THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 1056

November 2008

Structural Estimation and Solution of International Trade Models with Heterogeneous Firms

by

Edward J. Balistreri, Russell H. Hillberry & Thomas F. Rutherford

Department of Economics
The University of Melbourne
Melbourne   Victoria   3010
Australia.

Structural Estimation and Solution of International Trade Models with Heterogeneous Firms[*]

Edward J. Balistreri[†]
Colorado School of Mines

Russell H. Hillberry
University of Melbourne

Thomas F. Rutherford
The Swiss Federal Institute of Technology (ETH-Zürich)

November 2008

Abstract

We present an empirical implementation of a general-equilibrium model of international trade with heterogeneous manufacturing firms. The theory underlying our model is consistent with Melitz (2003). A nonlinear structural estimation procedure identifies a set of core parameters and unobserved firm-level trade frictions that best fit the geographic pattern of trade. Once the parameters are identified, we utilize a decomposition technique for computing general-equilibrium counterfactuals. We first assess the economic effects of reductions in measured tariffs. Taking the simple-average welfare change across regions the Melitz structure indicates welfare gains from liberalization that are nearly four times larger than in a standard trade policy simulation. Furthermore, when we compare the economic impact of tariff reductions with reductions in estimated fixed trade costs we find that policy measures affecting the fixed costs are of greater importance than tariff barriers. (JEL C68,F12)

1 Introduction

In canonical models of international trade, trade-policy changes induce factors of production to move between industries. Recent empirical and theoretical developments suggest that *within*-industry factor movements may also be an important channel through which policy affects economic outcomes. More productive plants are more likely to engage in exporting, so trade liberalization allows a reallocation of factors from less- to more-productive plants. Melitz (2003) derives a model that makes qualitative predictions about these phenomena. We evaluate the quantitative significance of these insights within the Melitz framework.[1]

Recent papers in the geography-of-trade literature use structural assumptions and the bilateral trade pattern to make inferences about the size of unobserved trade costs. Anderson and van Wincoop (2003) develop a structural estimation technique that can be used to infer trade costs.[2] Balistreri and Hillberry (2007) develop an extensive-form estimation strategy for the same model, and argue that it can be extended to most general equilibrium models of trade.[3] Balistreri and Hillberry (2008) link structural estimation techniques to established methods for calibrating general equilibrium models. We adapt these methods to calibrate a multi-sector, multi-region general equilibrium model in which the manufacturing sector has a Melitz-style market structure.

A key technical challenge in the development of this method is an algorithm for efficient solution of a multi-dimensional model with heterogeneity in plant level productivity. This is computationally difficult because the model requires joint solution of the trade equilibrium and a host of entry conditions identifying marginal firms. To solve this problem, we decom-

---

[1]Alvarez and Lucas (2007) evaluate the consequences of tariff reductions in the Eaton and Kortum (2003) model. Our exercise is motivated along similar lines, but focuses on the Melitz (2003) model.

[2]Eaton and Kortum (2003) and Behrens et al. (2008) engage in similar exercises, using structural assumptions to infer trade costs from the bilateral trade pattern.

[3]Su and Judd (2008) suggest a similar, extensive-form, technique as a general strategy for structural estimation of economic models.

pose the general equilibrium into an industry-specific module which determines the industrial organization and a general equilibrium module which evaluates relative prices, comparative advantage and the terms of trade. The full general equilibrium solution is achieved through a convergent iterative procedure in which the exchange equilibrium is solved conditional on the industrial organization.

Our structural estimation procedure uses the equations defined by the numeric model as a series of side constraints on the econometric objective. The model is underidentified, but assumptions about a few key structural parameters allow the econometric procedure to complete the identification of both the production technology and average bilateral trade costs. Leaning heavily on our structural model, we attribute differences between observed and fitted bilateral trade in manufactured goods to unobserved fixed costs of trade. This interpretation allows for an exact calibration of the model, without any role for idiosyncratic preferences in manufacturing trade.[4]

The key structural parameters of the model are the distance elasticity of ad valorem trade costs and the parameter defining the shape of the Pareto distribution of firm level productivities. We estimate these under three different sets of identifying assumptions. Our preferred specification ties down the distance elasticity of trade costs, using unexplained variation in the trade pattern to fit the implied shape of the productivity distribution. Our estimate of this parameter is largely consistent with estimates from the confidential plant-level data.

With our general equilibrium system fully parameterized, we proceed to a quantitative assessment of the effect of trade policy changes. Using a standard Armington structure as the benchmark, we consider a 50 percent reduction in manufacturing tariffs. As expected, endogenous productivity changes and growth in the number of imported varieties lead to larger welfare changes in the Melitz model than in the Armington baseline. Taking the simple-average welfare change across regions, the Melitz structure indicates welfare gains from lib-

---

[4]Hillberry et al. (2005) highlight and critique the role of idiosyncratic preference biases in determining the trade equilibrium in Computable General Equilibrium models.

eralization that are nearly four times larger than those from the baseline model. We also consider reductions in the inferred bilateral fixed costs, and find that the welfare gains from these changes are substantially larger.[5] Joint reductions of tariffs and the inferred fixed costs of trade generate even larger welfare gains. When we reduce both tariffs and fixed border costs by 50% the average welfare gain is nearly 30 times larger than in the case of tariff cuts in the Armington structure.

In section 2 we provide a review of the relevant literature. Section 3 provides a brief review of the Melitz (2003) theory as it relates to our application. Section 4 details a practical method for numerically solving the heterogeneous-firms model. In Section 5 we outline the nonlinear estimation procedure, and estimate the structural parameters that allow the model to best fit the data. These parameters are used to calibrate an operational general equilibrium model, which we employ to conduct counterfactual analysis. The results of this analysis appear in section 6. In section 7 we discuss implications of our work and directions for further research.

2   Literature Review

Two broad areas of the empirical trade literature motivate heterogeneous-firms models like those proposed in Melitz (2003) or Bernard et al. (2007).

First, an extensive literature documents heterogeneity across establishments in productivity, export behavior, and responses to trade shocks. Important findings from this literature include a) there is wide variation in productivity levels among coexisting plants;[6] b) only a small fraction of establishments engage in exporting, and exporters tend to be larger

---

[5]As in Anderson and van Wincoop (2003), Eaton and Kortum (2003), and Behrens et al. (2008), these exercises should be seen as illustrative calculations about the value of a significant move toward full integration, rather than as an assessment of particular trade policy changes.

