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

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

July 30, 2021

Abstract: We address two potentially related puzzles in the international trade literature. First, two parameters are central to several modern quantitative models of bilateral international trade flows: the elasticity of substitution in consumption, or Armington elasticity (σ), and the inverse index of heterogeneity of firms' productivities (θ). However, structural parameter estimation applications using the seminal Feenstra (1994) econometric methodology typically focus on estimates of only the Armington elasticity (σ) and a bilateral export supply elasticity – which we will term γ. Second, 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 address theoretically and quantitatively the importance of accounting for increasing marginal costs (via γ) – alongside firm heterogeneity – in understanding the relative impacts on trade, extensive margins, intensive margins, and welfare of reducing fixed trade costs and variable trade costs – the two elements common to modern trade agreements. One illustrative quantitative implication for U.S. trade policy is that, under constant marginal costs, fixed trade costs would have to be reduced by 88 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 15 percent.

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

JEL Classification Numbers: F1

* *Acknowledgements*: We are grateful to Scott Baier, 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 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

We address two potentially related puzzles in the international trade literature. 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, also referred to in trade as the "Armington elasticity," cf. Feenstra et al. (2018). Parameter σ 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). Although σ has a half-century-old presence, the last 20 years have witnessed the surfacing of a second important parameter in trade models, a (inverse) measure of heterogeneity of firms' productivities, θ. Motivated by 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) and is one of two sufficient statistics to measure welfare effects of trade liberalizations in a set of quantitative trade models with certain assumptions and restrictions, cf., Arkolakis et al. (2012).

The first puzzle is that a common assumption to all these models is *constant* marginal costs. Yet, by contrast, the most widely respected structural approach for estimating bilateral import demand (Armington) elasticities – the "Feenstra method" – assumes increasing marginal costs of exporting to foreign markets by assuming bilateral export supply prices are positive functions of the level of exports to foreign markets. Although σ and θ play central roles now in trade theory and calibration exercises of new quantitative trade models, this third parameter – the bilateral export supply elasticity – has been ignored. Yet, it has been crucial for structural estimation of σ – and potentially, in the future, of θ – using the seminal econometric methodology of Feenstra (1994), cf., Broda and Weinstein (2006) for computing numerically the gains from variety, Broda et al. (2008) and Ossa (2016) for estimating the relationship between optimal tariffs and market power, and Feenstra et al. (2018) for estimating upper and lower level elasticities of substitution. However, the bilateral export supply elasticity – which we will refer to as γ – 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), positively-sloped bilateral export supply curves were assumed even though the underlying theoretical Krugman model features horizontal supply curves. Moreover, in a recent study allowing firm heterogeneity based upon a standard Melitz model with constant marginal costs, Feenstra et al. (2018) introduce an equation that "plays the role of a supply curve" (p. 140). We will address this first puzzle by incorporating increasing marginal costs into a standard Melitz model of trade. This will generate extensive and intensive trade-margin elasticities with respect to variable *and* fixed trade costs that are functions of σ, θ, *and* γ and motivate bilateral import demand and export supply functions that are estimable using the Feenstra method (that can potentially estimate σ, γ, and θ *simultaneously*).

The second puzzle is that 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, summarized comprehensively in Hofmann et al. (2017). The database documents the extensive growth in deep provisions 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 (i.e., the intensive margin) 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. Furthermore, recent empirical work using gravity equations indicates economically and statistically significant effects of such provisions on trade flows, cf., Kohl et al. (2016), Baier and Regmi (2020), Crowley et al. (2020), and Fontagne et al. (2020).

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 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 *40 percent* reduction in bilateral fixed trade costs to achieve the equivalent gain in welfare as a 5 percent reduction in *ad valorem* variable trade costs. Such estimates suggest the welfare gains from trade liberalization are more readily attained via tariff-rate reductions. If so, why have countries increasingly pursued deep trade agreements? In the context of our Melitz model, we will show, based upon the empirical results from our novel econometric technique controlling for firm heterogeneity, that allowing for empirically-justified increasing marginal costs alters the *ad valorem* variable-trade-cost elasticity relative to the fixed-trade-cost elasticity, such that *much smaller reductions* in fixed trade costs lead to equivalent increases in welfare as small changes in variable trade costs. Thus, even though Arkolakis et al. (2012), or ACR, clarify that – for several modern

[1]We will discuss Balistreri et al. (2011) and Dixon et al. (2016) later.

quantitative trade models – only two statistics (the *ad valorem* trade elasticity from a gravity equation and changes in the domestic output share of expenditures) are sufficient to quantify the overall welfare gains from trade liberalization, precise and unbiased estimates of σ, γ, and θ remain important for conducting policy choices in a world with variable and fixed trade costs and deep trade agreements.

We summarize four tangible goals of this paper to address these two related puzzles. First, we generalize methodologically the seminal Feenstra-Broda-Weinstein econometric framework to allow potentially for simultaneous estimation of σ, γ, and θ. Second, to incorporate γ in a non *ad hoc* manner, we generalize a standard Melitz model of international trade to *allow* for the possibility of increasing marginal costs to supply destination output. Third, to illustrate our methodology (but constrained by availability of high-quality data), we use our augmented Feenstra-Broda-Weinstein approach to generate new estimates of σ and γ – accounting explicitly for firm heterogeneity. Specifically, we can control for firm heterogeneity, but data limitations preclude credible estimation of θ. Fourth, combining the first three contributions, we provide two numerical calibrations to illustrate the relevance of our first three contributions and relate these insights to understanding the increasing importance of deep international trade agreements in the world.

Because our sole modification of the Melitz model is to allow for the (empirically-motivated) possibility of increasing marginal costs in serving foreign markets, we illustrate first the role of positively-sloped bilateral export supply curves in the simplest (Armington or Krugman) trade model focusing only on the intensive margin. Figure 1 illustrates the attenuation of the intensive margin elasticity in the presence of a positively-sloped bilateral export supply curve, consistent with increasing marginal costs (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, $\Delta \ln IM_{ij} = \overline{AC} < \overline{AB}$. This figure illustrates clearly that – under CMC – the trade elasticity is a function solely of the elasticity of substitution. However, under IMC, the trade elasticity is a function of the elasticity of substitution in consumption *and* an index of the shape of the supply curve.

In econometric specifications to estimate σ, Feenstra (1994), Broda and Weinstein (2006), and Soderbery (2015) all allow for the possibility of positively-sloped bilateral export supply curves; in a domestic setting, Hottman et al. (2016) explicitly allows for IMC in supplying products. In all the trade studies, evidence surfaces that bilateral export supply curves *are* positively sloped; in fact, in Hottman et al. (2016), the median marginal cost elasticity of output has a value of 0.16, implying a supply elasticity of 6 – which is far less than ∞.[2]

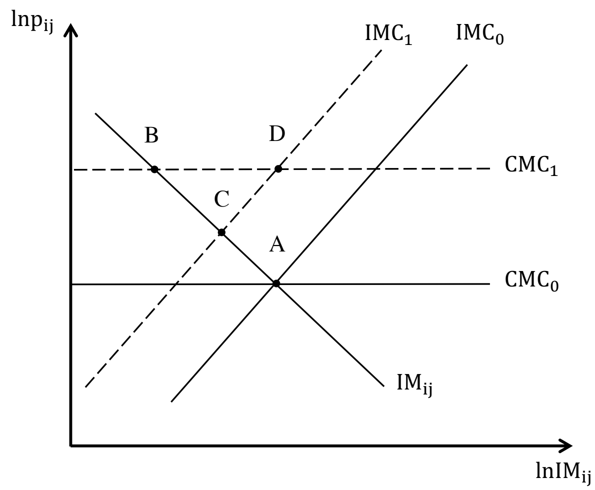

Figure 1: Increasing Marginal Costs vs. Constant Marginal Costs

We now summarize our paper's contributions chronologically. Motivated by the preceding discussion, our paper's first contribution is theoretical. We allow for IMC in an otherwise standard Melitz model of international trade.[3] We generate three novel theoretical findings. First, we derive a gravity equation that is very similar to the ones in Chaney (2008), Redding (2011), and ACR, except that the extensive margin elasticity – and the "trade elasticity" – with respect to (*ad valorem*) variable trade costs are magnified. The magnification results from given changes in *ad valorem* variable trade costs having larger impacts on export cutoff productivities under IMC relative to CMC, causing larger increases in the number of exporting firms and aggregate trade flows. Yet, the variable-trade-cost intensive margin elasticity is diminished (and a function of γ and σ), consistent with Figure 1. An important implication of this is that trade-policy liberalizations with IMC will have more firm entry and exit and labor reallocations than under CMC. Thus, our paper contributes to understanding better the empirical evidence on the dominant role of the extensive margin in response to variable-trade-cost changes.

Second, by accounting for IMC, our export-fixed-cost trade elasticity is magnified relative to that in Chaney (2008), Redding (2011), and ACR. The intuition is straightforward.

[2]Interestingly, the last column in Table 3.1 in Head and Mayer (2014) does acknowledge the assumption of increasing marginal cost curves in the Armington and Krugman models in Bergstrand (1985) and Baier and Bergstrand (2001); we return to this issue later.

[3]For our purposes, a Melitz (2003) model is preferable to an Eaton and Kortum (2002) framework for two reasons. First, the Melitz model has both an intensive and extensive margin, whereas the Eaton-Kortum model features only an extensive margin. Second, the Melitz model allows examination of fixed-trade-cost effects.

The export-fixed-cost elasticity in the earlier papers is $1 - \theta/(\sigma - 1)$, which is a function of the extensive-margin variable-trade-cost elasticity (θ) relative to the intensive-margin variable-trade-cost elasticity ($\sigma - 1$). With increasing marginal costs, the magnification of the variable-trade-cost extensive-margin elasticity and the attenuation of the variable-trade-cost intensive-margin elasticity cause the export-fixed-cost trade elasticity to increase (in absolute value). Thus, the export-fixed-cost trade elasticity is larger (in absolute terms) and a function of all three parameters: σ, θ, and γ.[4] Moreover, a further result that the fixed-trade-cost elasticity is also magnified relative to the variable-trade-cost elasticity will be important later in understanding the welfare-equivalent impacts of fixed-trade-cost liberalizations relative to variable-trade-cost liberalizations in deep trade agreements.

Third, in our framework, the trade elasticity is the heterogeneity parameter, θ, scaled by one plus the marginal cost elasticity of output, $1 + 1/\gamma$. Consequently, allowing for increasing marginal costs *diminishes* the welfare effect of a given change in the domestic trade share (and 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 $1 + 1/\gamma$.[5]

Our second contribution of this paper is methodological. We show that the Feenstra-Broda-Weinstein structural estimation framework can be generalized to include a micro-founded bilateral export supply elasticity (the inverse marginal cost elasticity with respect to destination output) and to account explicitly for firm heterogeneity. The seminal study articulating a structural approach to estimating elasticities of substitution in consumption using disaggregate international trade flow data and prices (unit values) is Feenstra (1994).[6]

[4] An important implication is that we address an outstanding empirical finding noted in Feenstra (2016) (page 168) that empirical estimates of the Pareto curvature parameter, θ, tend to fall below empirical estimates of the elasticity of substitution minus unity, $\sigma - 1$. In the context of the Melitz-Redding models, the former must exceed the latter to solve the Melitz-Redding model. However, with increasing marginal costs, θ can be less than $\sigma - 1$ *as long as* θ is larger than $(\sigma - 1)\frac{\gamma}{\sigma+\gamma}$.

[5] Feenstra (2010) demonstrates in a two-country model that the transformation curve between domestic and exported *output-adjusted* varieties is a concave constant-elasticity-of-transformation (CET) function. In our paper, we additionally derive the exact linear functional relationship between the marginal cost elasticity of output and the CET parameter discussed in Feenstra (2010).

[6] As discussed comprehensively in Hillberry and Hummels (2013), there are alternative approaches to estimating Armington elasticities (each with drawbacks). Hillberry and Hummels (2013) categorize these approaches into four topics: early time-series estimates of import demand functions using incomes and prices (which suffer from simultaneity), instrumental variables approach to estimating import demand and export supply, more recent cross-sectional and panel estimates using *ad valorem* trade-cost measures and using fixed effects for export supply variation, and the Feenstra-Broda-Weinstein method to estimate bilateral import demand and export supply functions. See that survey's section 18.3.

Broda and Weinstein (2006) extended Feenstra's approach to account for the introduction of completely new product categories. Soderbery (2015) extended both those studies by implementing a novel limited information maximum likelihood technique. Importantly, all three studies introduce positively-sloped bilateral export supply curves to a destination market by *assuming* that the price of the product-variety sold by the exporter to a particular importing country (the United States) is a log-linear function of the bilateral quantity supplied. Moreover, drawing off of the Krugman (1980) trade model using representative (homogenous) firms, the estimation of the Armington elasticities and bilateral export supply elasticities using disaggregated industry-level trade flow and import unit value data in Broda and Weinstein (2006) and Soderbery (2015) ignores the heterogeneity of firms' productivities. Although Feenstra et al. (2018) estimated Armington (micro) elasticities in the context of a Melitz model, that paper's theoretical foundation assumed constant marginal costs as in a standard Melitz model, providing an equation (18) that "plays the role of a supply curve" (p. 140). Accordingly, their resulting estimation technique for (micro) Armington elasticities and bilateral export supply elasticities involves only two right-hand-side (RHS) variables, analogous to the standard Feenstra-Broda-Weinstein approach (see their equations (19) and (20)).[7] In our paper allowing IMC, we show that the Feenstra-Broda-Weinstein structural estimation framework can be generalized to include a micro-founded bilateral export supply elasticity (the inverse marginal cost elasticity with respect to destination output), akin to Vannoorenberghe (2012), and to account explicitly for firm heterogeneity.[8] Unlike these previous studies, our approach distinctly recognizes the importance of how the mass of exporting firms depends not just on the exporting country's labor-force size but also on its 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. Our extension of the Feenstra-Broda-Weinstein technique motivates the inclusion of 20 independent variables (6 unique variables and 14

[7] The focus of Feenstra et al. (2018), however, is estimating separately the (macro) elasticity of substitution between domestic and imported goods from the (micro) elasticity of substitution between imported varieties. In this fuller specification, the regression equation has five RHS variables.

[8] Vannoorenberghe (2012) introduced increasing marginal (production) costs in a two-country partial equilibrium framework to show that firms likely do not face constant marginal costs and so do not maximize profits on different markets independently of each other. While nested in a short-run context, before demand shocks are realized, firms decide whether to produce – and to export – and pay the corresponding fixed costs. After demand shocks are realized, each firm decides how much labor to demand and how much to sell on each of the two markets. In our model with IMC at the firm level, profit maximization is not independent across markets as any change in trade costs between a country pair alters, in general equilibrium, the wage rate in the origin country, causing potential quantities supplied to each market by the firm to change as a result of shifts in the bilateral supply curves; we address this later in our paper. As Bartelme et al. (2019) note, firms in the closed-economy model established in Hopenhayn (1992), the model undergirding Melitz (2003), can be subject to decreasing returns (p. 1137). In fact, Hopenhayn (1992) argues that the positive fixed cost is "equivalent to the existence of a *fixed outside opportunity cost for some resource* (e.g., managerial ability) used by the firm" (p. 1130; italics added.)

interaction terms) to appropriately estimate σ, γ, and θ simultaneously.

