left(\frac{\gamma}{\gamma+\sigma} \right) (\sigma-1) \right]^{\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)} \left[\left(\frac{\gamma}{1+\gamma} \right) \left(\frac{\sigma-1}{\sigma} \right) \frac{1}{\theta\delta} \right]^{\frac{1}{\theta}\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)}.
$$

We invert equation (A.100) to solve for the mass of firms as a function of the adjusted mass:

$$
M_{ij} = \left(\frac{1}{k_0} \right)^{\frac{1+\eta}{\eta}} f_{ij}^{-\frac{\theta}{\eta}} \left(\frac{f_e}{L_i} \right)^{\frac{1}{\eta}} \tilde{M}_{ij}^{\frac{1+\eta}{\eta}} \qquad \text{(A.101)}
$$

where $\eta = \theta \left(\frac{1+\gamma}{\gamma}\right) \left(\frac{\sigma}{\sigma-1}\right) - 1$.

We can use equation (A.36), from section 5 of Appendix A, to express country i's labor stock as a linear transformation function of masses M_{ij}:

$$
L_i = \left(\frac{1+\gamma}{\gamma} \right) \left(\frac{\sigma}{\sigma-1} \right) \left[\frac{\theta(\sigma-1)\frac{\gamma}{\gamma+\sigma}}{\theta - (\sigma-1)\frac{\gamma}{\gamma+\sigma}} \right] \sum_{j=1}^{N} M_{ij} f_{ij}. \qquad \text{(A.102)}
$$

Substituting equation (A.101) into equation (A.102) yields country i's labor stock as a *concave* CET function of the “output-adjusted” masses:

$$L_i = k_1 (f^e)^{\frac{1}{1+\eta}} \left(\sum_j f_{ij}^{\frac{\eta-\theta}{\eta}} \tilde{M}_{ij}^{\frac{1+\eta}{\eta}} \right)^{\frac{\eta}{1+\eta}} \tag{A.103}$$

which is similar – but not identical – to (corrected) equation (3.24) in Feenstra (2010).[46] Note that k_1 is a constant that depends only parameters $\sigma, \gamma,$, θ, and k_0:

$$\begin{aligned} k_1 &= \frac{1}{k_0} \left[\left(\frac{1+\gamma}{\gamma} \right) \left(\frac{\sigma}{\sigma - 1} \right) \frac{\theta(\sigma-1)\frac{\gamma}{\gamma+\sigma}}{\theta - (\sigma-1)\frac{\gamma}{\gamma+\sigma}} \right]^{1-\frac{1}{\theta}\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)} \\ &= \frac{1}{k_0} \left[\frac{\theta\sigma\left(\frac{1+\gamma}{\gamma+\sigma}\right)}{\theta - (\sigma-1)\frac{\gamma}{\gamma+\sigma}} \right]^{1-\frac{1}{\theta}\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)} . \end{aligned}$$

[46] The exponent for f_{ij}, $1 - \frac{\theta}{\eta}$, differs from, and is a corrected version of, that in Feenstra (2010). Under CMC, the exponent in Feenstra (2010) should be $1 - \frac{\theta}{\omega}$, not $1 + \frac{\theta}{\omega}$ $\left(= 1 + \frac{(\omega+1)(\sigma-1)}{\omega\sigma}\right)$, and was confirmed with Robert Feenstra in email correspondence.

B Appendix B

B.1 The Bergstrand (1985) Model with Increasing Marginal Costs

As noted in numerous studies and in prominent surveys of the gravity equation in international trade, the first formal theoretical foundation for the gravity equation was Anderson (1979). Assuming a frictionless world, Anderson (1979) established theoretically one of the most enduring empirical relationships in international trade – that bilateral trade from i to j (X_{ij}) was proportional to the *product* of both countries' national outputs ($Y_i Y_j$) – using only four assumptions: every country i is endowed with a nationally differentiated output (Y_i), preferences are identical and homothetic across countries, the assumed absence of trade costs allows all prices to be identical across countries, and trade is balanced multilaterally (i.e., markets clear). The first three assumptions implied the demand for i's output in j was proportionate to j's output, $X_{ij} = b_i Y_j$, where b_i is every importer's demand for the good of i as a share of its expenditures. Assuming all output of each country is absorbed (i.e., markets clear), $X_{ij} = Y_i Y_j / Y_W$, where Y_W is world output. However, once Anderson (1979) introduced (positive) trade costs, he was unable to generate a transparent "structural" gravity equation, such as in Anderson and van Wincoop (2003). In fact, throughout the later sections including his appendix (using CES preferences), Anderson (1979) assumed inappropriately "the convention that all free trade prices are unity" despite his incorporating trade costs (cf., p. 115).

In contrast to Anderson (1979), the main motivation behind Bergstrand (1985) was to address the role of prices in the gravity equation, both theoretically and empirically. Unlike Anderson (1979), Bergstrand (1985) started with a CES utility function to emphasize that products from various markets were imperfect substitutes, as originally hypothesized by Armington. Moreover, he nested a CES utility function among importables inside a CES utility function between importables and the domestic good. On the supply side, he chose not to use the convention of constant marginal costs. Rather, he introduced a constant-elasticity-of-transformation (CET) function for producing output in the domestic market and foreign market, allowing a cost (in terms of labor) for output to be transformed between home and foreign markets. He also used a CET function to allow a cost for foreign output to be transformed between various export markets. He nested the latter CET function inside the former CET function. This formulation motivated upward-sloping supply curves *for each bilateral market* (including the domestic market). Assuming bilateral import demand values equaled bilateral export supply values in general equilibrium, this generated a system of $4N^2 + 3N$ equations in the same number of unknowns.

Assuming each bilateral market was small relative to the other $N^2 - 1$ markets and identical preferences and technologies across countries, Bergstrand (1985) derived the trade gravity equation:

$$X_{ij} = Y_i^{\frac{\sigma-1}{\gamma+\sigma}} Y_j^{\frac{\gamma+1}{\gamma+\sigma}} (C_{ij}T_{ij})^{-\sigma\frac{\gamma+1}{\gamma+\sigma}} E_{ij}^{\sigma\frac{\gamma+1}{\gamma+\sigma}} \left(\sum_{k=1,k\neq i}^{N} p_{ik}^{1+\gamma}\right)^{-\frac{(\sigma-1)(\gamma-\eta)}{(1+\gamma)(\gamma+\sigma)}} \left(\sum_{k=1,k\neq j}^{N} \overline{p}_{kj}^{1-\sigma}\right)^{\frac{(\gamma+1)(\sigma-\mu)}{(1-\sigma)(\gamma+\sigma)}}$$
$$\left[\left(\sum_{k=1,k\neq i}^{N} p_{ik}^{1+\gamma}\right)^{\frac{1+\eta}{1+\gamma}} + p_{ii}^{1+\eta}\right]^{-\frac{\sigma-1}{\gamma+\sigma}} \left[\left(\sum_{k=1,k\neq j}^{N} \overline{p}_{kj}^{1-\sigma}\right)^{\frac{1-\mu}{1-\sigma}} + p_{jj}^{1-\mu}\right]^{-\frac{\gamma+1}{\gamma+\sigma}}, \tag{B.1}$$

where $C_{ij} \geq 1$ is the gross transport (or c.i.f./f.o.b.) factor, $T_{ij} \geq 1$ is the gross tariff rate, E_{ij} is the spot exchange rate (value of j's currency in terms of i's), p_{ik} is the (free-on-board, or f.o.b.) price in i's currency of i's goods sold in k, $\bar{p}_{kj}$ is the (cost-insurance-freight, or c.i.f.) price of k's good in j (including tariffs), σ (μ) is the elasticity of substitution in consumption between importables (between importables and the domestic good), and γ (η) is the elasticity of transformation of output between export markets (between foreign markets and the domestic market).[47] The limitation in Bergstrand (1985) was that – due to the complexity of equation (B.1) – the market-clearing condition of Anderson (1979) could not be imposed.