[6]This is a robust feature of the data, as discussed in Bartelsman and Doms (2000).

and more productive than non-exporters;[7] c) there is considerable heterogeneity *among* exporters in the number of markets served per firm;[8] d) within-industry reallocation of market share from less-productive to more-productive establishments is an important component of aggregate productivity growth;[9] and e) productivity growth via shifting market shares (including the exit of the lowest productivity plants) is an important channel through which trade cost reductions induce aggregate productivity growth.[10]

Second, a recent literature has focussed attention on the *extensive* margin of trade. Several authors have linked variation in aggregate trade flows to variation in the number of a) firms trading, b) commodities traded, and c) trading partners.[11] Of particular interest in this literature is explaining trade growth via this *extensive* margin.[12]

In a critique of the performance of applied general equilibrium models commonly used in trade policy analysis, Kehoe (2005) argues that the models typically fail along two dimensions: they do not allow trade policy to affect aggregate productivity, and they do not allow trade policy to induce trade growth along the extensive margin. While some policy estimates include *ad hoc* productivity adjustments, these attempts do not typically specify the mechanism by which trade policy is meant to induce productivity growth.[13] Analysis of the extensive margin of trade remains outside the scope of most policy-oriented trade models.

While the empirical literature has demonstrated the relevance of within-industry productivity heterogeneity and trade growth via the extensive margin, what has been lacking until recently is a sound theoretical structure that formalizes the insights from the empirical literature. Melitz-type models with firm heterogeneity and fixed trade costs offer a useful

---

[7]See, for example, Bernard and Jensen (1999), Roberts and Tybout (1997), and Bernard et al. (2003).

[8]Eaton et al. (2004) document this using French data.

[9]See Foster et al. (2001) and Aw et al. (2001), among others. An important component of the contribution of shifting market share is the exit of less-productive establishments.

[10]See, for example, Bernard et al. (2006) and Pavcnik (2002).

[11]Eaton et al. (2004) and Hillberry and Hummels (2008) show that variation in the number of firms serving a market explains variation in exports to that market. Hummels and Klenow (2005) and Broda and Weinstein (2006) identify the extensive margin in terms of the role of added commodities/trading partners.

[12]See Kehoe and Ruhl (2002) and Evenett and Venables (2002).

[13]See Anderson et al. (2005) for an example.

framework for addressing Kehoe's critique. Trade policy changes affect industry productivity by shifting market share away from low-productivity non-exporters, and toward high-productivity exporters. The model also allows for trade growth along the extensive margin, and provides a mechanism by which such trade growth can be linked to policy changes.

3 Theory

Consumers have Cobb-Douglas utility over commodity bundles which are defined as constant-elasticity-of-substitution (CES) aggregates of differentiated products. Firms pay a fixed cost of entry. Entrants receive a random productivity draw. Firms with sufficiently low productivity draws exit, and the remaining firms produce with a technology exhibiting increasing returns to scale. Trade costs include *ad valorem* iceberg costs, revenue-generating tariffs, and a fixed cost of entering each market. Firms with higher levels of productivity will be able to profitably serve more markets. The model is simplified by isolating the characteristics and behavior of the average firm participating in each bilateral market. Melitz (2003) develops the critical links between the average and marginal firms, and shows how average firm characteristics relate to consumer utility.

3.1 Demand

Consumers in region $s \in R$ are assumed to have Cobb-Douglas preferences over composites from different sectors, A_{ks}, where the sector is indexed by k and α_k is the expenditure share;

$$U_s = \prod_k (A_{ks})^{\alpha_k}. \tag{1}$$

We drop the industry index at this point and isolate the Dixit-Stiglitz composite of manufactured goods consumed in region s,

$$A_s = \left[\sum_r \int_{\omega_{rs} \in \Omega_r} q_s(\omega_{rs})^\rho \, d\omega_{rs} \right]^{\frac{1}{\rho}}, \tag{2}$$

where ω_{rs} indexes the differentiated products sourced from region $r \in R$ (and Ω_r is the set of goods produced in r). Substitution across the products is indicated by $\rho = 1 - 1/\sigma$, where σ is the constant elasticity of substitution. The dual price index, P_s, is given by

$$P_s = \left[\sum_r \int_{\omega_{rs} \in \Omega_r} p_s(\omega_{rs})^{1-\sigma} \, d\omega_{rs} \right]^{\frac{1}{1-\sigma}}. \tag{3}$$

Defining this in terms of the average variety's price, $\tilde{p}_{rs}$, we have

$$P_s = \left[\sum_r N_{rs} \tilde{p}_{rs}^{1-\sigma} \right]^{1/(1-\sigma)} \tag{4}$$

where N_{rs} is the number of varieties shipped from r to s. Melitz (2003) obtains this simplification by noting that $\tilde{p}_{rs}$ is the price set by a small firm with the CES weighted average productivity $\tilde{\varphi}_{rs}$.[14] Demand for the average variety to be shipped from r to s at a gross of trade and tax price of $\tilde{p}_{rs}$ is

$$\tilde{q}_{rs} = \frac{\alpha E_s}{P_s} \left(\frac{P_s}{\tilde{p}_{rs}} \right)^\sigma \tag{5}$$

where E_s is the value of total expenditures in region s.[15]

---

[14]The weighted average productivity is given by

$$\tilde{\varphi}_{rs} = \left[\int_0^\infty \varphi_{rs}^{\sigma-1} \mu_{rs}(\varphi_{rs}) d\varphi_{rs} \right]^{\frac{1}{\sigma-1}},$$

where $\mu_{rs}(\varphi_{rs})$ is the distribution of productivities of each of the N_{rs} firms.

[15]One problem we face in reconciling the empirical model with the established theory is the discrepancy between gross expenditures and value added, because of intermediate inputs. To simplify we assume that intermediate inputs are purchases of the aggregate consumption commodity. Gross expenditures, E_s, less the value of intermediate inputs (to all industries) equals regional income.

3.2 Firm-level environment

We assume a single composite input price, c_r, associated with all fixed or marginal costs of manufacturing in region r. In application, we adopt an upstream Cobb-Douglas technology for generating the composite input. This is represented by a cost function of the form

$$c_r = (P_r^E)^{\beta_r^E} \prod_j (w_{jr})^{\beta_{jr}}, \tag{6}$$

where the w_{jr} are the prices of the factor inputs and P_r^E is the price of the composite intermediate input. Constant returns in the technology for forming the composite input indicates that the sum of the share parameters, the β, equals one.

Operating firms in a given market use the composite input to cover both fixed-operating and marginal costs, but firms also face an *entry* cost. The entry cost entitles the firm to a productivity draw. If the productivity draw is sufficiently high the firm will operate profitably. Let f_r^e indicate the entry cost (in composite-input units), and let M_r denote the number of entered firms in region r. Then each of the M_r firms incur the nominal entry payment $c_r f_r^e$, although this payment is spread across time (as there is a nonzero probability that the firm will survive beyond the current period).

Now consider the input technology for a firm from region r that finds it profitable to sell into market s. Let f_{rs} indicate the recurring fixed cost of operating on the r–s link, and let φ represent the firm-specific measure of productivity. A firm supplying q units to s uses

$$f_{rs} + \frac{q}{\varphi}$$

units of inputs. Higher productivity (higher φ) indicates lower marginal cost.