Our third contribution is empirical. Credible estimation of σ, γ, and θ in our extended, theory-motivated econometric specification requires detailed industry-level data on six characteristics in order to construct the 20 independent variables (including interaction terms). Unfortunately, at this time, reliable data is not available for all six variables. Specifically, for the first category of variables – trade flows and import unit values – reliable data exists. For a second category of variables novel to our approach – industry-level employment, wage rates, variable trade costs, and fixed trade costs – we can only control for these variables and their interactions using proxies. Consequently, our empirical implementation will allow us to accurately estimate σ and γ *controlling explicitly* for firm heterogeneity; however, data limitations preclude simultaneous credible estimation of θ at this time. Our empirical results further confirm the existence of IMC, with the (across-industry) median inverse marginal cost elasticity (γ) estimate ranging – across various specifications – between 5.74 and 6.31, implying marginal cost elasticities ranging between 0.16 and 0.17. Moreover, our marginal cost elasticity estimate of 0.16 is *precisely* the same median estimate in Hottman et al. (2016) using firm-level U.S. barcode data. An inverse elasticity of marginal costs to output of 6 is far below ∞, the latter used in the trade literature's benchmark model with CMC.

Finally, our fourth contribution is to illustrate the impact of recognizing increasing marginal costs on the estimated effects of deep trade agreements in the world. Goldberg and Pavcnik (2016) emphasized that trade economists have not paid sufficient attention to the study of the effects of trade-policy changes other than *ad valorem* tariff-rate changes. One of the most notable events concerning international trade since 1990 is the proliferation of economic integration agreements (EIAs), such as bilateral and plurilateral free trade agreements (FTAs). With most countries' Most-Favored-Nation tariff rates at 5 percent or less, increasingly such agreements have introduced liberalizations in the form of reductions of border and behind-the-border barriers that reduce *fixed* trade costs. Prominent examples are the proposed Transatlantic Trade and Investment Partnership (TTIP) agreement between the European Union and the United States and the replacement in 2020 of the North American Free Trade Agreement (NAFTA) with the United States-Mexico-Canada Agreement (USMCA).

While a major point in Goldberg and Pavcnik (2016) is the need for better measurement of such barriers and the changes in them, Goldberg and Pavcnik (2016) make clear that trade economists need to focus more on understanding the effects of reduced fixed trade costs on international trade and economic welfare. In this spirit, we conduct two numerical analyses. In the first exercise, we show that – similar to Feenstra (2010) – 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." 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. Yet, 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). So, 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 (empirically-justified) 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, in the case of the United States, we show that, under (empirically-unsupported) CMC, fixed trade costs would have to be reduced by 88 percent to provide the same increase in welfare as a reduction in variable trade costs of 3 percent. By contrast, under (empirically-justified) IMC, it would take only a *15 percent* reduction in fixed trade costs to increase welfare by the same amount as a 3 percent reduction in variable trade costs!

The remainder of this paper is as follows. In section 2, we introduce and solve our Melitz model with 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." Section 4 provides empirical estimates of the elasticity of substitution and bilateral export supply elasticity using our novel econometric approach. Section 5 provides 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. Section 6 concludes.

2 Theory

Our theoretical framework builds on the workhorse Melitz (2003) heterogeneous firms model of international trade. We depart from the particular framework in Melitz (2003) in two ways. First, as in Chaney (2008), Redding (2011), and numerous other papers, we allow for differences in countries' labor endowments and bilateral trade barriers and assume a Pareto distribution for productivity draws. The cross-country asymmetry and asymmetric trade barriers provide a more realistic framework for estimating the structural parameters of the model using disaggregated bilateral trade data. The Pareto distribution yields closed-form solutions that we can use to obtain clear theoretical predictions and to develop a novel econometric approach for the estimation. Second, as noted in the introduction, we allow for the possibility of increasing marginal costs of providing output to any market. It is reasonable intuitively to study a "more general" model – especially one that motivates the econometrically tractable structural bilateral import demand and bilateral export supply functions in Feenstra (1994), Broda and Weinstein (2006), and Soderbery (2015) – and let the data determine the slope of the bilateral export supply curve.

2.1 Preferences and Demand

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. 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:

$$U_j = \left[\int_{\nu \in \Omega_j} c_j(\nu)^{\frac{\sigma-1}{\sigma}} d\nu \right]^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1, \tag{1}$$

where $c_j(\nu)$ is the quantity consumed of variety ν, Ω_j is the (endogenous) set of varieties available for consumption in country j, and σ is the elasticity of substitution across varieties.

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

$$c_j(\nu) = E_j P_j^{\sigma-1} p_j^c(\nu)^{-\sigma}, \tag{2}$$

where E_j denotes aggregate expenditure in country j, $p_j^c(\nu)$ is the price of a unit of variety ν in country j facing the consumer, and P_j defined as:

$$P_j = \left[\int_{\nu \in \Omega_j} p_j^c(\nu)^{1-\sigma} d\nu \right]^{\frac{1}{1-\sigma}} \tag{3}$$

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 Firm Production

While seats sold in the United States ***appear nearly identical*** *to the seats we sell internationally, the internal components are very different.* Tom Downey, CEO of EVS Ltd.

For 25 years, the Feenstra-Broda-Weinstein econometric approach has assumed upward-sloping bilateral export supply curves in their econometric approach to estimate Armington bilateral import demand elasticities (σ) for various industries. Starting with Feenstra (1994), the bilateral (deterministic) "supply curve for these imports from country i" was specified (for some good g to some destination j) as $p_{ij} = q_{ij}^{\omega}$, cf., his equation (8).[9] As Soderbery (2018) recently summarized concisely, "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" in Feenstra and Weinstein (2017) 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.[10]

Because we are interested here in addressing simultaneously σ, γ, and θ, we need to introduce firm heterogeneity, as in Melitz (2003).[11] Feenstra and Romalis (2014) provides guidance. They developed an extension of the Melitz model to allow for endogenous quality choice by firms that provides a ready adaptable framework to allow for horizontal differentiation of a firm's good across multiple destinations. First, recall that Melitz (2003) cites likely sources of the important "export fixed costs" facing any firm considering selling to a foreign market (likely common across firms for any pair ij): exporting firms must employ labor to inform foreign buyers about their product (e.g., marketing); a firm may also have to use labor to "set up new distribution channels in the foreign country and conform to all the shipping rules specified by the foreign customs agency"; and labor is employed to research foreign standards and in many cases a firm "*adapts its product* to ensure that it conforms to foreign standards" (p. 1706; italics added). Feenstra and Romalis (2014) captured the latter element in their *Assumption 1*: "Firms may produce multiple products,

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

[10] Once again, we have modified notation in Feenstra and Weinstein (2017) to be consistent with that of the current paper.

[11] 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 likely vary across industries.

one for each potential market."[12] Second, we use this assumption in our model along with an adaptation of their *Assumption 2*: a firm with productivity φ producing in country i simultaneously chooses quantity $q_{ij}(\varphi)$ and free-on-board (fob) price $p_{ij}(\varphi)$ for each market j. Third, we adapt Feenstra and Weinstein (2017) and assume that marginal costs from each exporter in i to each importing country j ($mc_{ij}(\varphi)$) are increasing in output $q_{ij}(\varphi)$, or $mc_{ij}(\varphi) = (1/\gamma)_{ij0} q_{ij}^{1/\gamma}$.[13]

We now summarize the cost function formally. Production uses only one input, labor (or, as in Feenstra and Romalis (2014), a composite 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:

$$l_{ij}(\varphi) = f_{ij} + \frac{q_{ij}^{1+\frac{1}{\gamma}}}{\varphi}, \quad \gamma > 0, \tag{4}$$

where f_{ij} denotes the fixed costs of each firm in i to serve market j.[14] The special case of $i = j$ represents the demand for labor for domestic sales.[15] As implied by equation (4), fixed costs are common across firms for a given origin-destination pair, whereas marginal costs vary across firms for two reasons. First, as conventional to a Melitz model, more productive firms (with higher φ) need fewer workers to produce a given level of firm output.[16] Second, marginal costs are a function of output such that, all else equal, larger firms face higher

[12] For example, in correspondence with the CEO (Tom Downey) of EVS Ltd. of South Bend, Indiana, a domestic producer and exporter to several continents of specialized seating to the ambulance industry (EVS stands for "Emergency Vehicle Seating"), the firm not only faces export fixed costs but also slightly differentiates output in production due to regulatory constraints, making minor modifications to emergency vehicle seats *by destination market* to allow for different vehicle standards. Regarding export fixed costs, Downey notes, "First on the list was to attend several trade shows in countries such as Germany, United Arab Emirates, and Australia.... We realized that ... regulatory agencies, as well as compliance audits, were very different than within the United States." Regarding production, Downey noted, "In addition to investing time and money to learn what was required, we had to actually modify our products to meet international standards.... While seats sold in the United States *appear nearly identical* to the seats we sell internationally, the internal components are very different" (italics added).

[13] We also employ analogous assumptions to *Assumptions 4, 5* and *6* in Feenstra and Romalis (2014). For Assumption 4, we also assume productivity is Pareto distributed. For Assumption 5, we assume only *ad valorem* (iceberg) bilateral trade costs τ_{ij}. For Assumption 6, however, we assume fixed costs f_{ij} that are unrelated to productivity φ, as is more standard.

[14] 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.

[15] 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 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).

[16] 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.

marginal costs to provide output to each destination 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 Melitz (2003), ACR, and most workhorse trade models mentioned earlier.[17]

We note that our formulation provides an econometrically tractable microeconomic foundation for our empirical work to capture the positively-sloped bilateral export supply curves estimated using the Feenstra-Broda-Weinstein technique. This approach will allow us also to compare our empirical estimates of this elasticity using disaggregate international trade data to those in Hottman et al. (2016) using U.S. firm-level barcode data. At the same time, our specification in this paper of increasing marginal costs helps move the Melitz-Chaney-Redding framework in the direction of many small exporting countries in an empirically tractable manner, as addressed in Arkolakis (2010), noting "The basic idea put forward is that firms reach individual consumers rather than the market in its entirety" (p. 1152). Each market is assumed to be composed of individual consumers for which "the cost per consumer may differ depending on how many consumers have already been reached" (p. 1156). This assumption in his model captures the idea that "as marketing expenditures increase, efficiency declines" (p. 1156). As noted in Anderson (2011), the marketing element in Arkolakis (2010) effectively has a fixed-trade-cost component and a variable-trade-cost component subject to diminishing returns.[18] However, to be consistent with the Feenstra-Broda-Weinstein approach, we capture positively sloped bilateral export supply curves in an empirically tractable manner.

As standard, firms can sell their output to consumers in foreign markets, but face additional costs of shipment. As common in this literature, we model such trade barriers using iceberg trade costs. We assume firms in country i must ship $\tau_{ij} \geq 1$ units of output for each one unit of output to arrive 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} and q_{ij} denote the price received and

[17] To see the interpretation of the cost function in Feenstra and Weinstein (2017)'s equation (23), marginal costs in equation (4) above simplify to $mc_{ij}(\varphi) = \left[(1 + \frac{1}{\gamma}) \frac{w_i}{\varphi_{(ij)}} \right] q_{ij}(\varphi)^{1/\gamma}$ with $\gamma > 0$.

[18] Arkolakis (2010) introduced two components to export fixed costs. The first was increasing returns to scale; per consumer export fixed costs fall with destination population size. The second, "increasing marginal (market-penetration) costs," implied that per consumer export fixed costs increase with additional numbers of destination consumers. Because his framework introduces in a particular manner increasing marginal costs into the fixed-trade-cost term, the optimal pricing decision of firms corresponds to the traditional CMC case (cf., his equation (7)). One of the benefits of the assumption in Arkolakis (2010) that increasing marginal penetration costs are embedded in the export fixed cost terms is that – for his calibrations – he avoids having to specify "as many [export] fixed costs as destinations" (p. 1164). However, as noted in Anderson (2011), the introduction to his model of numerous additional parameters is useful for simulations, but "is not econometrically tractable" (p. 140). By contrast, our paper introduces only one new parameter – the elasticity of marginal costs with respect to destination output – to capture analogously "increasing marginal (market-penetration) costs" in an econometrically tractable manner consistent with the Feenstra-Broda-Weinstein approach.

the quantity shipped at the factory gate. 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 $p_{ij}^c = \tau_{ij} p_{ij}$.

Importantly, our approach above is strongly motivated by the fact that we will estimate Feenstra-Broda-Weinstein-type bilateral import demand and export supply functions to determine potentially values for σ, γ, and θ – and variable-trade-cost and export-fixed-cost trade elasticities – in a world with asymmetric country sizes and asymmetric bilateral trade costs. The assumptions yield tractable analytical solutions compared to a model with increasing marginal cost applied to total firm output. In the latter case, where firms' outputs are *not* distinguished by destination market, extending the Melitz model for increasing marginal costs in total output to allow for cross-country heterogeneity in size and asymmetric bilateral trade barriers increases the complexity of the model. In that setup, the total output of a firm depends on the choices of markets it serves which, in turn, depend on the marginal costs faced. The interdependence of production costs and selection is too complex with asymmetric countries and trade barriers to obtain closed-form solutions we can use in the empirical model and in calibration exercises. Nevertheless, this more traditional case of increasing marginal costs *in total output* needs to be considered. Accordingly, after we lay out the asymmetric case model in sections 2.3-2.5, and derive the important implications of the model under asymmetry in section 3 (including the *ad valorem* variable-trade-cost and export-fixed-cost trade elasticities), we derive similar implications for the case of IMC in total output in the special case of symmetry in (Online) Appendix C. In particular, we show that under an assumption of a large number of countries, the variable-trade-cost and export-fixed cost trade elasticities are *identical* in both models.

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.

Similar to Feenstra and Romalis (2014), firm profits in each market are given by revenues less labor costs:

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

where the second equality uses production function (4). We note that all country-i firms with productivity φ will charge the same price in destination j such that, henceforth, a consumer's variety ν can be identified by an origin country and a firm productivity only (i.e., $p_j(\nu) \equiv p_{ij}(\varphi)$).

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:[19]

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

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 $1 + 1/\gamma$ when marginal costs are increasing in output compared to one where marginal costs are constant (i.e., $\gamma \to \infty$). Second, prices are an increasing function of quantity; this provides a rationale for the upward-sloping bilateral export supply functions in Feenstra (1994), Broda and Weinstein (2006), and Soderbery (2015). 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:

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

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:

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

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

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

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

Because $\gamma/(\sigma+\gamma)$ and $\gamma/(1+\gamma)$ 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}^*).[20] 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. From equation (9), the ratio of export and domestic cutoff productivities is:

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

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.[21]

To anticipate general equilibrium considerations addressed later in this section, it is important to note that – as in Vannoorenberghe (2012) – the introduction of IMC causes exogenous shocks to any country pair (ij) to influence – not just behavior in market ij – but behavior in all bilateral markets, due to general equilibrium effects on wage rates; this applies also to selection into exporting. For instance, an exogenous decline in *ad valorem* bilateral trade costs, τ_{ij}, will reduce the price of firm φ's differentiated variety to consumers in j, which may raise demand for $q_{ij}(\varphi)$, if firm φ is already supplying that market (or may cause firm φ to enter that market). This rise in demand causes a movement up and along firm φ's bilateral export supply curve. However, the subsequent rise in labor demand raises wage rate w_i, which then has general equilibrium effects, such as shifting up the bilateral

[20] Detailed derivations are available in section 3 of Appendix A.

[21] The findings in Lu (2010) to the contrary are explained in Dai et al. (2016) as processing trade.

export supply curves of firm φ (and other firms) to other (non-j) markets (and even then shifting up the export supply curve of φ for market ij).