In the remainder of this appendix, we provide two theoretical results. First, we show that a special case of gravity equation (14) in Bergstrand (1985) – labeled equation (B.1) above – yields that the intensive-margin (and trade) elasticity with respect to τ_{ij} is *identical* to the intensive-margin elasticity in Section 3.1 of this paper (from our modified Melitz model). Second, we show that – allowing the non-nested (single) constant-elasticity-of-transformation in this case to equal infinity and assuming multilateral trade balance – a "structural gravity equation" results.

B.2 Reconciling the Intensive-Margin Elasticity in Bergstrand (1985) with Section 3.1's Intensive-Margin Elasticity

Before we reconcile equation (B.1) with structural gravity, a special case of Bergstrand (1985) yields an intensive-margin (and, in this homogeneous-firm context, trade) elasticity identical to that in Section 3.2. We need only two assumptions. First, assume the elasticities of substitution in consumption in equation (B.1) to be identical ($\sigma = \mu$). Second, assume the elasticities of transformation in equation (B.1) to be identical ($\gamma = \eta$). Simplifying notation

[47] We have replaced here some notation in the original article. We use X_{ij} for the nominal trade flow rather than PX_{ij} and we use p_{ij} rather than P_{ij} to denote bilateral prices.

in equation (B.1) by denoting $\tau_{ij} = C_{ij}T_{ij}/E_{ij}$, these two assumptions yield:

$$X_{ij} = Y_i^{\frac{\sigma-1}{\gamma+\sigma}} Y_j^{\frac{\gamma+1}{\gamma+\sigma}} (\tau_{ij})^{(1-\sigma)\frac{\gamma+1}{\gamma+\sigma}} \left[\left(\sum_{j=1}^{N} p_{ij}^{1+\gamma} \right)^{\frac{1}{1+\gamma}} \right]^{(1-\sigma)\frac{\gamma+1}{\gamma+\sigma}} \left[\left(\sum_{i=1}^{N} (p_{ij}\tau_{ij})^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \right]^{-(1-\sigma)\frac{\gamma+1}{\gamma+\sigma}} . \tag{B.2}$$

From equation (B.2), the (positively-defined) intensive-margin (and trade) elasticity with respect to τ_{ij} is:

$$\varepsilon_\tau = -\frac{\partial X_{ij}}{\partial \tau_{ij}} \frac{\tau_{ij}}{X_{ij}} = -\frac{1+\gamma}{\sigma+\gamma}(1-\sigma) = \frac{1+\gamma}{\sigma+\gamma}(\sigma-1). \tag{B.3}$$

This elasticity is identical to that in Section 3.1 of the current paper. Moreover, this trade elasticity is *scaled down* by $\frac{1+\gamma}{\sigma+\gamma}$ relative to the constant marginal cost case in Anderson (1979) (and analogously in Krugman (1980)). The intuitive explanation for this was provided in the paper's introduction, Section 1, and illustrated in Figure 1.

B.3 Reconciling the Gravity Equation in Bergstrand (1985) with Structural Gravity

The second theoretical result in this appendix is to show that a special case of gravity equation (14) in Bergstrand (1985) is consistent with the structural gravity equation in Anderson and van Wincoop (2003) and in Baier et al. (2017). Building upon the previous section B.2, add two more assumptions. First, assume production is now *costlessly* transformable between markets ($\gamma = \infty$). With this additional assumption, equation (B.2) above simplifies to:

$$X_{ij} = Y_j \left(\frac{p_i \tau_{ij}}{P_j} \right)^{1-\sigma} \tag{B.4}$$

where p_{ij} is replaced by p_i since output is now costlessly transformed between markets and:

$$P_j \equiv \left[\sum_{i=1}^{N} (p_i \tau_{ij})^{1-\sigma} \right]^{1/(1-\sigma)} . \tag{B.5}$$

Equation (B.4) is identical to equation (6) in Anderson and van Wincoop (2003) (ignoring the arbitrary preference parameter β_i in that paper) and to the bilateral import demand functions in structural gravity equations discussed in Baier et al. (2017). Second, structural gravity follows once one assumes also market clearance (trade balance), $Y_i = \sum_{j=1}^{N} X_{ij}$.

Following derivations in Anderson and van Wincoop (2003) and Baier et al. (2017):

$$X_{ij} = \frac{Y_i Y_j}{Y_W} \left(\frac{\tau_{ij}}{\Pi_i P_j} \right)^{1-\sigma} \tag{B.6}$$

where:

$$\Pi_i = \left[\sum_{j=1}^{N} \frac{Y_j}{Y_W} \left(\frac{\tau_{ij}}{P_j} \right)^{1-\sigma} \right]^{1/(1-\sigma)} \tag{B.7}$$

and:

$$P_j = \left[\sum_{i=1}^{N} \frac{Y_j}{Y_W} \left(\frac{\tau_{ij}}{\Pi_i} \right)^{1-\sigma} \right]^{1/(1-\sigma)}. \tag{B.8}$$

Thus, the simplifications of equation (B.1) above from Bergstrand (1985) – along with adding in the market-clearing condition – yield the same structural gravity equation as in Anderson and van Wincoop (2003) and Baier et al. (2017).

C Appendix C

The key distinguishing assumption of our model is that marginal costs are increasing in output. There are many ways to implement this. In section 2.2 of the paper, we motivated the case for marginal costs increasing with respect to destination-specific output. This is one extreme of a range of models. At the other extreme, marginal costs could depend exclusively on the overall output of the firm. In that case, all the destination-specific customization is captured in the fixed export costs, as more common to Melitz models. In this appendix, we develop a model that fits this type of increasing marginal costs, that is, marginal costs are allowed to increase with *total firm output*.

If the marginal costs depend on overall firm output, which itself depends on the endogenous set of countries to which the firm exports, we cannot solve analytically a model with asymmetric country size and asymmetric bilateral trade barriers. As a consequence, in this appendix we assume all countries are identical and develop an extension of the symmetric-country Melitz (2003) model with increasing marginal costs in total firm-level output and a Pareto distribution of firm productivity. We present only key results because the solution method is similar to the one we used to solve the model in the main text; we refer the reader to Appendix A for additional details.

Consider a world with $1 + J$ identical countries. The representative consumer in each country has CES preferences defined over differentiated varieties. The representative consumer maximizes utility subject to the standard income constraint. Hence, the optimal aggregate demand function for each variety ν is given by:

$$c(\nu) = EP^{\sigma-1}p^c(\nu)^{-\sigma}, \quad \text{with} \quad P = \left[\int_{\nu\in\Omega} p^c(\nu)^{1-\sigma}d\nu\right]^{\frac{1}{1-\sigma}} \tag{C.1}$$

where E denotes aggregate expenditure, $p^c(\nu)$ is the unit price of variety ν, and Ω is the set of varieties available for consumption.[48]

Firms face fixed production costs and increasing marginal costs, such that the total labor demand by a firm depends on its total output (q) and whether or not the firm exports as follows:

$$l(\varphi) = f + I_x J f_x + \frac{q^{1+\frac{1}{\gamma}}}{\varphi}, \tag{C.2}$$

where φ denotes the firm's productivity and q is total output defined as:

$$q = q_d + I_x J q_x,$$

[48] Different from the main text and Appendix A, we omit, for brevity, preference parameter b.

where q_d denotes domestic sales and q_x denotes sales to a foreign market.[49] The variable I_x is an indicator function equal to 1 if the firm exports and 0 otherwise. It is important to note that, because countries and (international bilateral) trade costs (τ and f_x) are symmetric, if a firm can export profitably to one market abroad, it will be able to export profitably to all foreign markets.