Once a firm incurs the *entry* cost, f_r^e, it is sunk and has no bearing on the firm's decision to operate in a given bilateral market. The profits earned by infra-marginal firms in the bilateral

markets do, however, give firms the incentive to incur the entry cost in the first place. There is no restriction on the markets that can be served by a given member of M_r. If a firm's productivity is high enough such that it is profitable to operate in multiple markets it can replicate itself, maintaining the same marginal cost but incurring the fixed operating cost, f_{rs}, for each of the s markets it serves.

The small firms, facing constant-elasticity demand for their differentiated products, follow the usual optimal markup rule. Let τ_{rs} indicate the *iceberg* transport-cost factor, and let t_{rs} indicate the tariff. Focusing on the average firm (with productivity draw $\tilde{\varphi}_{rs}$) shipping from r to s, optimal (gross) pricing is given by

$$\tilde{p}_{rs} = \frac{c_r \tau_{rs}(1 + t_{rs})}{\rho \tilde{\varphi}_{rs}}. \tag{7}$$

3.3 Operation, Entry, and the Average Firm

We assume that each of the M_r firms choosing to incur the entry cost receive their firm-specific productivity draw φ from a Pareto distribution with probability density

$$g(\varphi) = \frac{a}{\varphi} \left(\frac{b}{\varphi} \right)^a; \tag{8}$$

and cumulative distribution

$$G(\varphi) = 1 - \left(\frac{b}{\varphi} \right)^a, \tag{9}$$

where a is the shape parameter and b is the minimum productivity.

Considering the fixed cost of operating, f_{rs}, on the r–s link there will be some level of productivity, φ_{rs}^*, at which operating profits are zero. All firms drawing a φ above φ_{rs}^* will serve the s market, and firms drawing a φ below φ_{rs}^* will not. A firm drawing φ_{rs}^* is the marginal firm from r supplying region s. This leads us to the fundamental condition which determines the number of operating firms in a given market, N_{rs}. Let $r(\varphi) = p(\varphi)q(\varphi)$ indicate

the gross-of-tariff firm revenues as a function of the draw φ. Zero profits for the marginal firm requires

$$c_r f_{rs} = \frac{r(\varphi_{rs}^*)}{\sigma(1+t_{rs})}. \tag{10}$$

We would like, however, to define this condition in terms of the average operating firm rather than the marginal firm.

Following Melitz (2003) we define $\tilde{\varphi}$ as the productivity of a firm pricing at $\tilde{p}$, such that our simplification in equation (4) is consistent. The probability that a firm will operate is $1 - G(\varphi^*)$, so we find the CES weighted average productivity,

$$\tilde{\varphi}_{rs} = \left[\frac{1}{1 - G(\varphi_{rs}^*)} \int_{\varphi_{rs}^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right]^{\frac{1}{\sigma-1}}. \tag{11}$$

Using the Pareto distribution this becomes

$$\tilde{\varphi}_{rs} = \left[\frac{a}{a+1-\sigma} \right]^{\frac{1}{\sigma-1}} \varphi_{rs}^*. \tag{12}$$

Again, following Melitz (2003) optimal firm pricing and the input technology $(f_{rs} + q/\varphi)$ we establish the relationship between the revenues of firms with different productivity draws:

$$\frac{r(\varphi_1)}{r(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2} \right)^{\sigma-1}. \tag{13}$$

Using (12) and (13) to simplify (10) we derive the zero cutoff profit condition in terms of average-firm revenues and the parameters:

$$c_r f_{rs} + \pi_{rs}^c = \frac{\tilde{p}_{rs} \tilde{q}_{rs}}{(1+t_{rs})} \frac{(a+1-\sigma)}{a\sigma}. \tag{14}$$

The variable π_{rs}^c is introduced to track any extra profits that are generated when each of the M_r firms operate in a market. We term these profits *capacity rents*. The value of π_{rs}^c must

be zero in a steady-state, but if M_r is sticky a policy shock might lead to $N_{rs} = M_r$ indicating rents.[16]

Next we turn to the entry condition which determines the mass of firms, M_r. Firm entry requires a one-time payment of f_r^e, and entered firms face a probability δ in each future period of a *bad* shock, which forces exit. In a steady-state equilibrium δM_r firms are lost in a given period so total entry payments in that period must be $c_r \delta M_r f_r^e$. From an individual firm's perspective the *annualized* flow of entry payments is $c_r \delta f_r^e$.

Assuming risk neutrality and no discounting, firms enter to the point that expected operating profits equal the entry payment. A firm from r operating in market s can expect to earn the average profit in that market:

$$\tilde{\pi}_{rs} = \frac{\tilde{p}_{rs}\tilde{q}_{rs}}{\sigma(1+t_{rs})} - c_r f_{rs}. \tag{15}$$

Using the zero cutoff profit condition to substitute out the operating fixed cost this reduces to

$$\tilde{\pi}_{rs} = \frac{\tilde{p}_{rs}\tilde{q}_{rs}}{(1+t_{rs})} \frac{(\sigma-1)}{a\sigma}. \tag{16}$$

The probability that a firm in r will service the s market is simply given by the ratio of N_{rs}/M_r.[17] Setting the firm-level entry-payment flow equal to the expected profits from each potential market gives us the free entry condition

$$c_r \delta f_r^e = \sum_s \frac{N_{rs}}{M_r} \frac{\tilde{p}_{rs}\tilde{q}_{rs}}{(1+t_{rs})} \frac{(\sigma-1)}{a\sigma} \tag{17}$$

which determines the mass of firms, M_r.

---

[16]The value of π_{rs}^c is determined by the variational-inequality presented in the next section, equation (20). We are only concerned with steady-state equilibria (where $\pi_{rs}^c = 0$) in this study, but we found that the computational model performed better with the extended condition, which avoids numeric moves where $N_{rs} > M_r$.

[17]In Melitz (2003) the probability that a firm will operate, which equals the fraction of operating firms in equilibrium, is presented as $1 - G(\varphi^*)$.

Finally we can recover the marginal productivity as a function of the fraction of operating firms, $N_{rs}/M_r = 1 - G(\varphi^*)$. Applying the Pareto distribution and inverting we have

$$\varphi^*_{rs} = \frac{b}{\left(\frac{N_{rs}}{M_r}\right)^{1/a}}. \tag{18}$$

In the following section we formalize a computational model based on the outlined heterogenous-firms theory.

4 Solution Method

We represent the policy analysis model on the basis of two related equilibrium problems. The first is a partial equilibrium (PE) model which captures the heterogeneous-firms industrial organization in manufacturing and the associated impact on productivity and prices. The PE model takes aggregate income levels and supply schedules as given. The second module is a *constant-returns* general equilibrium (GE) model of global trade in all products. The GE model takes industrial structure as given and determines relative prices, comparative advantage and terms of trade. We iterate between these two models in policy simulations, letting the first module determine industrial structure and the second module establish regional incomes and relative costs. Industrial structure (numbers of firms operating within and across borders) are passed from the first module to the second whereas the structure of aggregate demand (income levels and supply prices) are passed back from the GE module to the PE module. Once the models are mutually consistent we have a solution to the multiregion general equilibrium with heterogeneous manufacturing firms. The four steps involved in the solution algorithm are depicted in Figure 1.