2.4 Trade Flows

We can now characterize equilibrium aggregate trade flows.[22] Imposing the labor-market-clearing condition, we can solve for the mass of incumbent firms in each country i that sell to each destination j:

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

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

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

We use the goods-market-clearing condition, $R_i = E_i$, to express trade flows as a gravity equation. It follows that total expenditure in country j can be expressed as:

$$E_j = \sum_{k=1}^{N} X_{kj} = \left[\frac{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}{\theta - \frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right]\left(\frac{1}{\delta f^e}\right)\sum_{k=1}^{N}\frac{w_k L_k f_{kj}}{(\varphi_{kj}^*)^\theta}. \tag{13}$$

Together the results in equations (12) and (13) imply that the share of country j's expenditure on goods supplied by country i is given by:

$$\lambda_{ij} \equiv \frac{X_{ij}}{E_j} = \frac{w_i L_i f_{ij}(\varphi_{ij}^*)^{-\theta}}{\sum_{k=1}^{N} w_k L_k f_{kj}(\varphi_{kj}^*)^{-\theta}}. \tag{14}$$

Using the definition of the productivity threshold in equation (9) to substitute for φ_{ij}^* (and analogously φ_{kj}^*) in equation (14) and rearranging yields:

$$X_{ij} = \left(\frac{L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}\tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} L_k w_k^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)}\tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}}\right) w_j L_j. \tag{15}$$

This expression for bilateral trade flows takes the usual gravity-equation form. It shows that trade between countries is increasing in the size of the trading partners ($w_i L_i$ and $w_j L_j$), decreasing in the trade barriers between them (τ_{ij} and f_{ij}), and a function of country j's multilateral price term (the denominator of equation (15)) that captures the influence of

[22] Derivation details are provided in sections 4–8 of Appendix A.

trade costs between all country pairs of j, including j's intra-national trade costs. We note that when $\gamma \to \infty$ the benchmark result obtains.[23]

As shown in section 11 of Appendix A, equation (15) 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.

2.5 General Equilibrium

In section 9 of 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 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 important 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. For a given set of structural parameters, the trade elasticities are different under IMC relative to the benchmark CMC. Moreover, the variable-trade-cost trade elasticity changes *relative* to the fixed-trade-cost trade elasticity, 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, 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."

[23] 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.

3.1 Trade Elasticities

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

$$\begin{aligned}\varepsilon_\tau \equiv -\frac{\partial X_{ij}}{\partial \tau_{ij}}\frac{\tau_{ij}}{X_{ij}} &= -\left[\underbrace{-\theta\left(\frac{1+\gamma}{\gamma}\right)}_{\text{extensive}} + \underbrace{(1-\sigma)\left(\frac{1+\gamma}{\sigma+\gamma}\right)}_{\text{intensive}} + \underbrace{(\sigma-1)\left(\frac{1+\gamma}{\sigma+\gamma}\right)}_{\text{compositional}}\right] \\ &= \theta\left(\frac{1+\gamma}{\gamma}\right).\end{aligned} \tag{16}$$

Following Head and Mayer (2014), we decompose this trade elasticity into extensive-margin, intensive-margin, and compositional-margin components.[24] 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.[25] 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 heterogeneous firm 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.

In contrast to the benchmark (Eaton-Kortum and Melitz-Chaney-Redding) models with firm heterogeneity – our elasticity of trade with respect to τ_{ij} depends on parameters governing the productivity distribution, θ, and the inverse marginal cost elasticity of output, γ. In the case of $\gamma < \infty$, the trade elasticity is magnified; the *ad valorem* trade elasticity depends on θ, as in the benchmark, but is scaled up by the additional term $1 + 1/\gamma$. The intuition can be traced back to equations (9) and (11). Equation (9) reveals that, with IMC,

[24] We note that this composition 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."

[25] 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.

a fall in τ_{ij} has a magnified effect of $1 + 1/\gamma$ 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 to j. Due to diminishing marginal returns, the trade elasticity is augmented and is now a nonlinear function of *both* parameters of the productivity distribution.

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

$$\varepsilon_f \equiv -\frac{\partial X_{ij}}{\partial f_{ij}}\frac{f_{ij}}{X_{ij}} = -\left[\underbrace{-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}_{\text{extensive}} + \underbrace{0}_{\text{intensive}} + \underbrace{1}_{\text{compositional}}\right] = \frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)} - 1. \quad (17)$$

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 $\varepsilon_f = \theta/(\sigma-1) - 1$.[26] 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 down of the denominator of this elasticity by $\frac{\gamma}{\sigma+\gamma}$ 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.

As mentioned earlier, we also 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 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 (16) and (17), respectively. This suggests our results may be a good approximation to a larger class of models.

We have shown that, conditional on a set of structural parameters, the elasticities of trade are magnified under IMC. As a result, any trade-policy liberalization or transport-cost

[26] 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 (17) 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 $0 < \frac{\gamma}{\sigma+\gamma} < 1$.

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 (16) and (17) 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. 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). To understand why fixed-trade-cost reductions have a relatively larger effect on trade than variable-trade-cost reductions with IMC, consider equations (9) and (12). 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 composition-margin effect). Due to IMC, the trade elasticity scales up θ by $1 + 1/\gamma$ due to diminishing marginal returns, cf., equation (8).

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^*_{uj} less than proportionately (i.e., φ^*_{ij} is proportionate to $f_{ij}^{1/(\sigma-1)}$). The introduction of IMC causes both the variable-trade-cost trade elasticity to increase from θ to $\theta(1 + \frac{1}{\gamma})$, but also the intensive-margin effect to decline from $\sigma - 1$ to $\frac{1+\gamma}{\sigma+\gamma}(\sigma - 1)$. This is confirmed by rewriting equation (9) as:

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

In the penultimate section of this paper, we conduct a counterfactual analysis to show – based upon median estimates of σ, γ, and θ – how much IMC increases the welfare benefit of a fixed-trade-cost liberalization relative to a variable-trade-cost one, providing insight into the increasing prominence of deep trade agreements.

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, a main result in ACR holds; two sufficient statistics to measure the welfare effects of international trade-cost shocks remain the domestic trade (or expenditure) share (i.e., the share of domestic expenditure on domestic output, or the "intra-national" trade share) and the "trade elasticity" (i.e., the elasticity of trade with respect to *ad valorem* variable trade costs, τ_{ij}). Second, we explain intuitively, and in the context of our model, why the "trade elasticity" under IMC is "magnified" relative to that under CMC.

First, in section 14 of 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:

$$\hat{W}_j = \hat{\lambda}_{jj}^{-1/[\theta(1+\frac{1}{\gamma})]} = \hat{\lambda}_{jj}^{-1/\varepsilon_\tau}, \tag{18}$$

where $\hat{\lambda}_{jj} \equiv \lambda'_{jj}/\lambda_{jj}$ is the (gross) change in the share of domestic expenditure 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:

$$G_j = 1 - \lambda_{jj}^{1/[\theta(1+\frac{1}{\gamma})]} = 1 - \lambda_{jj}^{1/\varepsilon_\tau}, \tag{19}$$

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, it is important to 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 according to:

$$q_{ij}(\varphi^*_{ij}) \propto \left(\varphi^*_{ij}\right)^{\frac{\gamma}{1+\gamma}} \tag{20}$$

due to diminishing marginal returns. This result shows that, in general, there is a concave relationship between the productivity cutoff and the output. This is the result of two opposing effects. On the one hand, the direct effect of an increase in productivity is to lower the cost of production of a given number of units. On the other hand, the indirect effect of an increase in productivity is to increase production and increase marginal cost. The impact of a change in cutoff on input is increasing in γ and, 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.[27]

4 Estimation

In this section, we extend the Feenstra-Broda-Weinstein approach using international disaggregate industry trade data to estimate σ and γ, *controlling explicitly* for firms' heterogeneous productivities.[28] As in Feenstra and Romalis (2014), our approach recognizes the importance of how a country's mass of exporting firms depends not just on that exporting country's labor-force size but also on the industry's zero-cutoff-profit productivity threshold. However, they also have endogenous quality choice; the interaction between productivity and quality is such that all firms charge the same price in equilibrium. Instead, we rely on *observed price heterogeneity* to estimate the model's parameters.

Due to our goal here of providing a novel methodological approach to estimate potentially – given availability of necessary high quality data – all three parameters (σ, γ, and θ) under our modified Melitz model framework, we omit allowing for heterogeneous (across exporter-importer pairs) bilateral export supply elasticities, as addressed recently in Soderbery (2018) and Farrokhi and Soderbery (2020); implicitly, we are considering a restricted case of Farrokhi and Soderbery (2020), as addressed in section 3 (page 21) of their paper.[29]

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

[28] Because we need estimates of all three parameters for the counterfactuals, we cannot use the method suggested by Caliendo and Parro (2015), which provides only the overall trade elasticity. In recent work, Fajgelbaum et al. (2020) uses 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 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 industry-specific demand and supply parameters.

[29] Soderbery (2018) extends estimation of the Feenstra framework in the direction of allowing *heterogeneous* (by country pair) bilateral export supply elasticities. Though still relying upon the same assumed bilateral export supply function as in the Feenstra-Broda-Weinstein models and Soderbery (2015), this paper 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) extends further the Soderbery (2018) estimation of heterogeneous (across exporter-importer pair) bilateral export supply elasticities. However, section 3 in Farrokhi and Soderbery (2020) shows that the Feenstra-Broda-Weinstein (FBW) approach is a restricted version of the more general model allowing external economies of scale and labor mobility across industries. Specifically, the FBW approach constrains the bilateral export supply elasticities to have positive slopes and assumes demand is not "convoluted by supply when using unit values." Extending our paper in these directions far exceeds the scope of our paper, which is to examine how increasing marginal costs influence the relative trade and welfare impacts of variable-trade-cost versus fixed-trade-cost liberalizations. 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). Consequently, to limit our paper's scope, we omit introducing bilateral export supply elasticities that are heterogeneous across country pairs.

4.1 Econometric Approach

We estimate a structural demand-and-supply system of equations. Although the estimation will be done separately for each industry using disaggregated trade data, we omit industry subscripts to minimize notation. Nevertheless, the empirical model developed in this section should be thought of as representing one of many industries (see footnote 11 above).

We begin by using the theoretical model to obtain equations for the (equilibrium) aggregate bilateral import-demand and bilateral export-supply functions. Our analysis shows that observed unit values, which are used in the standard approach of Feenstra (1994) and Broda and Weinstein (2006), are not appropriate measures of market prices in the presence of firm heterogeneity. However, it is still possible to extend the benchmark estimation method to account for firm heterogeneity because, as we demonstrate, the theoretically consistent market price is *proportional to observed unit values.*

4.2 Aggregate Bilateral Demand

Starting with demand equation (2), we can write aggregate nominal bilateral import demand for a variety as:

$$x_{ij}^{D}(\varphi) = p_{ij}^{c}(\varphi)c_{ij}(\varphi) = E_j P_j^{\sigma-1} p_{ij}^{c}(\varphi)^{1-\sigma}. \tag{21}$$

Using equation (21), we can write average import demand as:

$$\overline{x}_{ij}^{D} \equiv \int_{\varphi_{ij}^{*}}^{\infty} x_{ij}^{D}(\varphi)\mu_{ij}(\varphi)d\varphi = E_j P_j^{\sigma-1} \int_{\varphi_{ij}^{*}}^{\infty} p_{ij}^{c}(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi. \tag{22}$$

Hence, aggregate nominal bilateral import demand is:

$$X_{ij}^{D} = M_{ij}\overline{x}_{ij}^{D} = M_{ij}E_j P_j^{\sigma-1}\tilde{p}_{ij}^{c}, \tag{23}$$

where

$$\tilde{p}_{ij}^{c} \equiv \int_{\varphi_{ij}^{*}}^{\infty} p_{ij}^{c}(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi \tag{24}$$

is an index of firm-level prices.

Because the price index $\tilde{p}_{ij}^{c}$ is not observable in the data, we cannot use equation (23) to estimate the parameters of this model. To make progress, we express the observable average cost-insurance-freight (or cif) import unit value $\overline{p}_{ij}^{c}$ in terms of two unobservable price indexes as follows:

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

where $\tilde{p}^c_{ij}$ is defined in equation (24), and

$$\grave{p}^c_{ij} \equiv \int_{\varphi^*_{ij}}^{\infty} p^c_{ij}(\varphi)^{-\sigma} \mu_{ij}(\varphi) d\varphi \tag{26}$$

is another unobserved price index. In what follows, we use the theoretical model to obtain analytical expressions for each of the unobserved indexes, $\tilde{p}^c_{ij}$ and $\grave{p}^c_{ij}$. We then show that, by combining these two expressions, we can express aggregate nominal bilateral import demand as a function of the observable average import price $\overline{p}^c_{ij}$.

We proceed in three steps to obtain an estimable aggregate nominal bilateral import demand function that depends on the observable measure of bilateral import prices, $\overline{p}^c_{ij}$. The first step is to solve for firm-level (bilateral) prices $p^c_{ij}(\varphi)$ as functions of the productivity threshold φ^*_{ij}. Recalling $q_{ij}(\varphi)/\tau_{ij} = c_{ij}(\varphi)$ and $p^c_{ij}(\varphi) = \tau_{ij} p_{ij}(\varphi)$, we can use optimal demand equation (2) and optimal pricing rule (6) to show:

$$\frac{q_{ij}(\varphi)}{q_{ij}(\varphi^*_{ij})} = \left(\frac{\varphi}{\varphi^*_{ij}}\right)^{\sigma\left(\frac{\gamma}{\sigma+\gamma}\right)}. \tag{27}$$

Substituting into equation (27) using equation (8) for $q_{ij}(\varphi^*_{ij})$ yields:

$$q_{ij}(\varphi) = \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{\gamma}{1+\gamma}} (\varphi^*_{ij})^{-\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)} \varphi^{\sigma\left(\frac{\gamma}{\sigma+\gamma}\right)}. \tag{28}$$

Substituting equation (28) for $q_{ij}(\varphi)$ into optimal pricing rule (6) yields:

$$p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{1}{1+\gamma}} (\varphi^*_{ij})^{-\left(\frac{1}{1+\gamma}\right)\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1)} w_i \varphi^{-\frac{\gamma}{\sigma+\gamma}}. \tag{29}$$

In the second step, we compute the two unobservable average prices $\tilde{p}^c_{ij}$ and $\grave{p}^c_{ij}$ and show that the observable import unit value $\overline{p}^c_{ij}$ is proportional to the optimal price of the break-even exporter, $p^c_{ij}(\varphi^*_{ij})$. Using equation (29), optimal pricing function (6), the Pareto distribution assumption, and recalling $p^c_{ij}(\varphi) = \tau_{ij} p_{ij}(\varphi)$, we can solve for:

$$\tilde{p}^c_{ij} = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma) - \gamma(\sigma-1)}\right] p^c_{ij}(\varphi^*_{ij})^{1-\sigma}. \tag{30}$$

Using equation (26), optimal pricing function (6), and the Pareto distribution assumption, we can solve for:

$$\grave{p}^c_{ij} = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma) - \gamma\sigma}\right] p^c_{ij}(\varphi^*_{ij})^{-\sigma}. \tag{31}$$

Using these results and equation (25), we obtain:

$$\overline{p}_{ij}^{c} = \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right] p_{ij}^{c}(\varphi_{ij}^{*}), \tag{32}$$

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

The third and final step is straightforward. We can rewrite equation (32) with $p_{ij}^{c}(\varphi_{ij}^{*})$ as a function of the observable price import unit value $\overline{p}_{ij}^{c}$:

$$p_{ij}^{c}(\varphi_{ij}^{*}) = \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma+\gamma}{\theta(\sigma+\gamma)-\gamma\sigma}\right] \overline{p}_{ij}^{c}, \tag{33}$$

and substitute this last result into equation (30) to obtain:

$$\tilde{p}_{ij}^{c} = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right] \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right]^{\sigma-1} (\overline{p}_{ij}^{c})^{1-\sigma}. \tag{34}$$

We can now use this last result to express the aggregate nominal bilateral import demand, defined in equation (23), as a share of total expenditure as follows:

$$\lambda_{ij} \equiv \frac{X_{ij}}{E_j} = k_2 M_{ij} P_j^{\sigma-1} (\overline{p}_{ij}^{c})^{1-\sigma}, \tag{35}$$

where k_2 is a constant that depends only on the structural parameters σ, γ, and θ:

$$k_2 = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\,(\sigma-1)}\right] \left[\frac{\theta(\sigma+\gamma)-\gamma\sigma}{\theta(\sigma+\gamma)-\gamma(\sigma-1)}\right]^{\sigma-1}.$$

Finally, we remove the productivity threshold, φ_{ij}^{*}, and the mass of firms, M_{ij}, using equations (9) and (11), respectively, to yield:

$$\lambda_{ij} = k_3 L_i w_i^{-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} E_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} P_j^{(\sigma-1)+\theta\left(\frac{1+\gamma}{\gamma}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{-\theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} (\overline{p}_{ij}^{c})^{1-\sigma}. \tag{36}$$

where k_3 is a constant that depends only on the structural parameters σ, γ, θ, δ, and f^e:

$$k_3 = k_2 \left[\frac{\gamma}{1+\gamma}\frac{\sigma-1}{\sigma}\frac{1}{\delta\theta f^e}\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}\right)^{-\theta\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}}\left(\frac{\gamma}{\sigma+\gamma}(\sigma-1)\right)^{-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}\right].$$

This completes the derivation for the demand-side equation of the empirical model.