Using the labor-demand function in equation (C.2), we can express firm-level profits as:

$$\pi(\varphi) = p_d q_d + I_x J p_x q_x - w \left[f + I_x J f_x + \frac{1}{\varphi} \left(q_d + I_x J q_x \right)^{1+\frac{1}{\gamma}} \right], \tag{C.3}$$

where w is the wage rate. It is important to emphasize that, in contrast to the benchmark model, it is not possible to separate total profits into domestic and export components. This key distinction is a direct consequence of the technology. As shown in equation (C.2), labor demand is a non-linear function of total firm output such that it is not possible to separate the costs associated with output for domestic sales from the costs associated with output for foreign sales. As a result, the firm's production costs must be expressed as a function of total output as seen from the last term in square brackets. This implies that we cannot solve for the optimal behavior of a given firm in each market separately, that is, without also taking into account its behavior in other markets. Instead, we need to characterize the optimal behavior of firm as a function of both its market (domestic vs. foreign) and its type (domestic vs. exporter).

Markets are segmented, such that firms can charge different prices in the domestic and foreign markets. Therefore, the firm-level profit maximization problem takes the following form:

$$\max_{p_d, p_x} \pi(\varphi) = p_d q_d + I_x J p_x q_x - w \left[f + I_x J f_x + \frac{1}{\varphi} \left(q_d + I_x J q_x \right)^{1+\frac{1}{\gamma}} \right] \tag{C.4}$$

subject to the demand constraints defined in equation (C.1). The two first-order conditions imply the following pricing rules:

$$\begin{aligned} p_d^D(\varphi) &= \left(\frac{1+\gamma}{\gamma} \right) \left(\frac{\sigma}{\sigma - 1} \right) \frac{w}{\varphi} q_d^D(\varphi)^{\frac{1}{\gamma}}, \\ p_d^X(\varphi) &= \left(\frac{1+\gamma}{\gamma} \right) \left(\frac{\sigma}{\sigma - 1} \right) \frac{w}{\varphi} \left[\left(1 + J \tau^{1-\sigma} \right) q_d^X(\varphi) \right]^{\frac{1}{\gamma}}, \end{aligned} \tag{C.5}$$

where $p_d^D(\varphi)$ and $q_d^D(\varphi)$ denote, respectively, the optimal domestic sales price and output of a (pure) domestic firm (denoted with superscript D) with productivity φ producing and selling in the domestic market (denoted with subscript d). Let $p_d^X(\varphi)$ and $q_d^X(\varphi)$ denote, respectively, the optimal sales price and output of an exporting firm (denoted with superscript X) with

[49] Different from the main text and Appendix A, we omit, for brevity, the TFP factor A.

productivity φ selling in the domestic market (denoted with subscript d). The results in equation (C.5) imply that, conditional on productivity and total output, exporting firms (located in country d) can charge higher, equal, or lower prices at home relative to (pure) domestic firms, due to opposing effects from productivity differences versus scale effects. The higher productivity of an exporter tends to lower p_d^X relative to p_d^D. However, an exporter serves more markets, tending to raise p_d^X relative to p_d^D.

We define the profitability threshold φ^* as the productivity level at which a (pure) domestic firm makes zero profits: $\pi(\varphi^*|I_x = 0) = 0$, where the profit function $\pi(\cdot)$ is defined in equation (C.3). Using this condition, we can solve for the output and the price of the threshold pure domestic firm as follows:

$$\begin{aligned} q_d^D(\varphi^*) &= \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)f\varphi^*\right]^{\frac{\gamma}{1+\gamma}}, \\ p_d^D(\varphi^*) &= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)f\right]^{\frac{1}{1+\gamma}} w(\varphi^*)^{\frac{-\gamma}{1+\gamma}}. \end{aligned} \tag{C.6}$$

Similarly, if we define the export profitability threshold as the level of productivity φ_x^* required for an exporting firm to break even, $\pi(\varphi_x^*|I_x = 1) = 0$, we can solve for the domestic price and output of the threshold exporting firm as follows:

$$\begin{aligned} q_d^X(\varphi_x^*) &= \left(\frac{1}{1+J\tau^{1-\sigma}}\right)\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)(f+Jf_x)\varphi_x^*\right]^{\frac{\gamma}{1+\gamma}}, \\ p_d^X(\varphi_x^*) &= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)(f+Jf_x)\right]^{\frac{1}{1+\gamma}} w(\varphi_x^*)^{\frac{-\gamma}{1+\gamma}}. \end{aligned} \tag{C.7}$$

Note that, when $J = 0$, these last two solutions become equivalent to the domestic firms' solutions in (C.6), as they should. Substituting the results in (C.6) and (C.7) into the zero-profit conditions that define the productivity thresholds and rearranging, we obtain:

$$\begin{aligned} \pi(\varphi^*|I_x = 0) = 0 \quad &\Leftrightarrow \quad r_d^D(\varphi^*) = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\sigma w f, \\ \pi(\varphi_x^*|I_x = 1) = 0 \quad &\Leftrightarrow \quad r_d^X(\varphi_x^*) = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\left(\frac{\sigma w}{1+J\tau^{1-\sigma}}\right)(f+Jf_x). \end{aligned} \tag{C.8}$$

From the definition of revenues ($r(\varphi) = p(\varphi)q(\varphi)$) and the optimal demand function in (C.1), it follows that

$$\frac{r_d^D(\varphi)}{r_d^D(\varphi^*)} = \left[\frac{p_d^D(\varphi)}{p_d^D(\varphi^*)}\right]^{1-\sigma} \quad \text{and} \quad \frac{r_d^X(\varphi)}{r_d^X(\varphi_x^*)} = \left[\frac{p_d^X(\varphi)}{p_d^X(\varphi_x^*)}\right]^{1-\sigma}. \tag{C.9}$$

We can simplify these results using the definitions of prices in equations (C.6) and (C.7) to obtain analytical expressions for the revenue of any firm as a function of the revenue of the threshold firm. Combining these expressions with equations (C.8), it is possible to express the revenue of any domestic and exporting firms, respectively, as follows:

$$
\begin{aligned}
r_d^D(\varphi) &= \left(\frac{1+\gamma}{\sigma+\gamma}\right)\sigma w f \left(\frac{\varphi}{\varphi^*}\right)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}, \\
r_d^X(\varphi) &= \left(\frac{1+\gamma}{\sigma+\gamma}\right)\left(\frac{\sigma w}{1+J\tau^{1-\sigma}}\right)(f+Jf_x)\left(\frac{\varphi}{\varphi_x^*}\right)^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}.
\end{aligned} \tag{C.10}
$$

Using equations (C.8), we can obtain a first expression for the ratio of domestic threshold revenue and export threshold revenue,

$$
\frac{r_d^D(\varphi^*)}{r_d^X(\varphi_x^*)} = \left(1+J\tau^{1-\sigma}\right)\left(\frac{f}{f+Jf_x}\right). \tag{C.11}
$$

We can obtain a second expression for the ratio of domestic threshold revenue and export threshold revenue using the definition of revenue and the optimal demand function as follows:

$$
\frac{r_d^D(\varphi^*)}{r_d^X(\varphi_x^*)} = \left[\frac{p_d^D(\varphi^*)}{p_d^X(\varphi_x^*)}\right]^{1-\sigma}. \tag{C.12}
$$