In most policy modeling exercises, applied economists prefer to work with integrated equilibrium models formulated as systems of equations in which prices and quantities are

11

Figure 1: A Decomposition Algorithm

```
                    ┌──────────────────────────────┐
                    │ Step 2: Recalibrate Armington│
              ┌────▶│ demand functions in the GE   │─────┐
              │     │ model to reflect market      │     │
              │     │ structure.                   │     │
              │     └──────────────────────────────┘     │
              │                                          ▼
┌─────────────────────────┐              ┌─────────────────────────┐
│ Step 1: Solve one IRTS  │              │ Step 3: Solve the       │
│ spatial price equilibrium│             │ integrated CRTS general │
│ model *for each commodity*│            │ equilibrium model       │
└─────────────────────────┘              └─────────────────────────┘
              ▲                                          │
              │     ┌──────────────────────────────┐     │
              │     │ Step 4: Recalibrate resource │     │
              └─────│ supply schedules and demand  │◀────┘
                    │ functions in the PE model.   │
                    └──────────────────────────────┘
```
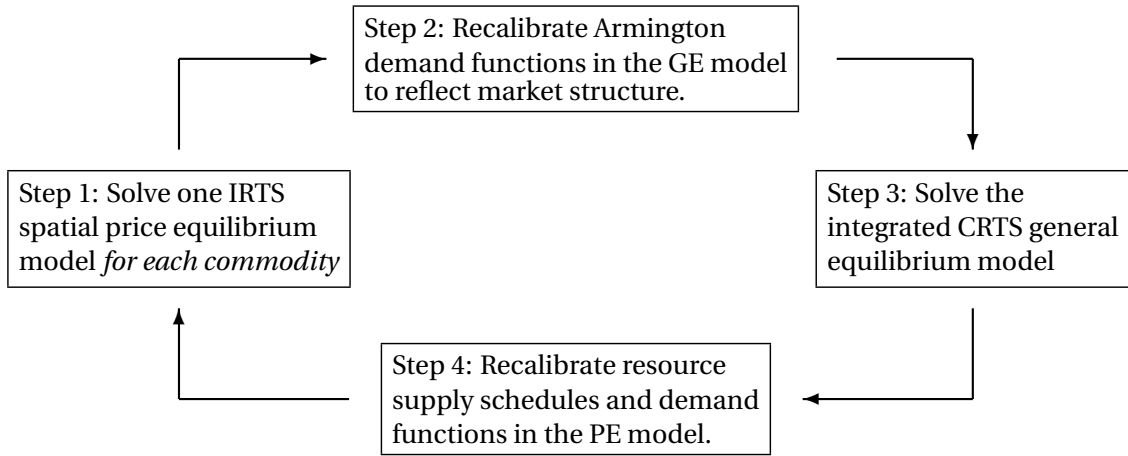
determined simultaneously. In the present application, however, dimensionality and non-convexities argue strongly in favor of decomposition. When we solve the industrial organization model on a market by market basis, we avoid dealing with excessively high dimensionalities which otherwise arise when there are large numbers of both goods and markets. In addition, we find that decomposition leads to a significant improvement in robustness of the solution method.

The Melitz model incorporates two types of non-convexity. The first is the conventional interaction of prices, quantities, and incomes. Income effects are the source of most of the difficulties in proving convergence for complementarity algorithms [Mathiesen (1987)]. The second non-convexity is associated with the Dixit-Stiglitz aggregation and productivity effects. While it is possible to solve general equilibrium models including Dixit-Stiglitz effects [Markusen (2002)], it is well known that even small instances of the problem class can be extremely difficult. Our decomposition approach seems to avoid these computational difficulties by a *divide-and-conquer* strategy in which income effects are handled in one submodule and productivity effects in a second module.

Table 1: PE module; multiregion heterogeneous-firms partial-equilibrium

Equilibrium Condition	(Equation)	Associated Variable	Dimensions
Zero cutoff profits (ZCP)	(14)	N_{rs} : Number of operating firms	$R \times R$
Free entry (FE)	(17)	M_r : Mass of firms taking a draw	R
Dixit-Stiglitz preferences	(4)	P_r : Price index	R
Firm-level demand	(5)	$\tilde{q}_{rs}$: Average-firm quantity	$R \times R$
Firm-level pricing	(7)	$\tilde{p}_{rs}$: Average-firm price	$R \times R$
CES wtd. Average φ	(12)	$\tilde{\varphi}_{rs}$: Average-firm productivity	$R \times R$
Pareto dist. Marginal φ	(18)	φ_{rs}^* : Marginal-firm productivity	$R \times R$
Input-market clearance	(19)	c_r : Composite-input price	R
Capacity constraint	(20)	π_{rs}^c : Capacity rents	$R \times R$
Total Dimensions:			$3R + 6R^2$

4.1 Partial Equilibrium Module

The exogenous links that make the PE module operational are the expenditure levels in each region, $\bar{E}_r$, (which establish demand for manufactured goods) and the prices, $\bar{c}_r$, and quantities, $\bar{Y}_r$, of the composite inputs to manufacturing. The model needs some flexibility to react to shocks, however, so we assume a constant-elasticity input-supply function centered (each iteration) on the quantity of inputs used by the sector in the general equilibrium ($\bar{Y}_r$). Input supply is thus $\bar{Y}_r (c_r/\bar{c}_r)^\eta$, where $\eta > 0$ is the elasticity. If the PE model is consistent with the general equilibrium $c_r = \bar{c}_r$, where c_r satisfies the equilibrium conditions in both modules.

Table 1 summarizes the nonlinear conditions in the PE module and establishes the complementarity between equations and associated variable. In addition to the conditions developed in the previous section we add the input-market clearance condition (which determines c_r)

$$\bar{Y}_r \left(\frac{c_r}{\bar{c}_r} \right)^\eta = \delta f_r^e M_r + \sum_s N_{rs} \left(f_{rs} + \frac{\tilde{q}_{rs} \tau_{rs}}{\tilde{\varphi}_{rs}} \right), \tag{19}$$

and the complementary-slack condition for determining capacity rents (π_{rs}^c)

$$M_r - N_{rs} \geq 0; \quad \pi_{rs}^c \geq 0; \quad \pi_{rs}^c (M_r - N_{rs}) = 0. \tag{20}$$

As noted above, in a steady-state equilibrium π_{rs}^c will equal zero, but the computational model benefits from an explicit constraint that prevents numeric moves where $N_{rs} > M_r$.[18]

4.2 General Equilibrium Module

The General Equilibrium Module (GE) is formulated as a standard constant-returns model of world trade in all products. Consumers have preferences over goods differentiated by region of origin (the Armington assumption). Consider the unit expenditure function associated with region-s purchases of goods of type k (we reintroduce the commodity index, $k \in K$, in the general equilibrium):

$$P_{ks} = \psi_{ks} \left[\sum_r \xi_{krs} [(1 + t_{krs}) c_{kr}]^{1-\sigma} \right]^{1/(1-\sigma)} \tag{21}$$

Notice that we define the Armington aggregation directly over the composite of the sector (k) inputs from region r, which trade at a price of c_{kr}. The total and relative productivity parameter ψ_{ks} and ξ_{krs} control the functional calibration. These are the instruments through which the GE module is affected by the PE solution.