4.3 Aggregate Bilateral Supply

We now turn our attention to the aggregate bilateral export-supply equation. We again proceed in several steps. We begin by inverting the optimal pricing function (6) to get an analytical expression for output as a function of the price:

$$q_{ij}(\varphi) = \left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{\varphi p_{ij}(\varphi)}{w_i}\right]^{\gamma}. \tag{37}$$

We then use this result to compute the average export supply from country i to country j:

$$\overline{q}_{ij} \equiv \int_{\varphi_{ij}^*}^{\infty} q_{ij}(\varphi)\mu_{ij}(\varphi)d\varphi = \left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{1}{w_i}\right]^{\gamma}\int_{\varphi_{ij}^*}^{\infty}\left[\varphi p_{ij}(\varphi)\right]^{\gamma}\mu_{ij}(\varphi)d\varphi. \tag{38}$$

Defining aggregate bilateral export supply (in physical units) as $Q_{ij} \equiv M_{ij}\overline{q}_{ij}$, then using equation (38) yields:

$$Q_{ij} = M_{ij}\left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\frac{1}{w_i}\right]^{\gamma}\breve{p}_{ij}, \tag{39}$$

where

$$\breve{p}_{ij} \equiv \int_{\varphi_{ij}^*}^{\infty}\left[\varphi p_{ij}(\varphi)\right]^{\gamma}\mu_{ij}(\varphi)d\varphi \tag{40}$$

is yet another unobservable price index.

Because $\breve{p}_{ij}$ is not observable, we need to obtain an expression for $\breve{p}_{ij}$ as a function of an observed average price $\overline{p}_{ij}$. The first step is to solve for $\breve{p}_{ij}$ as a function of the zero-cutoff-profit firm's price, $p_{ij}(\varphi_{ij}^*)$. Substituting the optimal price equation (29) into equation (40), assuming a Pareto distribution for productivities, and solving yields:

$$\breve{p}_{ij} = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]\left[\varphi_{ij}^* p_{ij}(\varphi_{ij}^*)\right]^{\gamma}. \tag{41}$$

Substituting in the analogue for equation (33) for $p_{ij}(\varphi_{ij}^*)$, we can write:

$$\breve{p}_{ij} = \left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]\left[\frac{\theta(\sigma+\gamma)-\gamma(\sigma-1)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]^{\gamma}\left[\varphi_{ij}^*\overline{p}_{ij}\right]^{\gamma}. \tag{42}$$

Substituting equation (42) for $\breve{p}_{ij}$ in equation (39) yields:

$$Q_{ij} = k_4 M_{ij}\left(\frac{\overline{p}_{ij}\varphi_{ij}^*}{w_i}\right)^{\gamma}, \tag{43}$$

where k_4 is a constant that depends only on the structural parameters σ, γ, and θ:

$$k_4 = \left[\left(\frac{\gamma}{1+\gamma}\right)\left(\frac{\sigma-1}{\sigma}\right)\right]^{\gamma}\left[\frac{\theta(\sigma+\gamma)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]\left[\frac{\theta(\sigma+\gamma)-\gamma(\sigma-1)}{\theta(\sigma+\gamma)-\gamma\sigma}\right]^{\gamma}. \tag{44}$$

Solving for average price, we get:

$$\overline{p}_{ij} = k_4^{-\frac{1}{\gamma}}\left(\frac{Q_{ij}}{M_{ij}}\right)^{\frac{1}{\gamma}}\frac{w_i}{\varphi^*_{ij}}. \tag{45}$$

For estimation purposes, we need to make the aggregate bilateral export-supply equation (45) comparable to the aggregate bilateral import-demand equation (36). The value of aggregate bilateral exports equals the value of aggregate bilateral imports, such that $M_{ij}\overline{p}_{ij}\overline{q}_{ij} = \lambda_{ij}E_j$, or $Q_{ij} = \lambda_{ij}E_j/\overline{p}_{ij}$. Using the fact that $\overline{p}^c_{ij} = \tau_{ij}\overline{p}_{ij}$, we get $Q_{ij} = \tau_{ij}\lambda_{ij}E_j/\overline{p}^c_{ij}$. Using these results to substitute for Q_{ij} and $\overline{p}_{ij}$ in equation (45) yields:

$$\overline{p}^c_{ij} = \tau_{ij}\overline{p}_{ij} = k_4^{-\frac{1}{\gamma}}\tau_{ij}\left(\frac{\tau_{ij}\lambda_{ij}E_j}{M_{ij}\overline{p}^c_{ij}}\right)^{\frac{1}{\gamma}}\frac{w_i}{\varphi^*_{ij}}. \tag{46}$$

Solving for $\overline{p}^c_{ij}$ yields:

$$\overline{p}^c_{ij} = k_4^{-\frac{1}{1+\gamma}}\tau_{ij}\left(\frac{\lambda_{ij}E_j}{M_{ij}}\right)^{\frac{1}{1+\gamma}}\left(\frac{w_i}{\varphi^*_{ij}}\right)^{\frac{\gamma}{1+\gamma}} \tag{47}$$

where k_4 is the constant defined in (44).

Finally, we eliminate the productivity threshold, φ^*_{ij}, and the mass of firms, M_{ij}, using the derivations in equations (9) and (11), respectively, to obtain:

$$\overline{p}^c_{ij} = k_5 L_i^{-\frac{1}{1+\gamma}} w_i^{\frac{\gamma}{1+\gamma}+\left(\frac{\theta}{\gamma}-1\right)\left(\frac{\sigma}{\sigma-1}\right)} E_j^{\frac{1}{1+\gamma}+\left(1-\frac{\theta}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)} P_j^{1-\frac{\theta}{\gamma}} \tau_{ij}^{\frac{\theta}{\gamma}} f_{ij}^{\frac{1}{\sigma-1}\left(\frac{\theta}{\gamma}-1\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)} \lambda_{ij}^{\frac{1}{1+\gamma}}, \tag{48}$$

where k_5 is a constant that depends only on the parameters $\sigma, \gamma, \theta, \delta$, and f^e:

$$k_5 = k_4^{-\frac{1}{1+\gamma}}\left[\frac{\gamma}{1+\gamma}\frac{\sigma-1}{\sigma}\frac{1}{\delta\theta f^e}\right]\left[\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}\right)^{\frac{\theta-\gamma}{\gamma}\frac{\sigma}{\sigma-1}}\left(\frac{\gamma}{\sigma+\gamma}(\sigma-1)\right)^{\frac{\sigma+\gamma}{1+\gamma}(\theta-\gamma)}\right].$$

This completes the derivation for the supply-side equation of the empirical model.

4.4 Deriving the Equation for Estimation

Together, aggregate bilateral import-demand equation (36) and the aggregate bilateral export-supply equation (48) form the basis of our method to estimate the structural parameters of the model. As shown in (Online) Appendix D, using equation (36), the *double-differenced* aggregate bilateral import-demand can be expressed as follows:

$$\begin{aligned}\Delta \ln \lambda_{ijt} = \Delta \ln L_{it} - \theta \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta \ln w_{it} - \left(\frac{\theta}{\sigma-1}\right)\left(\frac{\sigma+\gamma}{\gamma}\right)\Delta \ln f_{ijt} \\ - \theta \left(\frac{1+\gamma}{\gamma}\right)\Delta \ln \tau_{ijt} - (\sigma-1)\Delta \ln \overline{p}^{c}_{ijt} + \Delta \phi_{ijt},\end{aligned} \tag{49}$$

where $\Delta\phi_{ijt}$ is the double difference of a (demand-side) residual, ϕ_{ijt}.[30] By double-differencing, we remove time-invariant and exporter-specific effects. The operator Δ denotes the *double difference* with respect to time and reference exporting country k such that for a given variable a, $\Delta \ln a_{ijt} \equiv (\ln a_{ij,t} - \ln a_{ij,t-1}) - (\ln a_{kj,t} - \ln a_{kj,t-1})$. To obtain the supply-side equation, we first double-difference equation (48), then we substitute equation (49) for $\Delta \ln \lambda_{ijt}$, and solve to obtain:

$$\begin{aligned}\Delta \ln \overline{p}^{c}_{ijt} = \left(\frac{\gamma}{\sigma+\gamma}\right)\left[1 - \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{it} - \left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt}, \\ + \left(\frac{1}{\sigma+\gamma}\right)\Delta\phi_{ijt} + \Delta\psi_{ijt},\end{aligned} \tag{50}$$

where $\Delta\psi_{ijt}$ is the double difference of a (supply-side) residual, ψ_{ijt}.[31]

Our methodology relies on the orthogonality of the double-differenced residual terms – a standard identification condition in the trade literature (e.g., Feenstra (1994); Broda and Weinstein (2006); Soderbery (2015); and Feenstra and Weinstein (2017)) – which can be expressed as $\mathbb{E}(\Delta\phi_{ijt}\Delta\psi_{ijt}) = 0$, where $\mathbb{E}$ denotes the expectation operator. As shown in Appendix D, to take advantage of this moment condition, we use equations (49) and (50) to solve for $\Delta\phi_{ij,t}$ and $\Delta\psi_{ijt}$. We then multiply the two expressions and rearrange to obtain

$$Y_{ijt} = \sum_{k=1}^{20} \beta_k Z_{ijt,k} + \xi_{ijt}, \tag{51}$$

[30] In Feenstra (1994) and the subsequent literature, the residuals are derived from within the theoretical model by including demand shocks in the CES preferences, cf., equation (1). As discussed in Appendix D, it would be straightforward to extend our model to include these shocks. We chose to not include them to simplify the presentation of the theoretical model.

[31] As explained in section 2.2 above, Feenstra (1994) and much of the subsequent literature, *assume* the existence of a positively-sloped export supply-curve comprised of a deterministic and a random component. By contrast, similar to Farrokhi and Soderbery (2020), we provide a micro-foundation for the deterministic component (described in equation (48)). As discussed in Appendix D, it would be would be straightforward to extend our model to include random shocks, such that the residuals, ψ_{ijt}, would emerge from the theoretical model. However, as was the case for the demand shock, we chose not to include them to simplify the presentation.

where

$$
\begin{array}{lll}
Y_{ijt} = \left(\Delta \ln \overline{p}^{c}_{ijt}\right)^2, & Z_{ijt,1} = \left(\Delta \ln \lambda_{ijt}\right)^2, & Z_{ijt,2} = \Delta \ln \lambda_{ijt} \Delta \ln \overline{p}^{c}_{ijt}, \\
Z_{ijt,3} = \left(\Delta \ln \tau_{ijt}\right)^2, & Z_{ijt,4} = \Delta \ln \tau_{ijt} \Delta \ln \overline{p}^{c}_{ijt}, & Z_{ijt,5} = \Delta \ln \tau_{ijt} \Delta \ln \lambda_{ijt}, \\
Z_{ijt,6} = \Delta \ln \tau_{ijt} \Delta \ln L_{it}, & Z_{ijt,7} = \Delta \ln \tau_{ijt} \Delta \ln w_{it}, & Z_{ijt,8} = \Delta \ln L_{it} \Delta \ln \overline{p}^{c}_{ijt}, \\
Z_{ijt,9} = \Delta \ln L_{it} \Delta \ln \lambda_{ijt}, & Z_{ijt,10} = \Delta \ln w_{it} \Delta \ln \overline{p}^{c}_{ijt}, & Z_{ijt,11} = \Delta \ln w_{it} \Delta \ln \lambda_{ijt}, \\
Z_{ijt,12} = \Delta \ln L_{it} \Delta \ln w_{it}, & Z_{ijt,13} = \left(\Delta \ln L_{it}\right)^2, & Z_{ijt,14} = \left(\Delta \ln w_{it}\right)^2, \\
Z_{ijt,15} = \left(\Delta \ln f_{ijt}\right)^2, & Z_{ijt,16} = \Delta \ln f_{ijt} \Delta \ln \overline{p}_{ijt}, & Z_{ijt,17} = \Delta \ln f_{ijt} \Delta \ln \lambda_{ijt}, \\
Z_{ijt,18} = \Delta \ln f_{ijt} \Delta \ln \tau_{ijt}, & Z_{ijt,19} = \Delta \ln f_{ijt} \Delta \ln w_{it}, & Z_{ijt,20} = \Delta \ln f_{ijt} \Delta \ln L_{it},
\end{array}
$$

$\xi_{ij} = \frac{(\sigma+\gamma)\Delta\phi_{ijt}\Delta\psi_{ijt}}{(1+\gamma)(\sigma-1)}$ is a residual, and the β_ks are functions of the three structural parameters, σ, γ, and θ, only. The operator Δ will denote the *double difference*.

The coefficients of equation (51)) cannot be consistently estimated because the error term is correlated with the regressors. However, as explained in Feenstra (1994) and Broda and Weinstein (2006), it is possible to obtain consistent estimates by exploiting the panel structure of the data and assuming that the parameters are constant over time for each good. Following the literature, we estimate the model using averages over time:

$$\overline{Y}_{ij} = \sum_{k=1}^{20} \beta_k \overline{Z}_{ij,k} + \overline{\xi}_{ij}, \tag{52}$$

where the over-bar indicates that the variables are averages over time (e.g., $\overline{Z}_{ij} \equiv T^{-1} \sum_t Z_{ijt}$). As demonstrated in Feenstra (1994) and Broda and Weinstein (2006), the estimates are robust to the simplest form of measurement error (with equal variance across country-pair) if a constant term is added to equation (52). Equation (52) is a generalization of the estimating equation used by Feenstra (1994) and Broda and Weinstein (2006) to estimate the elasticities σ and γ by industry. Because of the double differencing, all the variables of equation (52) have null averages. This implies that the empirical model identifies the coefficients (β_k) from the second moments of the data (i.e., variances and covariances) as explained in Rigobon (2003) and Gervais and Richard (2021).

4.5 Data

For estimation, we need information on average unit import prices, trade shares, trade costs, employment, and wages. The primary data set is the Comtrade Database collected and maintained by the United Nations. This dataset collects both f.o.b. export values that correspond to the transaction value of the goods, as well as c.i.f. import values which include the value of services performed to deliver goods to the border of the importing country.[32] The dataset also contains information on quantities imported and exported.[33] Additional information on trade barriers comes from a database compiled by Feenstra and Romalis (2014), which collects information on the rates associated with most favored nation status or any preferential status available.[34] We use this information to construct measures of trade shares, import unit values, and trade costs for each importer-exporter-industry-year observation in the sample.

We define industries as four-digit Standard International Trade Classification (SITC4) categories. Information on employment is not available at this level of detail. To 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 Gross Domestic Product (GDP).[35] Information on employment and GDP for each country-year comes for the Penn World Tables (version 9.1). Finally, we use GDP per capita as our measure of wage rates. Our sample covers the period from 1999 to 2010 (we include 1999, so that after taking time-differences we end up with years 2000-2010).