Using the definitions of prices in equations (C.6) and (C.7), we obtain:

$$
\frac{r_d^D(\varphi^*)}{r_d^X(\varphi_x^*)} = \left(\frac{f}{f+Jf_x}\right)^{\frac{1-\sigma}{1-\gamma}}\left(\frac{\varphi^*}{\varphi_x^*}\right)^{(\sigma-1)\left(\frac{\gamma}{1+\gamma}\right)}. \tag{C.13}
$$

Combining our two expressions for the ratio of revenues, (C.11) and (C.13), we can solve for the ratio of the productivity thresholds as follows:

$$
\frac{\varphi_x^*}{\varphi^*} = \left(\frac{1}{1+J\tau^{1-\sigma}}\right)^{\left(\frac{1}{\sigma-1}\right)\left(\frac{1+\gamma}{\gamma}\right)}\left(\frac{f+Jf_x}{f}\right)^{\left(\frac{1}{\sigma-1}\right)\left(\frac{\sigma+\gamma}{\gamma}\right)}. \tag{C.14}
$$

When $\gamma \to \infty$, the relationship between the two thresholds is analogous to that in the benchmark Melitz (2003) model. We can use the definitions of revenue in (C.10) and the ratio in (C.14) to express average profits as a function of parameters of the model and the profitability threshold φ^*. Using the free entry condition that the expected value of entry is equal to the cost of entry, we can show that there exists a unique equilibrium threshold φ^*.

We are interested in defining the trade elasticities in our model. For convenience, we introduce the term XD to denote aggregate domestic absorption. As a first step, we can define aggregate domestic absorption as follows:

$$XD = M \int_{\varphi^*}^{\infty} r_d(\varphi)\mu(\varphi)d\varphi = M \left[\int_{\varphi^*}^{\varphi_x^*} r_d^D(\varphi)\mu(\varphi)d\varphi + \int_{\varphi_x^*}^{\infty} r_d^X(\varphi)\mu(\varphi)d\varphi \right], \qquad \text{(C.15)}$$

where M is the equilibrium mass of firms in each country and $\mu(\varphi)$, defined as:

$$\mu(\varphi) = \begin{cases} 0 & \text{if } \varphi < \varphi^*, \\ \frac{g(\varphi)}{1-G(\varphi^*)} & \text{if } \varphi \geq \varphi^*, \end{cases} \qquad \text{(C.16)}$$

denotes the equilibrium distribution of firm productivities. We assume that the following theoretical restriction on the parameters holds: $\theta > \frac{\gamma}{\sigma+\gamma}(\sigma - 1)$. Equation (C.15) shows that domestic absorption depends on the mass of firms M and the average sales of firms in their domestic market. The average sales per firm can be decomposed into the separate contributions of domestic firms and exporting firms, the first and second terms in square brackets, respectively.

To obtain an analytical solution, we assume that firms draw their productivity from a Pareto distribution with parameter θ, such that $G(\varphi) = 1 - \varphi^{-\theta}$. Using this assumption and the definitions of revenue in equation (C.10), we can solve for aggregate domestic absorption as:

$$XD = M \left[\frac{\theta}{\theta - (\sigma - 1)\left(\frac{\gamma}{\sigma+\gamma}\right)} \right] \left(\frac{1+\gamma}{\sigma+\gamma} \right) \sigma\theta w f \qquad \text{(C.17)}$$
$$\times \left[1 - \left(\frac{\varphi_x^*}{\varphi^*} \right)^{\frac{\gamma(\sigma-1)}{\sigma+\gamma} - \theta} + \left(\frac{1}{1 + J\tau^{1-\sigma}} \right) \left(\frac{f + Jf_x}{f} \right) \left(\frac{\varphi_x^*}{\varphi^*} \right)^{-\theta} \right].$$

In a second step, we introduce, for convenience, the term XX to denote aggregate expenditures on foreign goods, noting that – due to symmetry – aggregate imports (from the rest of the world) equal aggregate exports (to the rest of the world). We define aggregate expenditure on foreign goods as:

$$XX = M_x J \int_{\varphi_x^*}^{\infty} r_x^X(\varphi)\mu_x(\varphi)d\varphi = MJ \int_{\varphi_x^*}^{\infty} r_x^X(\varphi)\mu(\varphi)d\varphi, \qquad \text{(C.18)}$$

where $M_x = [1 - G(\varphi_x^*)]M$ is the equilibrium mass of exporting firms in each country and $\mu_x(\varphi)$, defined as:

$$\mu_x(\varphi) = \begin{cases} 0 & \text{if } \varphi < \varphi_x^*, \\ \frac{g(\varphi)}{1-G(\varphi_x^*)} & \text{if } \varphi \geq \varphi_x^*, \end{cases} \qquad \text{(C.19)}$$

denotes the equilibrium distribution of exporting firms' productivities, where $\mu_x(\varphi) = \frac{1-G(\varphi^*)}{1-G(\varphi_x^*)}\mu(\varphi)$. Substituting with the definition of revenue in equation (C.10) and using the fact that $r_x^X(\varphi) = \tau^{1-\sigma} r_d^X(\varphi)$ yields:

$$XX = M\left[\frac{\theta}{\theta - (\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\left(\frac{1+\gamma}{\sigma+\gamma}\right)\sigma\theta w(f + Jf_x)\left(\frac{J\tau^{1-\sigma}}{1+J\tau^{1-\sigma}}\right)\left(\frac{\varphi_x^*}{\varphi^*}\right)^{-\theta}. \quad \text{(C.20)}$$

We now have separate analytical expressions for expenditures on domestic and foreign goods.

Using E to denote aggregate expenditures ($E = XD + XX$), we can now compute the share of aggregate expenditures on foreign goods (XX/E). Using equations (C.14), (C.17) and (C.20), we obtain:

$$\frac{XX}{E} = \frac{XX}{XD+XX} = \frac{\frac{J\tau^{1-\sigma}}{1+J\tau^{1-\sigma}}}{1 + \left(\frac{1}{1+J\tau^{1-\sigma}}\right)^{\left(\frac{\theta}{\sigma-1}\right)\left(\frac{1+\gamma}{\gamma}\right)}\left(\frac{f+Jf_x}{f}\right)^{\left(\frac{\theta}{\sigma-1}\right)\left(\frac{\sigma+\gamma}{\gamma}\right)-1} - \left(\frac{1}{1+J\tau^{1-\sigma}}\right)^{\frac{1+\gamma}{\sigma+\gamma}}}. \quad \text{(C.21)}$$

We can use this last result to derive the trade elasticities. Note that:

$$\varepsilon_\tau \equiv -\frac{\partial(XX/JE)}{\partial\tau}\frac{\tau}{XX/JE} = -\frac{\partial(XX/E)}{\partial\tau}\frac{\tau}{XX/E}, \quad \text{(C.22)}$$

$$\varepsilon_f \equiv -\frac{\partial(XX/JE)}{\partial f_x}\frac{f_x}{XX/JE} = -\frac{\partial(XX/E)}{\partial f_x}\frac{f_x}{XX/E}. \quad \text{(C.23)}$$