Table 2 summarizes the full set of equilibrium conditions in the GE module. First we define the aggregate Cobb-Douglas expenditure function;

$$P_r^E = \prod_k (P_{kr})^{\alpha_k}. \tag{22}$$

The remaining transformation technologies are given by our characterization of the commodity-specific price index, (21), and the composite-input cost function, (6).

Each price (index) has an associated market. Let $\bar{e}_{jr}$ be the exogenous endowment of

---

[18]This also indicates how the model might be extended into an intertemporal context where M_r cannot adjust instantaneously.

Table 2: GE module: multiregion constant-returns general equilibrium

Equilibrium Condition	(Equation)	Associated Variable	Dimensions
Optimality conditions:			
Expenditure function	(22)	E_r : Total Expenditures	R
Zero-profits Armington Activity	(21)	A_{kr} : Armingtion activity level	$K \times R$
Zero-profits Production	(6)	Y_{kr} : Sectoral Output	$K \times R$
Market-clearance conditions:			
Input markets	(23)	w_{jr} : Factor price by type	$J \times R$
Product markets	(24)	c_{kr} : Composite-input price	$K \times R$
Armington-Composite markets	(25)	P_{kr} : Price index by commodity	$K \times R$
Gross supply and demand	(26)	P_r^E : Aggregate price index	R
Income balance:			
Final demand	(27)	W_r : Hicks welfare index	R
Factor income + tariff revenue	(28)	I_r : Nominal income	R
			$4R$
Total Dimensions			$+4(K \times R)$
			$+(J \times R)$

factor j in region r. This will equal the quantity demanded;

$$\bar{e}_{jr} = \sum_k \frac{\beta_{jkr} c_{kr} Y_{kr}}{w_{jr}}. \tag{23}$$

In turn the supply of the composite-input activity will equal demand (as derived from the Armington activity):

$$Y_{kr} = \sum_s \xi_{krs} \psi_{ks} A_{ks} \left(\frac{P_{ks}}{(1 + t_{krs}) c_{kr}} \right)^\sigma. \tag{24}$$

Supply of the Armington composite equals gross demand:

$$A_{kr} = \frac{\alpha_k E_r}{P_{kr}}. \tag{25}$$

Gross expenditures equal the value of final demand plus the value of intermediate use:

$$E_r = P_r^E W_r + \sum_k \beta_{kr}^E c_{kr} Y_{kr}. \tag{26}$$

The welfare index is calculated directly from the ratio of income to the price of the aggregate commodity:

$$W_r = \frac{I_r}{P_r^E}. \tag{27}$$

Income in a region equals the value of factor endowments plus tariff revenues:

$$I_s = \sum_j w_{js}\bar{e}_{js} + \sum_k \sum_r t_{krs}\xi_{krs}\psi_{ks}A_{ks}\left(\frac{P_{ks}}{(1+t_{krs})c_{kr}}\right)^\sigma. \tag{28}$$

4.3   Full Solution

The challenge to arriving at a fully consistent general equilibrium is to adjust the ψ_{ks} and ξ_{krs} (where $k =$ Manufacturing) such that aggregate supply of the manufacturing composite and relative demands for inputs are consistent with the PE solution. Changes in the number of firms will indicate total and relative productivity changes in the composite inputs embodied in the trade flow. Once these productivity changes are incorporated the GE module can be solved to find a new set of gross expenditures, input prices, and input quantities to pass back to the PE module. At the global solution there are no additional adjustments in the ψ and ξ, and the common variables across the PE and GE modules have the same solution values.

In passing information from the PE module to the GE, we first establish total factor productivity by relating ψ_{ks} to the ratio of the price index and the total value of manufacturing expenditures:

$$\psi_{ks} = \frac{P_{ks}}{\alpha_{ks}E_s}. \tag{29}$$

An increase in the total number of varieties consumed in region s, $\left(\sum_r N_{rs}\right)$, indicate a decrease in the computed PE price index. Through equation (29) the Dixit-Stiglitz effect is carried over to the GE Armington technology. Finding the relative productivity changes involves finding the set of ξ_{krs} that are consistent with the value of input demands in the PE

module. All payments are to the factors so we must have:

$$c_{kr}\xi_{krs}\psi_{ks}A_{ks}\left(\frac{P_{ks}}{(1+t_{krs})c_{kr}}\right)^{\sigma} = N_{krs}\tilde{p}_{krs}\tilde{q}_{krs}. \tag{30}$$

Solving for ξ_{krs}, and noting that A_{ks} is the inverse of the new ψ, we have

$$\xi_{krs} = \frac{N_{krs}\tilde{p}_{krs}\tilde{q}_{krs}}{c_{kr}\left(\frac{P_{ks}}{(1+t_{krs})c_{kr}}\right)^{\sigma}} \tag{31}$$

The recalibration of the constant returns GE (ψ_{ks} and ξ_{krs}) based on the heterogeneous-firms PE solution, and the subsequent recalculation of the $\bar{E}_r$, $\bar{c}_r$, and $\bar{Y}_r$, has proven to be a robust solution method. The iterative procedure stops at the point that all variables common to the PE and the GE are consistent and there is no further recalibration indicated.

5 Nonlinear Least-squares Estimation

5.1 Estimation strategy

Consider that the $B = 3R + 6R^2$ nonlinear conditions in the PE module presented above might be written as $F(\boldsymbol{x},\boldsymbol{\gamma}) = 0$, which implicitly maps a set of exogenous parameters , $\boldsymbol{\gamma} \in \mathbb{R}^A$, to a vector of endogenous variables $\boldsymbol{x} \in \mathbb{R}^B$. Let $\hat{\boldsymbol{\gamma}} \in \{\mathbb{R}^{\hat{A}} : \hat{A} \leq A\}$ denote a vector of core parameters to be estimated, and let $\hat{\boldsymbol{x}} \in \{\mathbb{R}^{\hat{B}} : \hat{B} \leq B\}$ denote a key endogenous series (e.g., bilateral trade flows). Our estimation strategy is to find the $\hat{\boldsymbol{\gamma}}$ that minimize the sum of the squared differences between the $\log \hat{\boldsymbol{x}}$ and observed $\log \boldsymbol{x}^0$ subject to $F(\boldsymbol{x},\boldsymbol{\gamma}) = 0$ and an

additional $A - \hat{A}$ direct assumptions about the values of the remaining parameters:

$$\min_{\{\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{x}}\}} \quad \|\log \hat{\boldsymbol{x}} - \log \boldsymbol{x}^0\|^2$$

$$\text{subject to:} \quad F(\boldsymbol{x}, \boldsymbol{\gamma}) = 0,$$

$$\text{and} \quad \bar{\boldsymbol{\gamma}} = \mathbf{k},$$

where $\bar{\boldsymbol{\gamma}}$ are the assumed parameters and $\mathbf{k}$ is a vector of constants.[19] The norm is defined in logs to be consistent with the empirical trade literature, which often assumes a log-linear form of the trade equation.