4.6 Empirical Implementation

As shown in Appendix D, the coefficients in estimating equation (52) for each industry, i.e., the β_k, depend only on the three structural parameters of the model. This implies that, *in principle*, our model allows us to identify all three structural parameters, σ, γ, and θ, from the estimated coefficients for each industry. However, to be consistent with the theory, this requires imposing a large number of constraints on the coefficients. Because some of the constraints are not linear in the coefficients, identifying the three structural parameters entails writing a non-linear optimization program that searches for the optimal values of three structural parameters over a non-trivial space, which is well outside of the scope of the current paper.

[32] Values are reported in current US dollars. We convert current dollars to 2005 dollars using information on the Consumer Price Index (CPI-U) data provided by the United States Department of Labor Bureau of Labor Statistic.

[33] 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.

[34] The dataset 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. Tariff rates are reported at the four-digit SITC level.

[35] Because we introduce increasing marginal costs, the relationship between labor and export is not linear. However, the association is still positive, such that it make sense to attribute larger values of labor to industries in large countries that export large volumes of a good. We also check the sensitivity of our estimates to this assumption using exporter fixed effects instead of direct measures of inputs.

In addition to the technical difficulties associated with such a procedure, our data may not be reliable enough to justify such an approach. The description of our empirical measures highlights the fact that there are three categories of measures to properly estimate equation (52). The first category contains two variables for which we have reliable information, the import unit values ($\overline{p}^c_{ijt}$) and the trade shares (λ_{ijt}). These correspond to the variables used in the benchmark models estimated by Feenstra (1994) and Broda and Weinstein (2006). The second category contains two variables for which we do not have direct information on, but that we can control for using proxies, industry-level employment (L_{it}) and input costs (w_{it}). The third category contains two trade-barriers variables, the fixed trade costs (f_{ijt}) and the variable trade costs (τ_{ijt}). These variables are even more problematic because, while we have reliable information on trade costs, it is unclear how to allocate freight costs between fixed and variable components. Moreover, because our identification strategy relies on double-differenced variables, the usual *time invariant* controls for fixed costs used in the literature (e.g., institutions' quality) cannot be used for estimation in our model (as they disappear in the first difference).

Given these considerations, we implement (52) as follows. As in Feenstra (1994) and Broda and Weinstein (2006), we use only the most two reliable measures, import unit values ($\overline{p}^c_{ij}$) and trade shares (λ_{ij}) to estimate the σ, and γ, and use various combinations of controls to account for the other independent variables and to control for the impact of the third structural parameter, θ. The estimating equation takes the following general form:

$$\overline{Y}_{ij} = \beta_1 \overline{Z}_{ij,1} + \beta_2 \overline{Z}_{ij,2} + Controls_{ij} + \overline{v}_{ij}, \tag{53}$$

where

$$\beta_1 = \frac{1}{(1+\gamma)(\sigma-1)}, \quad \text{and} \quad \beta_2 = \frac{\sigma-\gamma-2}{(1+\gamma)(\sigma-1)}. \tag{54}$$

Because the coefficients β_1 and β_2 are defined exactly as in Feenstra (1994) and Broda and Weinstein (2006), we use the same methodology to back out the structural parameters from the estimates. Two points are worth mentioning. First, although we do not provide empirical estimates of θ, we do control for the influence of firms' productivity heterogeneity in our estimates of σ and γ. Second, with a full complement of high-quality data and a properly specified constrained estimator, our empirical model could identify all three structural parameters of the model.

We employ three specification types to estimate equation (53) by industry. For our first specification, we use our *ad valorem* measures of trade costs (which includes both freight costs and tariff rates) to control for all the variable (τ) and the fixed trade costs (f), in light of our caveat above. This reduces the number of regressors to only the first 14 covariates (Z_k) in equation (52) – with τ representing both fixed and variable bilateral trade costs. While we cannot identify all 20 coefficients in the original specification, this simpler specification nevertheless controls for all the relevant variables and allows us to obtain estimates for the two structural parameters σ and γ associated with equations (53) and (54).

To alleviate concerns regarding our proxies for L_i and w_i and our lack of explicit controls for fixed trade costs, we also estimate two richer specifications. For our second specification, we add exporter and importer fixed effects to control for the exporter-specific and importer-specific components of trade barriers not captured by our ad valorem measures. Note that, for this specification, we remove the three exporter-specific terms Z_{12} to Z_{14} from our second specification, because they are subsumed in the exporter fixed effects. Therefore, our second specification includes only the first 11 regressors (Z_1 to Z_{11}) in equation (52), with τ again representing an index of trade costs that includes both fixed and variable barriers, along with exporter and importer fixed effects.

In our second specification just described, we removed the exporter-specific terms Z_{12} to Z_{14}, but kept the interaction terms that contain the measures of labor, L_i, and wages, w_i, i.e., Z_6 to Z_{11}. In our third and final specification, we replace the measures of L_i and w_i by the exporter fixed effects altogether. This last specification includes Z_1 to Z_5, with τ again representing an index of trade costs (thus implicitly controlling for Z_{15} to Z_{18}) , exporter fixed effects to control for Z_{12} to Z_{14}, and interaction terms *between* the exporter fixed effects and the three main variables: trade barriers (τ), price p and trade shares (λ) to control for the remaining terms Z_6 to Z_{11} and Z_{19} and Z_{20}.

4.7 Estimation Results

We estimate equation (53) separately for each four-digit SITC industry in our dataset using each of our three specifications. For purpose of comparison, we present first the results from using the "benchmark" estimation method of Feenstra (1994); this corresponds to estimating β_1 and β_2 in equation (53) *without* any additional controls. Because we obtain hundreds of estimates across industries, it would not be practical to report them all. Instead, we present only the distribution of the estimated coefficients.

As seen in Table 1, the median elasticity of substitution in consumption (σ) and the median inverse elasticity of marginal costs (γ) estimated using the benchmark Feenstra (1994) method (*without* controls for firm-heterogeneity) are 4.70 and 4.00, respectively, as shown in the first two columns of numbers. These accord with previous estimates in the literature. Turning to our estimates, we find that the median estimated elasticities of substitution for specifications 1 to 3 are 6.22, 6.43, and 6.39, respectively. The corresponding median estimated inverse elasticities of marginal costs are 5.98, 6.31, and 5.74. The results presented in Table 1 have three important implications. First, our estimates are economically quite different from the benchmark estimates. We find that using our richer specifications increases the estimated values of the elasticity of substitution and the bilateral export supply elasticity. Second, our estimates are robust to changes in specification. In addition to the median, the distributions of the estimates are quite similar across our three specifications. Third, the estimated parameters are distributed densely around the medians.

TABLE 1
DISTRIBUTION OF PARAMETER ESTIMATES

Percentile	Feenstra (1994)		Specification 1		Specification 2		Specification 3	
	σ	γ	σ	γ	σ	γ	σ	γ
1	2.64	0.57	2.55	0.66	2.59	0.71	2.54	0.62
5	3.00	1.29	3.21	1.51	3.20	1.56	3.41	1.57
10	3.29	1.68	3.77	1.98	3.82	2.00	4.03	2.23
25	3.95	2.41	4.71	3.41	4.79	3.57	4.97	3.49
50	4.70	4.00	6.22	5.98	6.43	6.31	6.39	5.74
75	5.99	6.85	9.14	11.00	9.35	12.37	8.87	10.16
90	8.63	13.02	16.11	21.83	16.30	24.22	13.63	20.61
95	11.72	20.67	22.16	38.46	23.98	43.51	18.74	35.02
99	26.75	47.27	74.25	93.91	67.43	150.69	32.97	118.74

Notes: This table presents the distributions of the estimated structural parameters of the model obtained from estimating equation (53) separately for each industry using four 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. For purpose of comparison, we report the results for the 549 industries for which we obtain estimates that conform to the restrictions of the theoretical model. About 30 percent of the industries in our sample are excluded. By comparison Broda and Weinstein (2006) exclude about 35 percent of their industries. As usual in this literature, we do not report standard errors for our estimated elasticities. As shown in equation (54), these are computed from non-linear functions of the coefficients obtained from estimating equation (53). The share of coefficients (i.e., the β_1s and β_2s) that are statistically significant at the 5 percent level ranges from about 70 to about 85 percent depending on the specification.

Although we are able to compare our specifications' estimates for the elasticities of substitution with previous empirical results using similar data, the literature provides few comparisons for our estimates of γ. For instance, 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 *in the range of our median estimates* generated here using industry-level international trade data, but controlling for firm heterogeneity. In the next section we address the importance of precise and unbiased estimates of γ for relevant policy-oriented quantitative comparative statics.

5 Numerical Analyses

Having established in the previous section strong empirical evidence using international data that firms actually face increasing marginal costs, we provide in this section two numerical analyses to illustrate the importance of allowing for IMC in welfare calculations and other counterfactuals. First, for a given set of parameters, we quantify the impact of increasing marginal costs on the fixed- and variable-trade-cost elasticities of trade. Second, we illustrate the importance of allowing for increasing marginal costs in welfare calculations. Third, 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.

5.1 Trade Elasticities

Table 2 reports the trade elasticities implied by the theoretical model using the estimated structural parameters of the model, σ and γ. As explained earlier, given the absence of sufficient quality data for all variables, we are not able at this time to estimate θ simultaneously; hence, the three values of θ used in Table 2 were chosen from representative studies. The median value, $\theta = 8.28$, came from Eaton and Kortum (2002), 6.53 was selected from Costinot et al. (2012), and 12.86 came from Eaton and Kortum (2002). We computed elasticities for 30 different scenarios, each using a different value of θ and γ as indicated in the table. The variable-trade-cost elasticity is independent of the value of the elasticity of substitution, σ. However, the fixed-trade-cost elasticity is a function of σ, γ, and θ. For Table 2, we chose our median estimate of the Armington elasticity, $\sigma = 6.33$.

Two insights are worth noting. First, as expected, for any given value of γ in the second column, the variable- and fixed-trade-cost trade elasticities are increasing in θ (going across columns), since θ is in the numerator of both trade elasticities. Second, going down any column, one observes that, as γ increases, the variable- and fixed-trade-cost trade elasticities converge to the benchmark CMC case. Comparing the trade elasticities at the median value of γ to the benchmark ($\gamma = \infty$) shows that increasing marginal costs have a quantitatively significant impact on the elasticity estimates. For example, (for $\theta = 8.28$) at the median estimated value of the bilateral export supply elasticity, $\gamma = 5.74$, the variable-trade-cost trade elasticity of 9.72 is 17 percent larger than that with constant marginal costs (8.28); this accords with intuition as the trade elasticity with IMC is scaled up by $1 + \frac{1}{\gamma}$ relative to that with CMC. As a clue to our second counterfactual later, note now that the fixed-trade-cost trade elasticity with IMC for the same values of γ and θ, 2.16, is *quadruple* that with CMC, 0.50.

In the next two sections, we provide two different quantitative counterfactual analyses, with the purpose of showing the quantitative importance of accounting for empirically-justified increasing marginal costs in the evaluation of: (1) the "gains from trade," and (2)

TABLE 2
ESTIMATED TRADE ELASTICITIES

		Variable trade elasticity θ			Fixed trade elasticity θ		
Percentile	γ	6.53	8.28	12.86	6.53	8.28	12.86
1	0.62	17.06	21.63	33.60	12.35	15.93	25.29
5	1.57	10.69	13.55	21.05	4.99	6.59	10.79
10	2.23	9.46	11.99	18.63	3.56	4.79	7.99
25	3.49	8.40	10.65	16.54	2.34	3.24	5.58
50	5.74	7.67	9.72	15.10	1.50	2.16	3.91
75	10.16	7.17	9.09	14.13	0.92	1.44	2.79
90	20.61	6.85	8.68	13.48	0.55	0.96	2.05
95	35.02	6.72	8.52	13.23	0.40	0.77	1.75
99	118.74	6.58	8.35	12.97	0.24	0.58	1.45
	∞	6.53	8.28	12.86	0.18	0.50	1.33

Notes: This table presents the distributions of the elasticities of trade estimated separately for each industry under 30 different scenarios, each with different values for the structural parameters as indicated in the table. The Pareto distribution parameter (θ) varies across columns, whereas the inverse elasticity of marginal costs (γ) varies across rows. The last row corresponds the benchmark constant marginal cost case ($\gamma = \infty$). For the fixed trade cost elasticity, we fix the elasticity of substitution at the sample median, such that $\sigma = 6.33$. For purpose of comparison, we report the results for the 477 industries for which the fixed trade cost elasticity is positive.

the trade and welfare impacts of fixed-trade-cost reductions relative to variable-trade-cost reductions, the two main elements of deep trade agreements.

5.2 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 output elasticity of marginal costs (γ) that the welfare gains from trade are reduced by *more than 10 percent* in the case of IMC relative to the case of CMC.

From equation (19), 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):

$$G_j = 1 - \lambda_{jj}^{1/\varepsilon},$$

where λ_{jj} is the domestic absorption share of GDP and $\varepsilon = \theta(1 + 1/\gamma)$.[36] Consequently, the

[36] 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 $G_j = 1 - \lambda_j^{1/\theta}$, which is identical to the measure of G_j in Costinot and Rodriguez-Clare (2014), p. 204.

only data needed for this numerical exercise is export shares. As in Feenstra (2010), we use information on nominal exports and nominal GDPs from the Penn World Tables to calculate export shares.[37] As discussed earlier, we did not estimate values for θ; we controlled for the influence of firm heterogeneity. Our selection of values for γ is based upon the *estimated* values for γ from Section 4. Moreover, 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.

Table 3 provides the results of our numerical analysis for the average welfare change; in our sample, the mean trade share is 39.3 percent, so we set $\lambda_{jj} = 60.7$. The table is organized as follows. The third, fourth, and fifth columns represent alternative values of θ, while the rows represent alternative values of γ. The last row illustrates the gains from trade (relative to autarky) under the various values of θ and the assumption of CMC ($\gamma = \infty$). As expected, as one moves across columns for any row the gains from trade decrease as the inverse index of firm heterogeneity (θ) rises. Moreover, for any given value of θ, the gains from trade increase as γ increases, moving down any column. As a representative case, for $\theta = 8.28$ and our median estimate of $\gamma = 5.74$ from Section 4, we find that the welfare gain from trade is 5.00 percent, which is a reduction of 14.5 percent from the welfare gain of 5.85 percent in the benchmark case of constant marginal costs ($\gamma = \infty$). These values are consistent with the "welfare-diminution" effect discussed in section 3.

Table 4 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; in the case of $\gamma = 5.74$, the reduction of welfare of 0.23 from 1.57 to 1.34 owing to increasing marginal costs is *15 percent.* Overall, the results presented in this section suggest that increasing marginal costs have substantive effects on welfare calculations.

[37] 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 3
WELFARE GAINS FROM TRADE, 2010

		θ		
Percentile	γ	6.53	8.28	12.86
1	0.62	2.88	2.28	1.47
5	1.57	4.56	3.61	2.34
10	2.23	5.14	4.08	2.64
25	3.49	5.77	4.58	2.97
50	5.74	6.30	5.00	3.25
75	10.16	6.72	5.34	3.47
90	20.61	7.03	5.59	3.63
95	35.02	7.16	5.69	3.70
99	118.74	7.30	5.80	3.77
100	∞	7.36	5.85	3.81

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 39.3, so we set $\lambda_{jj}=60.7$. We compute gains from trade under 30 different scenarios, each with different values for the structural parameters of the model as indicated in the table. The values for the Pareto distribution parameter (θ) vary across columns, whereas the values for the inverse elasticity of marginal costs (γ) vary across rows. The last row presents the benchmark constant marginal cost case, which corresponds to $\gamma=\infty$. We note that, the variable-trade-cost trade elasticity (ε_τ) is independent of the elasticity of substitution σ.

5.3 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.