It is useful to introduce additional notation to simplify the presentation. Define the following terms:

$$a = \left(\frac{1}{1+J\tau^{1-\sigma}}\right)^{\left(\frac{\theta}{\sigma-1}\right)\left(\frac{1+\gamma}{\gamma}\right)}\left(\frac{f+Jf_x}{f}\right)^{\left(\frac{\theta}{\sigma-1}\right)\left(\frac{\sigma+\gamma}{\gamma}\right)-1}, \quad \text{(C.24)}$$

$$b = \left(\frac{1}{1+J\tau^{1-\sigma}}\right)^{\frac{1+\gamma}{\sigma+\gamma}}, \quad \text{(C.25)}$$

$$c = \frac{J\tau^{1-\sigma}}{1+J\tau^{1-\sigma}}. \quad \text{(C.26)}$$

Then, it is possible to rewrite the share of expenditures on foreign goods (C.21) as follows:

$$\frac{XX}{E} = \frac{c}{1+a-b}. \quad \text{(C.27)}$$

After some tedious, but straightforward, algebra, we can show that:

$$\varepsilon_{\tau} = \theta \left(\frac{1+\gamma}{\gamma} \right) \left[\frac{a}{1+a-b} - \left(\frac{\sigma - 1}{\theta} \right) \left(\frac{\gamma}{\sigma + \gamma} \right) \frac{b}{1+a-b} \right] c, \tag{C.28}$$

$$\varepsilon_{f} = \left[\left(\frac{\sigma+\gamma}{\gamma} \right) \left(\frac{\theta}{\sigma - 1} \right) - 1 \right] \left(\frac{a}{1+a-b} \right) c. \tag{C.29}$$

To gain some insight into these complex equations, we consider the case of a large number of countries. In the limit, when J tends to infinity it can be shown that:[50]

$$\lim_{J \to \infty} a = \infty \,(\text{if}\, \theta > \gamma), \lim_{J \to \infty} b = 0, \quad \text{and} \quad \lim_{J \to \infty} c = 1. \tag{C.30}$$

Together, it can be shown that these results imply:

$$\lim_{J \to \infty} \frac{a}{1+a-b} = 1, \quad \text{and} \quad \lim_{J \to \infty} \frac{b}{1+a-b} = 0. \tag{C.31}$$

Using these results in the definition of the elasticities in (C.28) and (C.29), it follows that:

$$\varepsilon_{\tau} = \theta \left(\frac{1+\gamma}{\gamma} \right), \tag{C.32}$$

$$\varepsilon_{f} = \frac{\theta \left(\frac{1+\gamma}{\gamma} \right)}{\frac{1+\gamma}{\sigma+\gamma} (\sigma - 1)} - 1. \tag{C.33}$$

These results show that – as the number of countries increases – the trade elasticities in our symmetric model with increasing marginal costs *defined over total firm output* converge to the elasticities in our benchmark model with asymmetric countries and destination-specific increasing marginal costs.

[50] We provide evidence in Section 5 of the paper, comparing Tables 3 and 4, that estimates of θ exceed estimates of γ within the 10th-75th percentiles of the 568 four-digit industries.

D Appendix D

In this appendix, we provide details on the derivations to establish the (structural) bilateral import-demand equation (29), the (structural) bilateral import-unit value equation (36), and then the (extended F/BW) reduced-form estimation equation (44) (which builds upon reduced-form equation (42)).

D.1 Bilateral Import Demand

Because the price index $\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi$ in equation (24) is not observable, we cannot use equation (24) to estimate the parameters of the model. To make progress, we express the observable average cost-insurance-freight (or cif) import unit value $\overline{p}_{ijt}^c$ as the ratio of two unobservable price indexes ($\tilde{p}_{ij}^c$, $\grave{p}_{ij}^c$), as noted in equation (25):

$$\overline{p}_{ij}^c \equiv \frac{X_{ij}^D}{Q_{ij}^D} = \frac{\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi}{\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{-\sigma}\mu_{ij}(\varphi)d\varphi} \equiv \frac{\tilde{p}_{ij}^c}{\grave{p}_{ij}^c}. \tag{D.1}$$

In what follows, we use the theoretical model to obtain analytical expressions for each of the unobserved price indexes, $\tilde{p}_{ij}^c$ and $\grave{p}_{ij}^c$. We then show that, by combining these two expressions in conjunction with the Pareto distribution (and allowing for deviations from Pareto, e.g., e_{ij}^{P1}, etc.), we can express nominal bilateral import demand as a function of the observable bilateral import unit value ($\overline{p}_{ij}^c$), e_{ij}^X, e_{ij}^{P1}, etc.

We proceed in several steps. The first step is to solve for firm-level (bilateral) prices $p_{ij}^c(\varphi)$ as functions of the productivity threshold φ_{ij}^*. Recalling $q_{ij}(\varphi)/\tau_{ij} = c_{ij}(\varphi)$ and $p_{ij}^c(\varphi) = \tau_{ij}p_{ij}(\varphi)$, we can use optimal demand equation (2) and optimal pricing rule (6) to show:

$$\frac{q_{ij}(\varphi)}{q_{ij}(\varphi_{ij}^*)} = \left(\frac{\varphi}{\varphi_{ij}^*}\right)^{\sigma\left(\frac{\gamma}{\sigma+\gamma}\right)}. \tag{D.2}$$

Substituting into equation (D.2) using equation (8) for $q_{ij}(\varphi_{ij}^*)$ yields:

$$q_{ij}(\varphi) = \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)f_{ij}\right]^{\frac{\gamma}{1+\gamma}} (\varphi_{ij}^*)^{-\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)} \varphi^{\sigma\left(\frac{\gamma}{\sigma+\gamma}\right)}. \tag{D.3}$$

Substituting equation (D.3) for $q_{ij}(\varphi)$ into optimal pricing rule (6) yields:

$$p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)f_{ij}\right]^{\frac{1}{1+\gamma}} (\varphi_{ij}^*)^{-\left(\frac{1}{1+\gamma}\right)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)} \tilde{w}_i \varphi^{-\frac{\gamma}{\sigma+\gamma}} \tag{D.4}$$

where recall that $\tilde{w}_i = w_i/A_i$.

In the second step, we compute the two unobservable average prices $\tilde{p}_{ij}^c$ and $\grave{p}_{ij}^c$ and show that the observable import unit value $\overline{p}_{ij}^c$ is proportional to the optimal price of the break-even exporter, $p_{ij}^c(\varphi_{ij}^*)$. Using equation (D.4), optimal pricing function (6), the Pareto distribution assumption allowing deviations from Pareto, and recalling $p_{ij}^c(\varphi) = \tau_{ij} p_{ij}(\varphi)$, we can solve for:

$$\tilde{p}_{ij}^c = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right] p_{ij}^c(\varphi_{ij}^*)^{1-\sigma} e_{ij}^{P1} \tag{D.5}$$

where $e_{ij}^P \neq 1$ implies deviations from the Pareto distribution for $\tilde{p}_{ij}^c$. Using equation (D.1), optimal pricing function (6), and the Pareto distribution assumption allowing deviations e_{ij}^{P2}, we can solve for:

$$\grave{p}_{ij}^c = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right] p_{ij}^c(\varphi_{ij}^*)^{-\sigma} e_{ij}^{P2}. \tag{D.6}$$

Using these results and equation (D.1), we obtain:

$$\overline{p}_{ij}^c = \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right] p_{ij}^c(\varphi_{ij}^*) \left(\frac{e_{ij}^{P1}}{e_{ij}^{P2}}\right) \tag{D.7}$$

which shows that observable $\overline{p}_{ij}^c$ is proportional to the optimal price of the zero-cutoff-profit exporter and Pareto deviations.