We utilize data that is commonly employed in gravity estimations. The economic data includes gross manufacturing output by region, bilateral trade flows, and measured tariffs. Because we are interested in fitting a complete general equilibrium (including various non-manufacturing sectors), we take these data from the Global Trade Analysis Project (GTAP) [Dimaranan (2006)], a data set commonly employed in general equilibrium simulations of trade policy changes. These 2001 data track factor endowments, industry production, input-output relationships and bilateral trade for a broad set of countries. The GTAP data has been balanced for use in general equilibrium studies (household income equals expenditure, for example).[20]

We aggregate the data into twelve regions,

CHN China	JPN Japan	CAN Canada	USA United States
MEX Mexico	ANZ Australia and N. Zel.	KTW Korea and Taiwan	ROA Rest of Asia
LAM Latin America	EUR Europe	EER Eastern Europe and FSU	ROW Rest of World ;

and seven aggregate sectors,

---

[19]In addition to the assumed parameters, the general equilibrium values of gross expenditures and manufacturing output are given in the benchmark data; $\bar{E}_r$ and $\bar{c}_r \bar{Y}_r$ are observed. We estimate conditional on matching these values. The endogenous variable c_r is determined (and fixed in estimation) by selecting units such that $c_r = \bar{c}_r = 1$, which uniquely determines $\bar{Y}_r$ as it enters eq. (19).

[20]A key advantage of our data and approach, relative to other studies in the geography of trade literature, is that we are able to capture aspects of observable reality that are often missing in such studies. For example, we are able to account for a sizable role for the services sector in the economy, for observable cross-country differences in input bundles, and for geographical variation in intermediate demands. The cost of incorporating such richness is a reduced ability to incorporate a large number of regions in the analysis

AGR Agriculture   MTL Mtls-related industry   EIS Other Energy Intensive
MFR Manufacturing   SER Services   ENG Energy
CGD Savings good .

We focus our estimation of the heterogeneous-firms model on the aggregate manufacturing sector (the subscript k is thus suppressed in the remainder of our estimation description). The other sectors in the general equilibrium are assumed competitive and calibrated via the usual techniques.[21] The regional aggregation attempts to strike a balance between our desire to capture the relevant geographic trade patterns and provide useful reports given the data and computational limitations.[22]

In addition to the economic data we utilize distances between regions to inform transportation costs.[23] Consistent with the gravity literature we assume that iceberg trade costs take on the following form:

$$\tau_{rs} = d_{rs}^{\theta};$$

where θ is an estimated distance elasticity.[24] In addition, we include rent generating ad valorem tariffs as measured in the GTAP data, t_{rs}. The c.i.f. import prices, thus, include the variable trade costs $(1 + t_{rs})\tau_{rs}$.

Taking the GTAP data as given, and given our assumed structure of trade costs, we have the following candidates for inclusion in $\hat{\gamma}$:

---

[21]We assume an Armington trade structure, with constant returns and perfect competition, in the sectors other than manufacturing.

[22]Earlier versions of this paper reported results from a nine-region aggregation (which provided no country-level reports). In addition, we have explored variations on other regional aggregations of the GTAP data. The estimated distance elasticities, Pareto shape parameters, and overall welfare analysis of integration where not substantially altered by our choice of geographic aggregation. The region specific fixed costs of trade are not estimated with great precision, and are therefore more sensitive to the particular aggregation.

[23]Our baseline distance data are described in Mayer and Zignago (2006). In our cross-country aggregation we must aggregate distances, and we apply Mayer and Zignago (2006)'s *distw* method for aggregating across bilateral city pairs to aggregate over bilateral country pairs.

[24]A normalization of distance is required to pin down the absolute scale of distance-related costs. We scale distance such that $\tau = 1$ on the shortest link.

σ : the inter-variety elasticity of substitution,
δ : the probability of firm death,
a : shape parameter for the Pareto distribution,
b : minimum productivity parameter for Pareto distribution,
θ : distance elasticity of iceberg trade costs,
f_r^e : fixed entry cost,
f_{rs} : bilateral fixed cost of shipping from region r to region s.

Informing these parameters off observed bilateral flows is not meaningful unless we are willing to significantly reduce the parameter space (beyond our implicit assumptions that the core distribution, substitution, and transport cost parameters are identical across regions). With R regions there are potentially $\hat{B} = R^2$ observable flows, but there are at least $R^2 + R + 5$ parameters. We might eliminate σ from the list of parameters; noting that we are interested in identifying trade costs conditional on second-order curvature.[25] We assume that $\sigma = 3.8$ throughout our analysis following the plant-level empirical analysis of Bernard et al. (2003). In addition, we directly assume the values $\delta = 0.025$, $f_r^e = 2$, and $b = 0.2$ following Bernard et al. (2007).[26]

The primary assumption that we employ to reduce the parameter space is to impose structure on the fixed costs. Let f_r^p be a fixed cost that is specific to goods produced in region r, and f_s^x be a fixed cost that is specific to goods exported to region s. Now consider decomposing the bilateral fixed costs as follows:

$$f_{rs} = \begin{cases} f_r^p + f_s^x + f_{rs}^r, & \text{for } r \neq s; \\ f_r^p + f_{rs}^r, & \text{for } r = s. \end{cases}$$

When $r = s$ the f_s^x term drops out reflecting the idea that f_s^x is an outward trade barrier. The f_{rs}^r are idiosyncratic residual bilateral costs. In the initial estimation the f_{rs}^r are constrained

---

[25]Our approach is similar to Anderson and van Wincoop (2003) in that we estimate trade costs conditional on σ.

[26]Bernard et al. (2007) explain that changes in f^e rescales the mass of firms where as changes in δ rescale the mass of entrants relative to the mass of firms. For consistency we simply adopt their values.

to be zero. The number of parameters to be estimated is thus reduced to $\hat{A} = 2R + 2$. Once the core parameters are estimated, and locked down, the system can be used to calculate the matrix of residual f_{rs}^r which generate an exact fit on trade flows.[27]

5.2 Estimation Results

The primary purpose of our nonlinear estimation is to complete a calibration of the numerical model. This entails a complete enumeration of the structural parameters necessary to reconcile the structural model with observed data. The primary parameters of interest are those that are taken to be common across the world: the shape of the implied Pareto distribution of productivity draws, a, and the distance elasticity of trade costs, θ. The model links both these parameters to the geographic pattern of trade.[28]

We conduct three econometric calibrations of the model. In the first, we allow both θ and a to be free parameters; they take the values that minimize the econometric objective, subject to the constraints defined by the model and our choices of the parameters in $\bar{\gamma}$. Very good estimates of θ appear in the literature, and our second set of estimates constrains the estimation procedure to replicate a commonly accepted value, $\bar{\theta} = 0.27$. As a sensitivity check, our third set of estimates imposes the constraint $\bar{\theta} = 0.46$.[29] Our estimates of key structural parameters appear in Table 3.[30]

The interaction between θ and a is a key point of interest. Conditional on the bilateral

---

[27]Balistreri and Hillberry (2008) use a similar technique of stochastic estimation and subsequent exact-fit calibration. This entails a consideration that the econometric residuals might logically be interpreted as idiosyncratic calibration parameters (as opposed to measurement error).