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:

TABLE 4

WELFARE GAINS FROM TRADE FOR SELECTED COUNTRIES, 2010

			γ			
Name	GDPPC	Export Share	3.49	5.74	10.16	∞
Guinea	1,677	30.34	3.34	3.65	3.90	4.27
Mali	1,736	22.84	2.40	2.63	2.81	3.08
Nepal	1,807	9.58	0.94	1.03	1.10	1.21
Kyrgyzstan	2,863	51.55	6.58	7.18	7.66	8.38
Republic of Moldova	3,737	39.23	4.57	4.99	5.33	5.84
Congo	4,709	65.81	9.58	10.45	11.13	12.16
Guatemala	6,293	25.81	2.76	3.02	3.23	3.54
China	9,423	26.27	2.82	3.09	3.29	3.61
Thailand	13,109	66.49	9.75	10.64	11.33	12.37
Gabon	13,151	57.66	7.75	8.46	9.02	9.86
Brazil	13,623	10.74	1.06	1.16	1.24	1.36
Malaysia	20,192	86.93	17.39	18.88	20.05	21.79
Israel	30,538	35.02	3.97	4.34	4.63	5.07
Bahamas	31,413	34.95	3.96	4.33	4.62	5.06
Italy	35,936	25.19	2.69	2.94	3.14	3.44
Germany	40,481	42.25	5.02	5.49	5.86	6.42
Saudi Arabia	41,482	49.57	6.22	6.80	7.25	7.94
United States	49,907	12.32	1.23	1.34	1.43	1.57
Norway	57,900	39.73	4.64	5.07	5.41	5.93
Bermuda	62,290	49.69	6.25	6.82	7.28	7.96

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. We set $\theta = 8.28$ and let the inverse elasticity of marginal costs (γ) varies across columns as indicated in the table. The last column presents the benchmark constant marginal cost case, which corresponds to $\gamma = \infty$. We selected 20 countries that cover the range of incomes per capita and geographical regions.

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. such 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. While recent empirical studies noted in the introduction provide evidence of the non-trivial impact of lowering such barriers on trade flows, few studies have yet provided estimates of their impact on the extensive margin of trade.

While empirical studies are now starting to flourish given the World Bank's new data base on the "Content of Preferential Trade Agreements," 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 three 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) finds 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 44.8 billion (US) dollars. Hence, the welfare-equivalent reduction in fixed trade costs would be 36 percent, to match the 5 percent reduction in variable trade costs (or a ratio of approximately 7). 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 two other papers have considered CGE analyses using a Melitz framework, Balisteri et al. (2011) and Dixon et al. (2016). 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 those from Canada to Japan; moreover, the fixed trade costs of intra-national Japanese trade is 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). The only other 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. Hence, for comparison, our model is most similar to Zhai (2008).

In our second counterfactual, we are interested in measuring fixed-trade-cost changes, $\hat{f}_{ij}$, that are equivalent in welfare to removing a given (*ad valorem*) variable trade cost, $\hat{\tau}_{ij}$. In our model, as seen in equation (15), we can write:

$$\phi_{ij} = \tau_{ij}^{-\varepsilon_\tau} f_{ij}^{-\varepsilon_f}, \tag{55}$$

such that for a given value of $\hat{\tau}_{ij}$, the equivalent fixed-trade-cost change is $\hat{f}_{ij} = \hat{\tau}_{ij}^{\frac{\varepsilon_\tau}{\varepsilon_f}}$. This gives the change in fixed trade costs that is equivalent to a change in variable trade costs in terms of its impact on trade flows, $\hat{\lambda}_{ij}$, and welfare.

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

$$\frac{\varepsilon_\tau}{\varepsilon_f} = \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)} - 1}. \tag{56}$$

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 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 σ and γ using specification 3 from section 4, $\sigma = 6.53$ and $\gamma = 5.74$, as well as the preferred (median) value of $\theta = 8.28$

from Eaton and Kortum (2002). Substituting in these values yields:

$$CMC : \frac{\varepsilon_\tau}{\varepsilon_f} = \frac{\theta}{\frac{\theta}{\sigma-1} - 1} = 16.65 \tag{57}$$

$$IMC : \frac{\varepsilon_\tau}{\varepsilon_f} = \frac{\theta\left(\frac{1+\gamma}{\gamma}\right)}{\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)} - 1} = 4.41 \tag{58}$$

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 eliminating a country's – or an average of countries' – MFN tariff rates. From the WTO, the average MFN tariff rates applied by the G20 countries is about 5 percent. This implies that the equivalent fixed costs changes are:

$$CMC : \hat{f} = (1.05)^{16.65} = 2.25 \tag{59}$$

$$IMC : \hat{f} = (1.05)^{4.41} = 1.24. \tag{60}$$

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

Table 5 reports the distribution of the ratio of welfare-equivalent fixed-trade-cost changes across industries for a 5 percent reduction in the tariff rate, $0.05/1.05 \approxeq 0.05$.[38] To discipline the quantitative exercise, we use the mean tariff applied in each industry and consider a complete removal of the tariff barriers as our shocks, $\hat{\tau}$. We compute the welfare-equivalent change for two separate cases, the benchmark CMC case of $\gamma \to \infty$ and the IMC case. The table shows that, for the median industry, *ad valorem* tariffs are about 3 percent. The equivalent fixed-trade-costs change under CMC is 47 percent, whereas under IMC it is only 13 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, which corresponds to eliminating a 9 percent tariff, the equivalent fixed-trade-cost change is a reduction of *1,434 percent*. But under IMC, it is a much more reasonable 56 percent.

Table 6 reports the distribution of the ratio of welfare-equivalent fixed-trade-cost changes for selected countries. We first compute the average tariff rate imposed by each country for each industry separately. We then compute an *import-weighted-average* tariff rate for each country. We set the shock to eliminating the average tariff rate. As in Table 4, we compute the 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.

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

TABLE 5
EQUIVALENT FIXED COSTS CHANGE

		CMC		IMC	
Percentile	$\hat{\tau}$	$\varepsilon_\tau/\varepsilon_f$	$\hat{f}$	$\varepsilon_\tau/\varepsilon_f$	$\hat{f}$
1	1.00	0.93	1.00	0.65	1.00
5	1.01	2.82	1.04	1.40	1.02
10	1.01	3.56	1.09	1.91	1.03
25	1.02	6.12	1.20	2.64	1.07
50	1.03	10.45	1.47	3.65	1.13
75	1.05	23.12	2.72	5.29	1.28
90	1.09	52.67	15.34	7.43	1.56
95	1.12	134.50	126.26	9.86	2.13
99	1.23	711.54	2.41e+10	36.75	9.31

Notes: This table present the distribution of the average industry-level *ad valorem* trade barriers ($\hat{\tau}$) and the corresponding equivalent fixed cost changes ($\hat{f}$). We set $\theta = 8.28$ and let the elasticity of substitution (σ) and the inverse elasticity of marginal costs (γ) varies across industries. The last row presents the benchmark constant marginal cost case, which corresponds to $\gamma = \infty$. The equivalent fixed costs changes are obtained from $\hat{f}_{ij} = \hat{\tau}_{ij}^{\varepsilon_\tau/\varepsilon_f}$. We keep the 477 industries in the sample for which the fixed trade costs elasticities are positive.

We conclude by addressing the result for the United States. For the United States, the MFN tariff rate is only about 3 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 and γ (along with an external estimate of θ) allow us to construct an estimate of $\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}$. Given the framework above, we find that – under CMC – the welfare-equivalent reduction (to a removal of the 3 percent tariff rate) in fixed trade costs is 88 percent (7.46/8.46). By contrast, under IMC the welfare-equivalent reduction in fixed trade costs is *only 15 percent*. The latter makes deep trade agreements much more attractive to pursue, with a 15 percent reduction well below the reductions of 25 percent used in earlier analyses of TTIP in Berden et al. (2010).

6 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

TABLE 6
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	21.18	5.47	5.57	1.56
Mali	1,736	1.10	37.62	32.55	6.14	1.77
Nepal	1,807	1.14	369.65	4.17e+20	6.32	2.25
Kyrgyzstan	2,863	1.01	48.42	1.79	7.11	1.09
Republic of Moldova	3,737	1.03	99.00	16.88	6.34	1.20
Congo	4,709	1.17	22.04	29.81	4.28	1.93
Guatemala	6,293	1.05	75.18	34.46	6.14	1.34
China	9,423	1.09	28.11	11.33	6.12	1.70
Thailand	13,109	1.09	30.27	12.78	5.09	1.54
Gabon	13,151	1.17	22.60	33.62	4.35	1.97
Brazil	13,623	1.11	62.80	620.23	5.71	1.79
Malaysia	20,192	1.08	61.59	142.58	5.43	1.55
Israel	30,538	1.06	73.65	60.13	6.49	1.44
Bahamas	31,413	1.28	94.90	1.18e+10	4.74	3.19
Italy	35,936	1.01	55.27	2.02	6.37	1.08
Germany	40,481	1.01	60.92	2.46	5.27	1.08
Saudi Arabia	41,482	1.10	227.70	2.98e+09	6.25	1.82
United States	49,907	1.03	63.53	8.46	4.70	1.17
Norway	57,900	1.01	48.47	1.89	5.23	1.07
Bermuda	62,290	1.18	42.42	1124.48	3.99	1.94

Notes: This table present the distribution of the average country-level *ad valorem* trade barriers ($\hat{\tau}$) and the corresponding equivalent fixed cost changes ($\hat{f}$) for selected countries for year 2010. We set $\theta = 8.28$ and let the inverse elasticity of marginal costs (γ) varies across rows as indicated in the table. The equivalent fixed costs changes are obtained from $\hat{f}_{ij} = \hat{\tau}_{ij}^{\varepsilon_\tau/\varepsilon_f}$. We keep the 477 industries in the sample for which the fixed trade costs elasticities are positive. To the extent possible, we choose the same countries as in Table 3.1 of Feenstra (2010) to facilitate comparison.

of output, implying that the welfare gain from trade are reduced as diminishing marginal returns interact with the Pareto shape parameter to lower the average productivity gains from trade liberalizations. Adapting Table 3.1 in Head and Mayer (2014), Table 7 contrasts the *ad valorem* variable-trade-cost intensive-margin elasticities, *ad valorem* variable-trade-cost trade elasticities, export-fixed-cost trade elasticities, and welfare effects from the large class of models addressed in Arkolakis et al. (2012) with those from this paper.

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 ranging from about 5.7 to 6.4 across various specifications – *far below* ∞, which is assumed in the benchmark models in the trade literature assuming CMC.

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 3 percent reduction in *ad valorem* variable trade costs was 47 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*.

We offer three 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, availability of higher quality data for industrial employment, input costs, variable trade costs, and fixed trade costs would enable credible estimation of σ, γ, *and* θ simultaneously; this would provide a significant enhancement of the Feenstra-Broda-Weinstein econometric approach. Third, 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.

TABLE 7

Elasticities and Welfare Measures by Model

Model	Intensive-margin elasticity	Trade elasticity (ε_τ)	Fixed-trade-cost elasticity (ε_f)	Welfare-change 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 export costs (Eaton-Kortum, 2002)	n.a.	θ	n.a.	$\hat{\lambda}_{jj}^{-\frac{1}{\theta}}$
Heterogeneity with fixed export costs and Pareto (Chaney, 2008)	$\sigma - 1$	θ	$\frac{\theta}{\sigma-1} - 1$	$\hat{\lambda}_{jj}^{-\frac{1}{\theta}}$
Heterogeneity with fixed export costs, Pareto, and IMC (Current paper)	$\frac{1+\gamma}{\sigma+\gamma}(\sigma-1)$	$\theta\left(\frac{1+\gamma}{\gamma}\right)$	$\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)} - 1$	$\hat{\lambda}_{jj}^{-\frac{1}{\theta\left(\frac{1+\gamma}{\gamma}\right)}}$

Notes: This table reports the (positively-defined) *ad valorem* variable-trade-cost intensive-margin elasticity, the *ad valorem* variable-trade-cost trade elasticity, the export-fixed-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 section B of the Supplemental Appendix (NIFP) for explanation. n.a. denotes not applicable.

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.

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.

Baier, S. and J. Bergstrand (2001). The growth of world trade: Tariffs, transport costs, and income similarity. *Journal of International Economics 53*(1), 1–27.

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.

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.

Caliendo, L. and F. Parro (2015). Estimates of the trade and welfare effects of NAFTA. *Review of Economic Studies 82*, 1–44.

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

Costinot, A., D. Donaldson, and I. Komunjar (2012). What goods do countries trade? A quantitative exploration of Ricardo's ideas. *Review of Economic Studies 79*(2).

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).

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.

Fontagne, L., N. Rocha, M. Ruta, and G. Santoni (2020). A general equilibrium assessment of the economic impact of deep trade agreements. *Working Paper*.

Gervais, A. and P. Richard (2021). Measuring international prices with correlated (unobserved) demand and supply shifts. Unpublished working paper.

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*.

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

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

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.

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*.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics 85*(4), 777–792.

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.

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.

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"

July 30, 2021

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}:

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

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:

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

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

$$\max_{p_{ij}} \pi_{ij}(\varphi) = E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} p_{ij}(\varphi)^{1-\sigma} - w_i f_{ij} - \frac{w_i}{\varphi} \left[E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} p_{ij}(\varphi)^{-\sigma} \right]^{\frac{1+\gamma}{\gamma}}.$$

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:

$$\frac{\partial \pi_{ij}}{\partial p_{ij}} = (1-\sigma)E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} p_{ij}(\varphi)^{-\sigma} + \sigma \left(\frac{1+\gamma}{\gamma} \right) \frac{w_i}{\varphi} \left(E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} \right)^{\frac{1+\gamma}{\gamma}} p_{ij}(\varphi)^{-\sigma\left(\frac{1+\gamma}{\gamma}\right)-1} = 0,$$

Simplifying the equation above yields:

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

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:

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

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

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

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

$$p_{ij}^c(\varphi) \equiv \tau_{ij}p_{ij}(\varphi) = \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi}q_{ij}(\varphi)^{\frac{1}{\gamma}}. \tag{A.5}$$

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

$$\lim_{\gamma\to\infty} p_{ij}(\varphi)^c = \left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi}.$$

A.2 Firm Profits

From equation (A.1), we have:

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

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 $\gamma \to \infty$:

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

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:

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right) \frac{r_{ij}(\varphi_{ij}^*)}{\sigma} = w_i f_{ij}. \tag{A.7}$$

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

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right) \left(\frac{1}{\sigma}\right) \left(\frac{1+\gamma}{\gamma}\right) \left(\frac{\sigma}{\sigma-1}\right) \frac{w_i q_{ij}(\varphi_{ij}^*)^{\frac{1+\gamma}{\gamma}}}{\varphi_{ij}^*} = w_i f_{ij}, \tag{A.8}$$

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:

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

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

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

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

$$\left(\frac{\sigma+\gamma}{1+\gamma}\right) \frac{E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} p_{ij}(\varphi_{ij}^*)^{1-\sigma}}{\sigma} = w_i f_{ij}. \tag{A.11}$$

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

$$
\begin{aligned}
w_i f_{ij} &= \left(\frac{\sigma+\gamma}{1+\gamma}\right) \frac{E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma} \left\{ \left(\frac{1+\gamma}{\gamma}\right) \left(\frac{\sigma}{\sigma-1}\right) \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{1}{1+\gamma}} w_i (\varphi_{ij}^*)^{\frac{-\gamma}{1+\gamma}} \right\}^{1-\sigma}}{\sigma} \\
\Rightarrow (\varphi_{ij}^*)^{(\sigma-1)\left(\frac{\gamma}{1+\gamma}\right)} &= \left(\frac{1+\gamma}{\sigma+\gamma}\right) \left(\frac{\sigma w_i f_{ij}}{E_j P_j^{\sigma-1} \tau_{ij}^{1-\sigma}}\right) \left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right) w_i\right]^{\sigma-1} \left[\left(\frac{\gamma}{\sigma+\gamma}\right)(\sigma-1) f_{ij}\right]^{\frac{\sigma-1}{1+\gamma}} \\
\Rightarrow \varphi_{ij}^* &= \left[\frac{\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1} w_i\right)^{\sigma}}{E_j P_j^{\sigma-1}}\right]^{\frac{1}{\frac{\gamma}{1+\gamma}(\sigma-1)}} \left[\frac{\gamma}{\sigma+\gamma}(\sigma-1) f_{ij}\right]^{\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}} \tau_{ij}^{\frac{1+\gamma}{\gamma}}. \qquad \text{(A.12)}
\end{aligned}
$$

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

$$
\begin{aligned}
\lim_{\gamma\to\infty} \varphi_{ij}^* &= \left[\left(\frac{\sigma}{\sigma-1}\right)^{\sigma} (\sigma-1) \frac{f_{ij} w_i^{\sigma}}{E_j P_j^{\sigma-1}}\right]^{\frac{1}{\sigma-1}} \tau_{ij} = \frac{\sigma^{1+\frac{1}{\sigma-1}} w_i^{1+\frac{1}{\sigma-1}} f_{ij}^{\frac{1}{\sigma-1}} \tau_{ij}}{(\sigma-1) E_i^{\frac{1}{\sigma-1}} P_j} \\
&= \left(\frac{\sigma}{\sigma-1}\right) \frac{w_i \tau_{ij}}{P_j} \left(\frac{\sigma w_i f_{ij}}{E_j}\right)^{\frac{1}{\sigma-1}}.
\end{aligned}
$$

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 $\gamma = \infty$ 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:

$$\pi_{ij}(\varphi^*_{ij}) = 0 \quad \Leftrightarrow \quad r_{ij}(\varphi^*_{ij}) = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\sigma w_i f_{ij}. \tag{A.15}$$

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

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

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:

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

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 equilibrium 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:

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

This result is analogous to the zero-cutoff-profit condition in Melitz (2003), with $\overline{\pi}_i$ 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:

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

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:

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

A.5 Masses of Firms

Consumers have no taste for leisure, so the supply of labor is fixed at L_i. There are two sources of demand for labor, labor for production and labor for entry costs. Therefore, the labor-market-clearing condition is given by:

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

where M_i^e is the mass of firms attempting to enter the industry in country i and 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 equilibrium distribution of firm productivity.