The third step is straightforward. We can rewrite equation (D.7) with $p_{ij}^c(\varphi_{ij}^*)$ as a function of the observable price import unit value $\overline{p}_{ij}^c$:

$$p_{ij}^c(\varphi_{ij}^*) = \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma+\gamma}{\theta(\sigma+\gamma)-\gamma\sigma}\right] \overline{p}_{ij}^c \left(\frac{e_{ij}^{P2}}{e_{ij}^{P1}}\right) \tag{D.8}$$

and substitute this last result into equation (26) to obtain:

$$\tilde{p}_{ij}^c = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right] \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right]^{\sigma-1} (\overline{p}_{ij}^c)^{1-\sigma} \left(\frac{e_{ij}^{P2}}{e_{ij}^{P1}}\right)^{1-\sigma} e_{ij}^{P1}. \tag{D.9}$$

We can now use this last result to express the aggregate nominal bilateral import demand, defined in equation (24), as a share of total expenditure as follows:

$$\frac{X_{ij}^D}{E_j} = k_2 M_{ij} P_j^{\sigma-1} (\overline{p}_{ij}^c)^{1-\sigma} \left(\frac{e_{ij}^{P2}}{e_{ij}^{P1}}\right)^{1-\sigma} e_{ij}^{P1} \tag{D.10}$$

where k_2 is a constant that depends only on the structural parameters σ, γ, and θ:

$$k_2 = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\,(\sigma-1)}\right] \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right]^{\sigma-1}.$$

In the fourth step, we remove the productivity threshold, φ^*_{ij}, using equation (8) and the mass of firms, M_{ij}, using an extended version of equation (11) allowing deviations from Pareto to yield:

$$\frac{X^D_{ij}}{E_j} = k_3 A_i^{1+\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} L_i w_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} b_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} E_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)}$$

$$P_j^{(\sigma-1)+\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{\frac{-\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}} (\overline{p}^c_{ij})^{1-\sigma} e^D_{ij} \tag{D.11}$$

where k_3 is a constant that depends only on the structural parameters σ, γ, θ, δ, and f^e:

$$k_3 = \frac{k_2}{\delta f^e} \left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \left[\frac{\gamma}{\sigma+\gamma}\left((\sigma-1)\right)\right]^{\frac{-\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}$$

and $e^D_{ij} \equiv e^{P1}_{ij} \left(\frac{e^{P2}_{ij}}{e^{P1}_{ij}}\right)^{1-\sigma} e^{P3}_{ij}$. While the first two RHS terms of e^D_{ij} were motivated above, the Pareto deviation e^{P3}_{ij} is associated with the mass of firms in the presence of deviations from Pareto. Referring back to section A.5 of Online Appendix A, the mass of firms M_{ij} turns out to be an extension of equation (A.41) with e^{P3}_{ij} appended to the RHS. Importantly, e^{P3}_{ij} has two components, one of which is e^{P4}_{ij} which surfaces because $1-G(\varphi^*_{ij}) = (\varphi^*_{ij})^{-\theta} e^{P4}_{ij}$ in the presence of deviations from Pareto. e^{P4}_{ij} is important for e^{P3}_{ij}, and hence e^D_{ij}, because of its particular influence on small exporters that tend to be near the cutoff productivity, consistent with the evidence that deviations from Pareto tend to surface for small exporters. The superscript D in e^D_{ij} refers to the role of deviations from Pareto on the "demand" side ($s_{ij} \equiv \frac{X^D_{ij}}{E_j}$).

This completes the derivation for the demand-side equation of the empirical model.

D.2 Bilateral Export Supply

The derivations for the bilateral import-unit value $\overline{p}^c_{ij}$ equation (36) are largely in the text and use part of section D.1 above. In those derivations, we use a constant k_4, defined as:

$$k_4 \equiv \left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\right]^{\gamma} \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right] \left[\frac{\theta(\gamma+\sigma)-\gamma\sigma+\gamma}{\theta(\gamma+\sigma)-\gamma\sigma}\right]^{\gamma} \tag{D.12}$$

and a constant k_5 defined as:

$$k_5 = \left(\frac{k_4}{\delta f^e}\right)^{-\frac{1}{1+\gamma}} \left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]^{\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \left[\frac{\gamma}{\sigma+\gamma}(\sigma-1)\right]^{\frac{\frac{\theta-\gamma}{\gamma}}{\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)}}. \tag{D.13}$$

D.3 Reduced-Form Specification

In this subsection, we provide derivations associated with a complete specification of theoretical coefficients based upon the model that are associated with estimating equations (42) or (44); in the interest of brevity, we provide this reduced-form equation where the underlying variables include only Group 1 and Group 2 variables associated with Specification 2, "IMC-Partial." Adapting equations (37) and (38) for this specification, ϵ_{ijt} simplifies to:

$$\epsilon_{ijt} = \Delta^k \ln s_{ijt} + (\sigma - 1)\Delta^k \ln \overline{p}^c_{ijt} + \theta \left(\frac{1+\gamma}{\gamma}\right) \Delta^k \ln tar_{ijt} + \theta \left(\frac{1+\gamma}{\gamma}\right) \Delta^k \ln trans_{ijt}, \tag{D.14}$$

and adapting equations (40) and (41) for this specification, ψ_{ijt} simplifies to:

$$\psi_{ijt} = -\frac{1}{1+\gamma}\Delta^k \ln s_{ijt} + \Delta^k \ln \overline{p}^c_{ijt} - \frac{\theta-\gamma}{\gamma}\Delta^k \ln tar_{ijt} - \frac{\theta-\gamma}{\gamma}\Delta^k \ln trans_{ijt}. \tag{D.15}$$

Taking the product of ϵ_{ijt} and ψ_{ijt} yields an equation with $\epsilon_{ijt}\psi_{ijt}$ on the LHS and 16 products on the RHS. Consolidating common RHS products yields:

$$\begin{aligned}
\epsilon_{ijt}\psi_{ijt} = {} & (\sigma - 1)\left(\Delta^k \ln \overline{p}^c_{ijt}\right)^2 - \frac{1}{1+\gamma}\left(\Delta^k \ln s_{ijt}\right)^2 + \left(1 - \frac{\sigma-1}{1+\gamma}\right)\left[\left(\Delta^k \ln s_{ijt}\right)\left(\Delta^k \ln \overline{p}^c_{ijt}\right)\right] \\
& + \left[\theta\left(\frac{1+\gamma}{\gamma}\right) - \frac{\theta-\gamma}{\gamma}(\sigma-1)\right]\left[\left(\Delta^k \ln \overline{p}^c_{ijt}\right)\left(\Delta^k \ln tar_{ijt}\right)\right] \\
& - \left(\frac{\theta}{\gamma} + \frac{\theta-\gamma}{\gamma}\right)\left[\left(\Delta^k \ln s_{ijt}\right)\left(\Delta^k \ln tar_{ijt}\right)\right] \\
& - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\left(\Delta^k \ln tar_{ijt}\right)^2 \\
& + \left[\theta\left(\frac{1+\gamma}{\gamma}\right) - \frac{\theta-\gamma}{\gamma}(\sigma-1)\right]\left[\left(\Delta^k \ln \overline{p}^c_{ijt}\right)\left(\Delta^k \ln trans_{ijt}\right)\right] \\
& - \left(\frac{\theta}{\gamma} + \frac{\theta-\gamma}{\gamma}\right)\left[\left(\Delta^k \ln s_{ijt}\right)\left(\Delta^k \ln trans_{ijt}\right)\right] \\
& - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\left(\Delta^k \ln trans_{ijt}\right)^2 \\
& - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\left[\left(\Delta^k \ln tar_{ijt}\right)\left(\Delta^k \ln trans_{ijt}\right)\right].
\end{aligned} \tag{D.16}$$