[28]θ governs the degree to which delivered prices rise over distance. Chaney (2008) shows that a exerts a substantial influence on the geography of bilateral trade, via the extensive margin.

[29]These latter two estimates are taken from Hummels (2001), who estimates θ directly off observed transportation cost margins. 0.27 is Hummels' central estimate, using data from 7 countries that report transport margins in their international trade statistics. 0.46 is the elasticity of air freight charges with respect to distance in U.S. data, and Hummels uses this as a plausible upper bound on θ. Our unconstrained estimate of θ lies well below Hummels' central estimate, and we treat this as a lower bound.

[30]In order to characterize the degree to which our procedure fits the observed trade pattern, we conduct a log linear regression of observed flows on fitted flows. This regression returns an R^2 of 0.939, 0.897, and 0.827, respectively.

Table 3: Nonlinear estimation results (dependent variable is log bilateral flows; core fixed parameters are $\sigma = 3.8$, $\delta f^e = 0.05$, and $b = 0.2$)

		Specification		
		$\theta =$free	$\bar{\theta} = 0.27$	$\bar{\theta} = 0.46$
Pareto shape parameter: a		5.171	4.582	3.924
		(0.336)	(0.227)	(0.019)
Distance elasticity: θ		0.155	0.27	0.46
		(0.028)		
Source-specific fixed cost: f_r^p				
CHN	China	0.036	0.026	0.430
		(0.385)	(0.396)	(0.220)
JPN	Japan	0.300	0.340	0.519
		(0.239)	(0.212)	(0.237)
CAN	Canada	0.347	0.303	0.228
		(0.305)	(0.210)	(0.103)
USA	United States	0.299	0.172	0.328
		(0.337)	(0.369)	(0.065)
MEX	Mexico	0.024	0.063	0.316
		(0.185)	(0.153)	(0.088)
ANZ	Australia and N. Zel.	0.039	0.030	0.019
		(0.014)	(0.004)	(0.000)
KTW	Korea, and Taiwan	0.047	0.027	0.023
		(0.088)	(0.070)	(0.008)
ROA	Rest of Asia	0.047	0.026	0.016
		(0.074)	(0.020)	(0.003)
LAM	Latin America	0.218	0.255	0.120
		(0.262)	(0.170)	(0.008)
EUR	Europe	0.503	0.330	0.753
		(0.322)	(0.312)	(0.157)
EER	Eastern Europe and FSU	1.300	0.562	0.222
		(0.518)	(0.338)	(0.022)
ROW	Rest of World	3.666	1.280	0.237
		(1.449)	(0.738)	(0.015)
Destination-specific fixed cost: f_s^x				
CHN	China	0.258	0.147	20.647
		(4.402)	(6.379)	(15.872)
JPN	Japan	0.872	0.493	0.035
		(0.986)	(0.461)	(0.102)
CAN	Canada	15.153	13.453	0.0 **bound**
		(11.352)	(7.838)	(0.006)
USA	United States	0.714	0.149	0.0 **bound**
		(1.021)	(0.707)	(0.000)
MEX	Mexico	0.0 **bound**	0.0 **bound**	0.0 **bound**
		(0.300)	(0.052)	(0.000)
ANZ	Australia and N. Zel.	6.291	12.004	37.512
		(3.184)	(6.714)	(1.324)
KTW	Korea, and Taiwan	0.913	0.346	0.0 **bound**
		(1.003)	(0.673)	(0.001)
ROA	Rest of Asia	1.602	2.067	2.708
		(1.618)	(1.497)	(1.723)
LAM	Latin America	8.254	23.532	33.984
		(9.412)	(20.009)	(0.992)
EUR	Europe	1.339	0.293	0.0 **bound**
		(1.293)	(0.403)	(0.000)
EER	Eastern Europe and FSU	55.005	35.110	18.628
		(18.955)	(3.000)	(0.505)
ROW	Rest of World	142.158	98.561	59.633
		(47.798)	(6.241)	(1.294)

Table 4: Heterogeneity in the productivity distribution: $\frac{\varphi}{b}$ for selected values of a

	Percentiles			
	50th	75th	90th	95th
Fitted values of a				
$\hat{a}$=5.171	1.143	1.307	1.561	1.785
$\hat{a}$=4.582	1.163	1.353	1.653	1.923
$\hat{a}$=3.924	1.193	1.424	1.798	2.146
Implied estimate from Eaton et al. (2004)				
a=4.2	1.179	1.391	1.730	2.041
Value used in Bernard et al. (2007)				
a=3.4	1.226	1.503	1.968	2.414

trade pattern, our procedure must assign responsibility for trade reductions to these two parameters (along with the fixed costs). Our unconstrained estimate of θ is $\hat{\theta} = 0.155$, while $\hat{a} = 5.171$. This is a relatively low estimated distance elasticity, and a somewhat high estimate for the Pareto distribution parameter.[31] As we constrain $\bar{\theta}$ to higher values, the estimated value of a falls. For $\bar{\theta} = 0.27$, $\hat{a} = 4.582$, and for $\bar{\theta} = 0.46$, $\hat{a} = 3.924$.

The lower values of a that occur in our restricted estimates imply greater heterogeneity in firm productivities. Table 4 illustrates some features of the productivity distributions implied by different values of a. For our unconstrained estimate of a, a firm with a productivity draw at the median of the distribution would be 1.143 times as productive as a firm with the minimum draw. As a falls, the productivity distribution flattens out. In our subsequent counterfactual scenarios, we will be employing the constrained estimate $a = 4.582$, the estimate corresponding to $\bar{\theta} = 0.27$. In this case, the median productivity draw is 1.163 times the size of the minimum draw.

The structural estimation results in Table 3 also contain estimated values of source- and destination-specific fixed costs (f_r^p and f_s^x, respectively). Just as other structural procedures impute values for model-consistent P_s, our fitting procedure imputes model-consistent fixed costs (as well as P_s). Model consistency includes two requirements. First, the size of $f_r^p + f_s^x$

---

[31]As noted earlier, Hummels' central estimate of θ is 0.27. Estimates of a - which are taken from distributions of plant/firm level market shares - vary, and are conditional on a choice of σ. Bernard et al. (2007) choose $a = 3.4$, and the estimates in Eaton et al. (2004) imply $a = 4.2$ under our maintained assumption that $\sigma = 3.8$.

should be such that the average firm pays $\frac{a+1-\sigma}{a\sigma}$ of revenue, net of tariffs, in fixed costs.[32]

Second, the elasticity of fitted trade with respect to $f_r^p + f_s^x$ should equal $1 - \frac{a}{\sigma-1}$.[33] Conditional on other trade resistances, the model attributes *home bias* in fitted trade flows to f_s^x. The source-specific fixed costs, f_r^p, are, in effect, source-specific fixed effects that help the procedure to best fit the data under these constraints.