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

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

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

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

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

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

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:

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

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

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

We can now solve for $\sum_{j=1}^{N} M_{ij} f_{ij}$. 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, 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 finite 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:

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

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

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

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

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

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

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

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

which implies that:

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

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:

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

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

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

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:

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

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:

$$
\begin{aligned}
p_{ij}^c(\varphi) &= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi}q_{ij}(\varphi)^{\frac{1}{\gamma}} \\
&= \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi}\left[E_jP_j^{\sigma-1}p_{ij}^c(\varphi)^{-\sigma}\right]^{\frac{1}{\gamma}} \\
&= \left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi}\right]^{\frac{\gamma}{\sigma+\gamma}}E_j^{\frac{1}{\sigma+\gamma}}P_j^{\frac{\sigma-1}{\sigma+\gamma}} \qquad \text{(A.42)}
\end{aligned}
$$

Substituting this result into the definition of the price index

$$
P_j = \left[\int_{\nu\in\Omega_j} p_j^c(\nu)^{1-\sigma}d\nu\right]^{\frac{1}{1-\sigma}}, \qquad \text{(A.43)}
$$

and rearranging, we obtain:

$$
\begin{aligned}
P_j^{1-\sigma} &= \int_{\nu\in\Omega_j} p_j^c(\nu)^{1-\sigma}d\nu = \sum_i M_{ij}\int_{\varphi_{ij}^*}^{\infty} p_{ij}^c(\varphi)^{1-\sigma}\mu_{ij}(\varphi)d\varphi \\
&= \sum_i M_{ij}\int_{\varphi_{ij}^*}^{\infty}\left\{\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi}\right]^{\frac{\gamma}{\sigma+\gamma}}E_j^{\frac{1}{\sigma+\gamma}}P_j^{\frac{\sigma-1}{\sigma+\gamma}}\right\}^{1-\sigma}\mu_{ij}(\varphi)d\varphi \\
&= \sum_i M_{ij}\left\{\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)w_i\tau_{ij}\right]^{\frac{\gamma}{\sigma+\gamma}}E_j^{\frac{1}{\sigma+\gamma}}P_j^{\frac{\sigma-1}{\sigma+\gamma}}\right\}^{1-\sigma}\int_{\varphi_{ij}^*}^{\infty}\varphi^{(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\mu_{ij}(\varphi)d\varphi \\
&= \sum_i M_{ij}\left\{\left[\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\frac{w_i\tau_{ij}}{\varphi_{ij}^*}\right]^{\frac{\gamma}{\sigma+\gamma}}E_j^{\frac{1}{\sigma+\gamma}}P_j^{\frac{\sigma-1}{\sigma+\gamma}}\right\}^{1-\sigma}\left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right] \\
&= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\sum_i M_{ij}[p_{ij}^c(\varphi_{ij}^*)]^{1-\sigma} \\
&= \left[\frac{\theta}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\sum_i M_{ij}\tau_{ij}^{1-\sigma}[p_{ij}(\varphi_{ij}^*)]^{1-\sigma}. \qquad \text{(A.44)}
\end{aligned}
$$

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:

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

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

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

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

$$
\lambda_{ij} \equiv \frac{X_{ij}}{E_j} = \frac{w_i L_i f_{ij}(\varphi_{ij}^*)^{-\theta}}{\sum_{k=1}^{N} w_k L_k f_{kj}(\varphi_{kj}^*)^{-\theta}}. \qquad (A.47)
$$

Adapting equation (A.12), we know:

$$
\varphi_{kj}^* = \left[\frac{\left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}w_k\right)^{\sigma}}{E_j P_j^{\sigma-1}}\right]^{\frac{1}{\frac{\gamma}{1+\gamma}(\sigma-1)}}\left[\frac{\gamma}{\sigma+\gamma}(\sigma-1)f_{kj}\right]^{\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}\tau_{kj}^{\frac{1+\gamma}{\gamma}}
$$

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}):

$$\lambda_{ij} = \frac{w_i L_i f_{ij} \left[w_i^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tau_{ij}^{\frac{1+\gamma}{\gamma}} f_{ij}^{\left(\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right)} \right]^{-\theta}}{\sum_{k=1}^{N} w_k L_k f_{kj} \left[w_k^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tau_{kj}^{\frac{1+\gamma}{\gamma}} f_{kj}^{\left(\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right)} \right]^{-\theta}}$$

$$= \frac{L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}}{\sum_{k=1}^{N} L_k w_k^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tau_{kj}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{kj}^{1-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}}. \tag{A.48}$$

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):

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

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:

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

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_i^e, must be equal to the cost of entry such that:

$$V_i^e = [1 - G(\varphi_{ii}^*)]\frac{\overline{\pi}_i}{\delta} = w_i f^e. \tag{A.53}$$

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 the wage rate, 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 $G(\varphi) = 1 - \varphi^{-\theta}$. As shown in section A.4 of the Supplemental Appendix (NIFP), 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:

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

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 (14), 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 with equation (A.12) for the productivity threshold in the solution for bilateral trade flows in equation (A.45) yields:

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

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

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

where

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

Rearranging this last result to solve for $R_i \tilde{\Pi}_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)}$ and substituting into (A.57), we get:

$$X_{ij} = R_i \tilde{\Pi}_i^{\theta\left(\frac{1+\gamma}{\gamma}\right)} \left(E_j P_j^{\sigma-1}\right)^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta}{\sigma-1}\right)} \phi_{ij} \tag{A.60}$$

where

$$\phi_{ij} = \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\frac{\theta}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}} \tag{A.61}$$

denotes the trade barriers component of bilateral trade flows. We can define $\tilde{\Phi}_j$ such that:

$$E_j \tilde{\Phi}_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)} = \left(E_j P_j^{\sigma-1}\right)^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta}{\sigma-1}\right)}. \tag{A.62}$$

Substituting with this last result into (A.60) yields:

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

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:

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

Finally, by definition of expenditure and equation (A.60) it follows that:

$$E_j = \sum_{i=1}^{N} X_{ij} = L_i \left(E_j P_j^{\sigma-1}\right)^{\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\theta}{\sigma-1}\right)} \sum_{i=1}^{N} R_i \tilde{\Pi}_i^{\varepsilon_\tau} \phi_{ij}. \tag{A.65}$$

This result implies that:

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

If we relabel using $\tilde{\Pi}_i^{-\varepsilon_\tau} = \Pi_i$ and $\tilde{\Phi}_j^{-\varepsilon_\tau} = \Phi_j$, then 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:

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

Finally, 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:

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

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:

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

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:

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

which is the result in the paper.

A.14 Welfare

In the model, welfare 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 the purchasing. 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:

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

Substituting demand function (A.2) into the equation above yields:

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

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

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

Substituting in the equation above with the mass of firms from equation (A.41) and then the cutoff productivity term from equation (A.12), we can solve for:

$$\begin{aligned}
P_j^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} &= \left[\frac{1}{\theta-(\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)}\right]\left(\frac{\gamma}{\sigma+\gamma}\right)\left(\frac{\sigma-1}{\delta f_e}\right) A^{-\theta} E_j^{\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)-1} \\
&\quad \times \sum_{i=1}^{N} L_i w_i^{1-\theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)} \tau_{ij}^{-\theta\left(\frac{1+\gamma}{\gamma}\right)} f_{ij}^{1-\theta\left(\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}\right)},
\end{aligned} \tag{A.87}$$

where

$$A = \left(\frac{1+\gamma}{\gamma}\frac{\sigma}{\sigma-1}\right)^{\frac{\sigma}{\frac{\gamma}{1+\gamma}(\sigma-1)}}\left[\frac{\gamma}{\sigma+\gamma}(\sigma-1)\right]^{\frac{1}{\frac{\gamma}{\sigma+\gamma}(\sigma-1)}}$$

is a constant that depends only on parameters σ and γ. Having defined the first component of welfare, we turn to the second component: wage rates. From equation (A.48), we have:

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

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

where $\tau_{jj} = 1$. Substituting with our results in equations (A.87) and (A.88) into our definition of welfare in equation (A.84), and simplifying, yields:

$$\begin{aligned} W_j &= \left\{ \left[\frac{1}{\theta - (\sigma-1)\left(\frac{\gamma}{\sigma+\gamma}\right)} \right] \left(\frac{\gamma}{\sigma+\gamma}\right) \left(\frac{\sigma-1}{\delta f^e}\right) A^{-\theta} \right\}^{\frac{1}{\theta}\left(\frac{\gamma}{1+\gamma}\right)} L_j^{\frac{1}{\sigma-1}} f_{jj}^{\left(\frac{1}{1+\gamma}\right)\left(\frac{\gamma}{\theta}-\frac{\sigma+\gamma}{\sigma-1}\right)} \lambda_{jj}^{-\frac{1}{\theta}\left(\frac{\gamma}{1+\gamma}\right)} \\ &= B L_j^{\frac{1}{\sigma-1}} f_{jj}^{\left(\frac{1}{1+\gamma}\right)\left(\frac{\gamma}{\theta}-\frac{\sigma+\gamma}{\sigma-1}\right)} \lambda_{jj}^{-\frac{1}{\theta}\left(\frac{\gamma}{1+\gamma}\right)} \end{aligned}$$

where

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

is a constant that depends only on parameters $\sigma, \gamma, \theta, \delta$, and f^e (A is defined above, just after equation (A.87)). Hence, for any *foreign* shock (i.e., holding constant L_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 $\lambda_{jj}'/\lambda_{jj}$.

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*.[39]

To make this point, Feenstra (2010) derives a transformation curve between masses of varieties for sale to different markets, M_{ij}, and show 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

[39]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 the A_{ij}s are demand-shift parameters that depend only on parameters of the model. 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 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

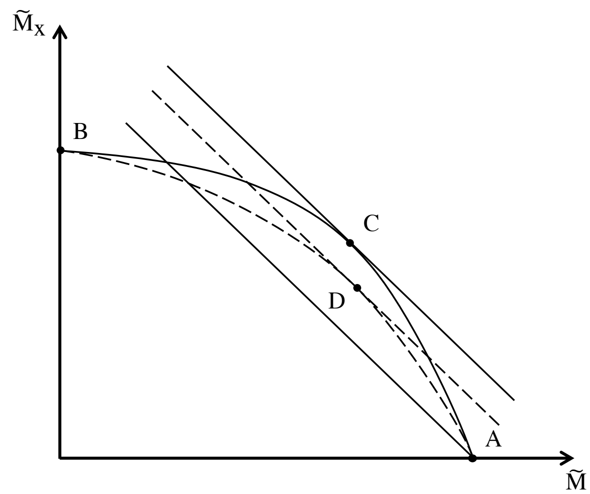

Figure A.1: CET Frontier with Increasing Marginal Costs and Constant Marginal Costs

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).[40] Note that k_1 is a constant that depends only parameters $\sigma, \gamma,, \theta$, and k_0:

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

[40]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_iY_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, $X_{ij} = b_iY_j$, 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), $X_{ij} = Y_iY_j/Y_W$, 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).[41] 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

[41] 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.

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:

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

where q_d denotes domestic sales and q_x denotes sales to a foreign market. 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 each market. 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 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 such equation 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:

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

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

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

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:

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

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:

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

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

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

denotes the equilibrium distribution of exporting firms' productivities. 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:

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

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

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:

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

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

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 follows that:

$$\lim_{J\to\infty} b = 0, \quad \text{and} \quad \lim_{J\to\infty} c = 1. \tag{C.30}$$

Together, these results imply that:

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

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

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

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

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.

D Appendix D

In this appendix, we provide details on the derivations required to go from the aggregate bilateral import-demand equation (36) and the aggregate bilateral export-supply equation (48) to estimating equation (51) in the paper, which we can use to estimate the structural parameters of the model.

We proceed in two main steps. In the first step, we remove the time-invariant and importer-specific effects by double-differencing. Taking logs of aggregate bilateral import-demand equation (36) yields:

$$\begin{aligned}\ln \lambda_{ijt} = \ln k_3 + \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\ln E_{jt} + \left[\sigma - 1 + \theta\left(\frac{1+\gamma}{\gamma}\right)\right]\ln P_{jt} \\ + \ln L_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\ln w_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\ln\ \tau_{ijt} \\ - \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\ln f_{ijt} - (\sigma-1)\ln \overline{p}^c_{ijt} + \phi_{ijt}.\end{aligned} \tag{D.1}$$

where ϕ_{ijt} is a demand-side residual that we added to the model.

In the literature beginning with Feenstra (1994), the demand-side residuals are introduced in the model by including (Armington) demand parameters in the CES preferences. It would be straightforward to include them in our theoretical model, but it would complicate the analytical expressions without having any impact on our main theoretical findings. Suppose that we include demand parameter a_{ij} in the CES as follows

$$U_j = \left[\int_{\nu\in\Omega_j} (a^{\sigma}_{ij} c_j(\nu))^{\frac{\sigma-1}{\sigma}} d\nu\right]^{\frac{\sigma}{\sigma-1}},$$

then the optimal demand for each variety is given by:

$$c_j(\nu) = E_j P_j^{\sigma-1}\left(\frac{p^c_j(\nu)}{a_{ij}}\right)^{-\sigma}.$$

Because a_{ij} is a constant, it would simply carry through all the computation and eventually show up in the aggregate bilateral import-demand equation.[42] Because the sole purpose of including the Armington parameters is to generate residual terms for the empirical model, we chose not to include them in the theoretical model.