Rearranging terms to isolate $(\sigma - 1)(\Delta^k \ln \overline{p}^c_{ijt})^2$ on the LHS and dividing through by $(\sigma - 1)$

yields the estimating equation (where $\xi_{ijt} = \epsilon_{ijt}\psi_{ijt}$):

$$
\begin{aligned}
\left(\Delta^k \ln \overline{p}^c_{ijt}\right)^2 = {} & \frac{1}{(1+\gamma)(\sigma-1)} \left(\Delta^k \ln s_{ijt}\right)^2 + \left(\frac{\sigma-\gamma-2}{(1+\gamma)(\sigma-1)}\right) \left[\left(\Delta^k \ln s_{ijt}\right)\left(\Delta^k \ln \overline{p}^c_{ijt}\right)\right] \\
& - \left[\frac{\theta}{\sigma-1}\left(\frac{1+\gamma}{\gamma}\right) - \frac{\theta-\gamma}{\gamma}\right] \left[\left(\Delta^k \ln \overline{p}^c_{ijt}\right)\left(\Delta^k \ln tar_{ijt}\right)\right] \\
& + \left(\frac{\theta}{\gamma(\sigma-1)} + \frac{\theta-\gamma}{\gamma(\sigma-1)}\right) \left[\left(\Delta^k \ln s_{ijt}\right)\left(\Delta^k \ln tar_{ijt}\right)\right] \\
& + \theta\left(\frac{1+\gamma}{\gamma(\sigma-1)}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\left(\Delta^k \ln tar_{ijt}\right)^2 \\
& - \left[\frac{\theta}{\sigma-1}\left(\frac{1+\gamma}{\gamma}\right) - \frac{\theta-\gamma}{\gamma}\right] \left[\left(\Delta^k \ln \overline{p}^c_{ijt}\right)\left(\Delta^k \ln trans_{ijt}\right)\right] \\
& + \left(\frac{\theta}{\gamma(\sigma-1)} + \frac{\theta-\gamma}{\gamma(\sigma-1)}\right) \left[\left(\Delta^k \ln s_{ijt}\right)\left(\Delta^k \ln trans_{ijt}\right)\right] \\
& + \theta\left(\frac{1+\gamma}{\gamma(\sigma-1)}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\left(\Delta^k \ln trans_{ijt}\right)^2 \\
& + \theta\left(\frac{1+\gamma}{\gamma(\sigma-1)}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\left[\left(\Delta^k \ln tar_{ijt}\right)\left(\Delta^k \ln trans_{ijt}\right)\right] + \xi_{ijt}.
\end{aligned}
\tag{D.17}
$$

D.4 Moment and Identification Conditions' Derivations

Estimation of equation (44) produces consistent coefficient estimates under two conditions. The first is the moment condition, $\mathbb{E}(\xi_{ijt}) \equiv \mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = 0$; alternatively, the expectation can equal a constant as long as equation (44) includes an intercept (β_0). Recalling $\epsilon_{ijt} \equiv \sigma\Delta^k \ln e^{P1}_{ijt} + (1-\sigma)\Delta^k \ln e^{P2}_{ijt} + \Delta^k \ln e^{P3}_{ijt}$ and $\psi_{ijt} \equiv -\frac{1}{1+\gamma}\Delta^k \ln e^{P3}_{ijt} - \frac{1}{1+\gamma}\Delta^k \ln e^{P5}_{ijt}$:

$$
\begin{aligned}
\mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = {} & -\frac{\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P1}_{ijt}, \Delta^k \ln e^{P3}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P1}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]\right] \\
& - \frac{1-\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P2}_{ijt}, \Delta^k \ln e^{P3}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P2}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]\right] \\
& - \frac{1}{1+\gamma}\left[\text{var}(\Delta^k \ln e^{P3}_{ijt}) + [\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]^2\right] \\
& - \frac{\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P1}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P1}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& - \frac{1-\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P2}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P2}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& - \frac{1}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P3}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& = -\left(\frac{1}{1+\gamma}\right)\text{var}(\Delta^k \ln e^{P3}_{ijt}) \equiv -4\left(\frac{1}{1+\gamma}\right)\text{var}(\ln e^{P3}_{ijt}) \equiv -4\left(\frac{1}{1+\gamma}\right)\sigma^2_{\ln e^{P3}_{ij}}
\end{aligned}
\tag{D.18}
$$

where σ still represents the elasticity of substitution in consumption but σ_z^2 represents the variance over time of the variable z. Note that we assume the expected values of the time-differenced (as well as reference-exporter-country differenced) deviations from the Pareto distributions of the underlying variables are zero (e.g., $\mathbb{E}(\Delta^k \ln e_{ijt}^{P1}) = 0$) and the double-differenced deviations have constant variances. Accordingly, we assume the covariances are zero as well (e.g., $\text{cov}[\Delta^k \ln e_{ijt}^{P1}, \Delta^k \ln e_{ijt}^{P3}] = 0$). The moment condition is satisfied as the RHS in the equation above, $-\left(\frac{1}{1+\gamma}\right) \text{var}(\Delta^k \ln e_{ijt}^{P3})$, is a constant.

The second condition necessary for consistent estimates of the coefficients is the identification condition. Following Feenstra (1994), this condition is equation (48) in the text. In the context of our model in section 4.1.5 (ignoring $\Delta \ln f_{ijt}^R$), the necessary condition for identification is equation (49). Equation (49) is obtained by recalling again $\epsilon_{ijt} \equiv \sigma \Delta^k \ln e_{ijt}^{P1} + (1-\sigma)\Delta^k \ln e_{ijt}^{P2} + \Delta^k \ln e_{ijt}^{P3}$ and $\psi_{ijt} \equiv -\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3} - \frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5}$. Using ϵ_{ijt}:

$$
\begin{aligned}
\text{var}(\epsilon_{ijt}) &= \text{var}\left[\sigma \Delta^k \ln e_{ijt}^{P1} + (1-\sigma)\Delta^k \ln e_{ijt}^{P2} + \Delta^k \ln e_{ijt}^{P3}\right] \\
&= \sigma^2 \, \text{var}(\Delta^k \ln e_{ijt}^{P1}) + (1-\sigma)^2 \, \text{var}(\Delta^k \ln e_{ijt}^{P2}) + \text{var}(\Delta^k \ln e_{ijt}^{P3}) \\
&\quad + 2 \, \text{cov}\left[\sigma \Delta^k \ln e_{ijt}^{P1}, (1-\sigma)\Delta^k \ln e_{ijt}^{P2}\right] \\
&\quad + 2 \, \text{cov}\left[\sigma \Delta^k \ln e_{ijt}^{P1}, \Delta^k \ln e_{ijt}^{P3}\right] \\
&\quad + 2 \, \text{cov}\left[(1-\sigma)\Delta^k \ln e_{ijt}^{P2}, \Delta^k \ln e_{ijt}^{P3}\right] \\
&= \sigma^2 \sigma^2_{\Delta^k \ln e_{ij}^{P1}} + (1-\sigma)^2 \sigma^2_{\Delta^k \ln e_{ij}^{P2}} + \sigma^2_{\Delta \ln e_{ij}^{P3}}.
\end{aligned}
\tag{D.19}
$$

The latter can be inserted into the LHS of equation (48) to produce the LHS of equation (49). Using ψ_{ijt}:

$$
\begin{aligned}
\text{var}(\psi_{ijt}) &= \text{var}\left[-\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3} - \frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5}\right] \\
&= \left(\frac{1}{1+\gamma}\right)^2 \text{var}(\Delta^k \ln e_{ijt}^{P3}) + \left(\frac{1}{1+\gamma}\right)^2 \text{var}(\Delta^k \ln e_{ijt}^{P5}) \\
&\quad + 2 \, \text{cov}\left[-\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3}, -\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5}\right] \\
&= \left(\frac{1}{1+\gamma}\right)^2 \left[\sigma^2_{\Delta^k \ln e_{ij}^{P3}} + \sigma^2_{\Delta^k \ln e_{ij}^{P5}}\right].
\end{aligned}
\tag{D.20}
$$

The latter can be inserted into the RHS of equation (48) to produce the RHS of equation (49).