We remove time-invariant effects from equation (D.1) by taking a first-difference over

[42] In fact, we can keep the algebra intact and replace the current definition of price, $p^c_j(\nu)$ with $p^{c\prime}_j(\nu) = p^c_j(\nu)/a_{ij}$ until the vey end. Then unpack the two components of $p^{c\prime}_j(\nu)$ when deriving the estimating equation.

time to obtain:

$$\begin{aligned}
\Delta_t \ln \lambda_{ijt} &\equiv \ln \lambda_{ijt} - \ln \lambda_{ij,t-1} \\
&= \theta \left(\frac{1+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\Delta_t \ln E_{jt} + \left[\sigma - 1 + \theta\left(\frac{1+\gamma}{\gamma}\right)\right]\Delta_t \ln P_{jt} \\
&+ \Delta_t \ln L_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta_t \ln w_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta_t \ln \tau_{ijt} \\
&\quad - \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\Delta_t \ln f_{ijt} - (\sigma-1)\Delta_t \ln \overline{p}^c_{ijt} + \Delta_t \phi_{ijt}.
\end{aligned}$$

where Δ_t refers to the time-differencing. Next, we remove *importer-specific* effects by taking a first difference with respect to a reference country k to obtain:

$$\begin{aligned}
\Delta \ln \lambda_{ijt} &\equiv \Delta_t \ln \ \lambda_{ijt} - \Delta_t \ln \lambda_{kjt} = (\ln \lambda_{ijt} - \ln \lambda_{ij,t-1}) - (\ln \lambda_{kjt} - \ln \lambda_{kj,t-1}) \\
&= \Delta \ln L_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta \ln w_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta \ln \tau_{ijt} \\
&\quad - \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} - (\sigma-1)\Delta \ln \overline{p}^c_{ijt} + \Delta \phi_{ijt},
\end{aligned} \tag{D.2}$$

where Δ refers to the *double-differencing.*

We repeat the same process for the aggregate bilateral export-supply equation (48). First, we take logs:

$$\begin{aligned}
\ln \overline{p}^c_{ijt} = \ln k_5 &+ \left[\frac{1}{1+\gamma} + \left(1 - \frac{\theta}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\right]\ln E_{jt} + \left(1 - \frac{\theta}{\gamma}\right)\ln P_{jt} \\
&- \left(\frac{1}{1+\gamma}\right)\ln L_{it} + \left[\frac{\gamma}{1+\gamma} + \left(\frac{\theta}{\gamma} - 1\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\ln w_{it} + \left(\frac{\theta}{\gamma}\right)\ln \tau_{ijt} \\
&+ \left[\left(\frac{1}{\sigma-1}\right)\left(\frac{\theta}{\gamma} - 1\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)\right]\ln f_{ijt} + \left(\frac{1}{1+\gamma}\right)\ln \lambda_{ijt} + \mu_{ijt},
\end{aligned} \tag{D.3}$$

where μ_{ijt} is a supply-side residual that we added to the model.

In Feenstra (1994) and the subsequent literature, these residuals come from random productivity shocks in the supply curve. It is important to note that the supply curve in that literature is not obtained from technologies and firm behavior; it is simply *assumed.* By contrast, we develop a general equilibrium model from which we derive the (deterministic component of the) aggregate bilateral supply curve. We could also include random productivity shocks to our theoretical model to generate supply-side residuals.[43] However, as was

[43] For example, there could be a random component to firm-level productivity. If firms maximize their expected value, their output (or labor demand) decisions would only depend on the deterministic component, φ, as described in the model. However, the actual production costs (or output) would not necessarily be as predicted because of the random productivity component.

the case for the demand-side residuals, we determined that the added complexity was a distraction given the main objectives of the paper.

From our last result, equation (D.3), we take a first-difference over time and rewrite the coefficients for $\ln w_{it}$ and $\ln f_{ijt}$ to obtain:

$$\begin{aligned}
\Delta_t \ln \overline{p}^c_{ijt} = &\left[\frac{1}{1+\gamma} + \left(1 - \frac{\theta}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\right]\Delta_t \ln\ E_{jt} + \left(1 - \frac{\theta}{\gamma}\right)\Delta_t \ln P_{jt} \\
&- \left(\frac{1}{1+\gamma}\right)\Delta_t \ln L_{it} + \left(\frac{1}{1+\gamma}\right)\left[\gamma + (\theta-\gamma)\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta_t \ln w_{it} \\
&+ \left(\frac{\theta}{\gamma}\right)\Delta_t \ln \tau_{ijt} + \left[\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)\left(\frac{1}{\sigma-1}\right)\right]\Delta_t \ln f_{ijt} \\
&+ \left(\frac{1}{1+\gamma}\right)\Delta_t \ln \lambda_{ijt} + \Delta_t \mu_{ijt},
\end{aligned}$$

and, finally, a first difference with respect to a reference country k to obtain:

$$\begin{aligned}
\Delta \ln \overline{p}^c_{ijt} = -\left(\frac{1}{1+\gamma}\right)\Delta \ln L_{it} &+ \left(\frac{1}{1+\gamma}\right)\left[\gamma + (\theta-\gamma)\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{it} \\
&+ \left(\frac{\theta}{\gamma}\right)\Delta \ln \tau_{ijt} + \left[\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)\left(\frac{1}{\sigma-1}\right)\right]\Delta \ln f_{ijt} \\
&+ \left(\frac{1}{1+\gamma}\right)\Delta \ln \lambda_{ijt} + \Delta \mu_{ijt}.
\end{aligned}$$

We can eliminate $\ln \lambda_{ijt}$ from this last result using the definition in equation (D.2). This yields:

$$\begin{aligned}
\Delta \ln \overline{p}^c_{ijt} = -\left(\frac{1}{1+\gamma}\right)\Delta \ln L_{it} &+ \left(\frac{1}{1+\gamma}\right)\left[\gamma + (\theta-\gamma)\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{it} \\
&+ \left(\frac{\theta}{\gamma}\right)\Delta \ln \tau_{ijt} + \left[\left(\frac{\theta-\gamma}{\gamma}\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)\left(\frac{1}{\sigma-1}\right)\right]\Delta \ln f_{ijt} \\
&+ \left(\frac{1}{1+\gamma}\right)\Bigg[\Delta \ln L_{it} - \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta \ln w_{it} \\
&- \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta \ln \tau_{ijt} - \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} \\
&- (\sigma-1)\Delta \ln \overline{p}^c_{ijt} + \Delta \phi_{ijt}\Bigg] + \Delta \mu_{ijt}.
\end{aligned}$$

Finally, after some algebra, we obtain:

$$\Delta \ln \overline{p}^c_{ijt} = \left(\frac{\gamma}{\sigma+\gamma}\right)\left[1-\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{ijt} - \left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} + \left(\frac{1}{\sigma+\gamma}\right)\Delta\phi_{ijt} + \Delta\psi_{ijt}, \tag{D.4}$$

where $\Delta\psi_{ijt} \equiv \left(\frac{1+\gamma}{\sigma+\gamma}\right)\Delta\mu_{ijt}$.

In the second step, we exploit the moment condition: $\mathbb{E}\left(\Delta\phi_{ijt}\Delta\psi_{ijt}\right) = 0$. To do so, we first need to define the doubled-differenced residuals, $\Delta\phi_{ijt}$ and $\Delta\psi_{ijt}$. From equation (D.2), we obtain:

$$\Delta\phi_{ijt} = \Delta \ln \lambda_{ijt} - \Delta \ln L_{it} + \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta \ln w_{it} + \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta \ln \tau_{ijt} + \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} + (\sigma-1)\Delta \ln \overline{p}^c_{ijt} \tag{D.5}$$

and from equation (D.4), we get:

$$\Delta\psi_{ijt} = \Delta \ln \overline{p}^c_{ijt} - \left(\frac{\gamma}{\sigma+\gamma}\right)\left[1-\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{it} + \left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} - \left(\frac{1}{\sigma+\gamma}\right)\Delta\phi_{ijt}.$$

We use equation (D.5) to substitute for $\Delta\phi_{ijt}$ in the previous result and obtain:

$$\begin{aligned}\Delta\psi_{ijt} = \Delta \ln \overline{p}^c_{ijt} &- \left(\frac{\gamma}{\sigma+\gamma}\right)\left[1-\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{it} + \left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} \\ &- \left(\frac{1}{\sigma+\gamma}\right)\Bigg[\Delta \ln \lambda_{ijt} - \Delta \ln L_{it} + \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\Delta \ln w_{it} \\ &+ \theta\left(\frac{1+\gamma}{\gamma}\right)\Delta \ln \tau_{ijt} + \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right)\Delta \ln f_{ijt} + (\sigma-1)\Delta \ln \overline{p}^c_{ijt}\Bigg].\end{aligned}$$

After collecting terms, this expression reduces to:

$$\begin{aligned}\Delta\psi_{ijt} = \left(\frac{1+\gamma}{\sigma+\gamma}\right)\Delta \ln \overline{p}^c_{ijt} &- \left(\frac{1}{\sigma+\gamma}\right)\left[\gamma + (\theta-\gamma)\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right]\Delta \ln w_{it} \\ &+ \left(\frac{1}{\sigma+\gamma}\right)\Delta \ln L_{it} - \theta\left(\frac{1}{\sigma+\gamma}\right)\left(\frac{1+\gamma}{\gamma}\right)\Delta \ln \tau_{ijt} \\ &+ \left(\frac{1}{\sigma-1}\right)\left(\frac{\theta-\gamma}{\gamma}\right)\Delta \ln f_{ijt} - \left(\frac{1}{\sigma+\gamma}\right)\Delta \ln \lambda_{ijt}.\end{aligned} \tag{D.6}$$

For convenience, we simplify the notation in equations (D.5) and (D.6), and re-order terms, to obtain, respectively:

$$\Delta\phi_{ijt} = a_1\Delta\ln\lambda_{ijt} + a_2\Delta\ln L_{it} + a_3\Delta\ln w_{it} + a_4\Delta\ln f_{ijt} + a_5\Delta\ln\overline{p}^c_{ijt} + a_6\Delta\ln\tau_{ijt} \quad \text{(D.7)}$$

$$\Delta\psi_{ijt} = b_1\Delta\ln\overline{p}^c_{ijt} + b_2\Delta\ln w_{it} + b_3\Delta\ln L_{it} + b_4\Delta\ln\tau_{ijt} + b_5\Delta\ln f_{ijt} + b_6\Delta\ln\lambda_{ijt}, \quad \text{(D.8)}$$

where the coefficients are defined as follows:

$$\begin{array}{ll}
a_1 = 1 & b_1 = \frac{1+\gamma}{\sigma+\gamma} \\
a_2 = -1 & b_2 = -\left(\frac{1}{\sigma+\gamma}\right)\left[\gamma + (\theta-\gamma)\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right)\right] \\
a_3 = \theta\left(\frac{1+\gamma}{\gamma}\right)\left(\frac{\sigma}{\sigma-1}\right) & b_3 = \frac{1}{\sigma+\gamma} \\
a_4 = \theta\left(\frac{\sigma+\gamma}{\gamma}\right)\left(\frac{1}{\sigma-1}\right) & b_4 = -\theta\left(\frac{1}{\sigma+\gamma}\right)\left(\frac{1+\gamma}{\gamma}\right) \\
a_5 = \sigma - 1 & b_5 = \left(\frac{1}{\sigma-1}\right)\left(\frac{\theta-\gamma}{\gamma}\right) \\
a_6 = \theta\left(\frac{1+\gamma}{\gamma}\right) & b_6 = \frac{-1}{\sigma+\gamma}
\end{array}$$

We can now obtain an expression for the product of the double-differenced exogenous demand and supply shocks, $\Delta\phi_{ijt}\Delta\psi_{ijt}$. From equations (D.7) and (D.8), we obtain (after some simplifications):

$$\begin{aligned}
\Delta\phi_{ijt}\Delta\psi_{ijt} = {} & (a_1b_1 + a_5b_6)\Delta\ln\lambda_{ijt}\Delta\ln\overline{p}^c_{ijt} + (a_1b_2 + a_3b_6)\Delta\ln\lambda_{ijt}\Delta\ln w_{it} \\
& + (a_1b_3 + a_2b_6)\Delta\ln\lambda_{ijt}\Delta\ln L_{it} + (a_1b_4 + a_6b_6)\Delta\ln\lambda_{ijt}\Delta\ln\tau_{ijt} \\
& + (a_1b_5 + a_4b_6)\Delta\ln\lambda_{ijt}\Delta\ln f_{it} + (a_1b_6)(\Delta\ln\lambda_{ijt})^2 \\
& + (a_2b_1 + a_5b_3)\Delta\ln L_{it}\Delta\ln\overline{p}^c_{ijt} + (a_2b_2 + a_3b_3)\Delta\ln L_{it}\Delta\ln w_{it} \\
& + (a_2b_5 + a_4b_5)\Delta\ln L_{it}\Delta\ln f_{ijt} + (a_2b_4 + a_6b_3)\Delta\ln L_{it}\Delta\ln\tau_{ijt} \\
& + (a_2b_3)(\Delta\ln L_{it})^2 + (a_3b_1 + a_5b_2)\Delta\ln w_{it}\Delta\ln\overline{p}^c_{ijt} \\
& + (a_3b_2)(\Delta\ln w_{it})^2 + (a_3b_4 + a_6b_2)\Delta\ln w_{it}\Delta\ln\tau_{ijt} \\
& + (a_3b_5 + a_4b_2)\Delta\ln w_{it}\Delta\ln f_{ijt} + (a_4b_1 + a_5b_5)\Delta\ln f_{ijt}\Delta\ln\overline{p}^c_{ijt} \\
& + (a_4b_4 + a_6b_5)\Delta\ln f_{ijt}\Delta\ln\tau_{ijt} + (a_4b_5)(\Delta\ln f_{ijt})^2 \\
& + (a_5b_1)(\Delta\ln\overline{p}^c_{ijt})^2 + (a_5b_4 + a_6b_1)\Delta\ln\overline{p}^c_{ijt}\Delta\ln\tau_{ijt} \\
& + (a_6b_4)(\Delta\ln\tau_{ijt})^2.
\end{aligned}$$

Applying the moment condition $\mathbb{E}\left(\Delta\phi_{ijt}\Delta\psi_{ijt}\right) = 0$ to this last result yields the equation for estimation in the main text, equation (51)

Regarding equation (53) in the paper, we note that the coefficients on $\overline{Z}_{ij,1}(= \Delta(\ln\lambda_{ijt})^2)$

and $\overline{Z}_{ij,2}(= \Delta \ln \lambda_{ijt} \Delta \ln \overline{p}^{c}_{ijt})$ are equal, respectively, to:

$$\begin{aligned}\beta_1 \equiv -\frac{a_1 b_6}{a_5 b_1} &= -\left[\frac{(1)\left(\frac{-1}{\sigma+\gamma}\right)}{(\sigma-1)\left(\frac{1+\gamma}{\sigma+\gamma}\right)}\right] = \left(\frac{1}{\sigma+\gamma}\right)\left(\frac{1}{\sigma-1}\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right) \\ &= \frac{1}{(\sigma-1)(1+\gamma)} \end{aligned} \tag{D.9}$$

and

$$\begin{aligned}\beta_2 \equiv -\left(\frac{a_1 b_1 + a_5 b_6}{a_5 b_1}\right) &= -\left[\frac{(1)\left(\frac{1+\gamma}{\sigma+\gamma}\right) + (\sigma-1)\left(\frac{1}{\sigma+\gamma}\right)}{(\sigma-1)\left(\frac{1+\gamma}{\sigma+\gamma}\right)}\right] \\ &= -\left[\left(\frac{1+\gamma-\sigma+1}{\sigma+\gamma}\right)\left(\frac{1}{\sigma-1}\right)\left(\frac{\sigma+\gamma}{1+\gamma}\right)\right] = -\left[\frac{2+\gamma-\sigma}{(\sigma-1)(1+\gamma)}\right] \\ &= \frac{\sigma-\gamma-2}{(\sigma-1)(1+\gamma)}\end{aligned}$$

which are exactly the same definitions as in Feenstra (1994).