D.5 Moment and Identification Conditions' Derivations Including $\Delta \ln f_{ijt}^R$

In section 4.1.5, we addressed fixed trade-costs measurement. In that section, we noted that the majority of fixed trade costs are exporter specific or importer specific. However, for estimation, a residual measurement error exists, which we labeled f_{ijt}^R. With the introduction of this additional error term, we can readily modify the moment and identification conditions to accommodate this additional error term. As we will see, this has inconsequential effects on the moment and identification issues addressed earlier.

In this case, we need to redefine ϵ_{ijt} as:

$$\epsilon_{ijt} \equiv \sigma \Delta^k \ln e_{ijt}^{P1} + (1-\sigma)\Delta^k \ln e_{ijt}^{P2} + \Delta^k \ln e_{ijt}^{P3} + a^D \Delta \ln f_{ijt}^R \tag{D.21}$$

where a^D is a constant, and redefine ψ_{ijt} as:

$$\psi_{ijt} \equiv -\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3} - \frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5} + a^S \Delta \ln f_{ijt}^R \tag{D.22}$$

where a^S is another constant. Hence, the moment condition becomes:

$$\begin{aligned}
\mathbb{E}(\epsilon_{ijt}\psi_{ijt}) = & -\frac{\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P1}_{ijt}, \Delta^k \ln e^{P3}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P1}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]\right] \\
& - \frac{1-\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P2}_{ijt}, \Delta^k \ln e^{P3}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P2}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]\right] \\
& - \frac{1}{1+\gamma}\left[\text{var}(\Delta^k \ln e^{P3}_{ijt}) + [\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]^2\right] \\
& - \frac{a^D}{1+\gamma}\left[\text{cov}[\Delta^k \ln f^{R}_{ijt}, \Delta^k \ln e^{P3}_{ijt}] + [\mathbb{E}(\Delta^k \ln f^{R}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})]\right] \\
& - \frac{\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P1}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P1}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& - \frac{1-\sigma}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P2}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P2}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& - \frac{1}{1+\gamma}\left[\text{cov}[\Delta^k \ln e^{P3}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& - \frac{a^D}{1+\gamma}\left[\text{cov}[\Delta^k \ln f^{R}_{ijt}, \Delta^k \ln e^{P5}_{ijt}] + [\mathbb{E}(\Delta^k \ln f^{R}_{ijt})][\mathbb{E}(\Delta^k \ln e^{P5}_{ijt})]\right] \\
& + a^D a^S\left[\text{var}(\Delta^k \ln f^{R}_{ijt}) + [\mathbb{E}(\Delta^k \ln f^{R}_{ijt})]^2\right] \\
& + \sigma a^S\left[\text{cov}[\Delta^k \ln e^{P1}_{ijt}, \Delta^k \ln f^{R}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P1}_{ijt})][\mathbb{E}(\Delta^k \ln f^{R}_{ijt})]\right] \\
& + (1-\sigma) a^S\left[\text{cov}[\Delta^k \ln e^{P2}_{ijt}, \Delta^k \ln f^{R}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P2}_{ijt})][\mathbb{E}(\Delta^k \ln f^{R}_{ijt})]\right] \\
& + a^S\left[\text{cov}[\Delta^k \ln e^{P3}_{ijt}, \Delta^k \ln f^{R}_{ijt}] + [\mathbb{E}(\Delta^k \ln e^{P3}_{ijt})][\mathbb{E}(\Delta^k \ln f^{R}_{ijt})]\right] \\
= & -\left(\frac{1}{1+\gamma}\right)\text{var}(\Delta^k \ln e^{P3}_{ijt}) + a^D a^S\,\text{var}(\Delta^k \ln f^{R}_{ijt}) \qquad \text{(D.23)}
\end{aligned}$$

which still satisfies the moment condition.

The second condition necessary for consistent estimates of the coefficients is the identification condition. The extension has inconsequential effects on the identification condition.

Using the redefined ϵ_{ijt} from above:

$$\begin{aligned}
\text{var}(\epsilon_{ijt}) &= \text{var}\left[\sigma\Delta^k \ln e_{ijt}^{P1} + (1-\sigma)\Delta^k \ln e_{ijt}^{P2} + \Delta^k \ln e_{ijt}^{P3} + a^D \Delta^k \ln f_{ijt}^{R}\right] \\
&= \sigma^2 \text{var}(\Delta^k \ln e_{ijt}^{P1}) + (1-\sigma)^2 \text{var}(\Delta^k \ln e_{ijt}^{P2}) + \text{var}(\Delta^k \ln e_{ijt}^{P3}) + (a^D)^2 \text{var}(\Delta^k \ln f_{ijt}^{R}) \\
&\quad + 2\,\text{cov}\left[\sigma\Delta^k \ln e_{ijt}^{P1}, (1-\sigma)\Delta^k \ln e_{ijt}^{P2}\right] \\
&\quad + 2\,\text{cov}\left[\sigma\Delta^k \ln e_{ijt}^{P1}, \Delta^k \ln e_{ijt}^{P3}\right] \\
&\quad + 2\,\text{cov}\left[(1-\sigma)\Delta^k \ln e_{ijt}^{P2}, \Delta^k \ln e_{ijt}^{P3}\right] \\
&\quad + 2\,\text{cov}\left[\sigma\Delta^k \ln e_{ijt}^{P1}, a^D \Delta^k \ln f_{ijt}^{R}\right] \\
&\quad + 2\,\text{cov}\left[(1-\sigma)\Delta^k \ln e_{ijt}^{P2}, a^D \Delta^k \ln f_{ijt}^{R}\right] \\
&\quad + 2\,\text{cov}\left[\Delta^k \ln e_{ijt}^{P3}, a^D \Delta^k \ln f_{ijt}^{R}\right] \\
&= \sigma^2 \sigma^2_{\Delta^k \ln e_{ij}^{P1}} + (1-\sigma)^2 \sigma^2_{\Delta^k \ln e_{ij}^{P2}} + \sigma^2_{\Delta \ln e_{ij}^{P3}} + (a^D)^2 \sigma^2_{\Delta^k \ln f_{ij}^{R}}. \qquad \text{(D.24)}
\end{aligned}$$

Using the redefined ψ_{ijt} from above:

$$\begin{aligned}
\text{var}(\psi_{ijt}) &= \text{var}\left[-\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3} - \frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5} + a^S \Delta^k \ln f_{ijt}^{R}\right] \\
&= \left(\frac{1}{1+\gamma}\right)^2 \text{var}(\Delta^k \ln e_{ijt}^{P3}) + \left(\frac{1}{1+\gamma}\right)^2 \text{var}(\Delta^k \ln e_{ijt}^{P5}) + (a^S)^2 var(\Delta^k \ln f_{ijt}^{R}) \\
&\quad + 2\,\text{cov}\left[-\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3}, -\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5}\right] \\
&\quad + 2\,\text{cov}\left[-\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P3}, a^S \Delta^k \ln f_{ijt}^{R}\right] \\
&\quad + 2\,\text{cov}\left[-\frac{1}{1+\gamma}\Delta^k \ln e_{ijt}^{P5}, a^S \Delta^k \ln f_{ijt}^{R}\right] \\
&= \left(\frac{1}{1+\gamma}\right)^2 \sigma^2_{\Delta^k \ln e_{ij}^{P3}} + \left(\frac{1}{1+\gamma}\right)^2 \sigma^2_{\Delta^k \ln e_{ij}^{P5}} + (a^S)^2 \sigma^2_{\Delta^k \ln f_{ij}^{R}}. \qquad \text{(D.25)}
\end{aligned}$$

The latter two results can be inserted into the RHS of equation (48) to produce an easily adjusted version of the RHS of equation (49).