The results in Table 3 indicate that both relative and absolute estimates of fixed costs vary with θ and a.[34] This is unsurprising, as both the adding up conditions of the model and the marginal conditions derived by Chaney (2008) link fixed costs to the values of these parameters. Some destinations (i.e. China) have substantial variation across columns in the implied fixed costs of importing, f_s^x. Substantial differences across columns are understandable. Observed levels of import penetration are being replicated across successive calibrations, even as changes in θ are substantially shifting the burden of trade resistance from distance related ad valorem costs to fixed costs of trade.

Our non-linear estimating system fits the data in a manner quite similar to a conventional gravity equation. An OLS regression of $\log(\hat{x})$ on (logged) d_{rs}, $(1 + t_{rs})$, $(f_r^p + f_s^x)$, $c_{kr}Y_{kr}$, $\alpha_k E_s$, and P_s returns a perfect fit. This intuitive OLS specification relies, of course, on our non-linear estimates of f_r^p, f_s^x, and P_s.[35] Following Hillberry and Hummels (2008) we decompose trade flows into extensive and intensive margins, regressing $\log N_{rs}$ and $\tilde{p}_{rs}\tilde{q}_{rs}$ on the determinants of bilateral trade. We report these results in Table 5. The results show

---

[32] See eq. (14).

[33] This relationship is derived in Chaney (2008).

[34] Standard errors for the estimated fixed costs are generally tighter when the distance elasticity is fixed. Fixing θ allows a to be more precisely estimated (exploiting any unexplained variation in the distance elasticity of trade, for example). More precise estimates of a allow more precise estimates of the fixed costs in the model.

[35] The fitted values of trade in the regression are fully consistent with optimizing behavior of the agents in the model. As Helpman et al. (2008) show, a Pareto distribution for firm heterogeneity coupled with an assumption that fixed costs are exporter- and importer-specific (which is true in the case of our fitted values) generates a multiplicative gravity relationship (and thus a perfect fit in a properly specified log-linear regression of structurally fitted flows on the appropriate dependent variables). The estimates here link the flows to variables that appear in the general equilibrium trade model; there is no need for a summary measure like *outward multilateral resistance* as described in Anderson and van Wincoop (2004).

Table 5: Intensive and Extensive Margins in Fitted Bilateral Trade Flows

	$\theta = free$		$\theta = 0.27$		$\theta = 0.46$	
Regressand	N_{rs}	$\tilde{p}_{rs}\tilde{q}_{rs}$	N_{rs}	$\tilde{p}_{rs}\tilde{q}_{rs}$	N_{rs}	$\tilde{p}_{rs}\tilde{q}_{rs}$
Regressor						
d_{rs}	-0.799	0	-1.236	0	-1.804	0
	(0.001)	(0)	(0.001)	(0)	(0.001)	(0)
$(1+t_{rs})$	-7.037	1	-6.240	1	-5.349	1
	(0.006)	(0)	(0.005)	(0)	(0.005)	(0)
$(f_r^p + f_s^x)$	-1.846	1	-1.636	1	-1.401	1
	(0.000)	(0)	(0.000)	(0)	(0.000)	(0)
$c_{kr}Y_{kr}$	1.000	0	1.000	0	1.001	0
	(0.000)	(0)	(0.000)	(0)	(0.000)	(0)
$\alpha_{ks}E_s$	1.846	0	1.635	0	1.401	0
	(0.000)	(0)	(0.001)	(0)	(0.001)	(0)
P_s	5.165	0	4.578	0	3.921	0
	(0.003)	(0)	(0.002)	(0)	(0.002)	(0)
constant	-11.328	2.114	-9.800	2.281	-8.074	2.585
	(0.004)	(0)	(0.004)	(0)	(0.004)	(0)

Note: All variables are in natural logarithms. Standard errors in parentheses.
All regressions have 144 observations and $R^2 = 1$.

that average firm revenues (c.i.f.) rise in tandem with tariffs and fixed trade costs.[36] Most trade responses to the gravity-type variables are observed as changes in the number of firms serving a particular market. The regression coefficients in Table 5 can be decomposed in each case into theory-consistent functions of the structural parameters a, σ, and θ.[37]

Once the core parameters reported in Table 3 are established we can freeze these at their point estimates and find a set of residual bilateral costs, f_{rs}^r, that give us perfect consistency with observed trade flows.[38] From the perspective of the nonlinear estimation these are

---

[36]Note that the implications for the intensive margin follow directly from eq. (15).

[37]Derivations in Chaney (2008) follow through for all but the coefficients on $(1+t_{rs})$ and $\alpha_k E_s$. Chaney does not consider rebatable tariff revenues, and in our framework the elasticity of bilateral trade to $(1+t_{rs})$ is $1 - \frac{a\sigma}{\sigma-1}$. Our estimated elasticity of N_{rs} with respect to destination expenditures takes the value $\frac{a}{\sigma-1}$, whereas in Chaney (2008) it is 1. The difference is that in our framework, high productivity firms are able to replicate their technology and serve each foreign market with a dedicated export facility. This treatment makes N_{rs} more sensitive to market size in our framework.

[38]There are a number of potential matrices of residual fixed costs that are consistent with observed trade and

Table 6: Total Bilateral Fixed Trade Costs (f_{rs})

	CHN	JPN	CAN	USA	MEX	Destination ANZ	KTW	ROA	LAM	EUR	EER	ROW
Source												
CHN	0.028	1.005	1.950	0.080	0.010	8.399	2.781	1.813	2.511	0.293	8.962	1.933
JPN	1.986	0.313	2.826	0.111	0.581	4.243	1.539	0.380	1.304	0.227	23.574	2.518
CAN	0.323	0.103	0.048	0.184	4.319	4.739	0.279	2.093	4.894	0.611	35.851	8.298
USA	0.654	0.341	5.180	0.043	3.403	5.077	0.098	0.384	4.893	0.478	64.550	3.970
MEX	1.668	1.146	2.571	0.133	0.138	19.710	0.706	4.606	4.018	1.153	30.763	38.193
ANZ	0.195	0.010	0.183	0.056	0.094	0.026	0.017	0.087	1.788	0.016	3.002	0.397
KTW	0.703	3.016	0.734	0.034	0.110	2.436	0.021	0.244	0.296	0.091	8.074	0.508
ROA	0.449	0.123	0.621	0.017	0.178	2.187	0.080	0.022	2.389	0.068	18.781	1.478
LAM	0.258	0.185	5.982	0.251	1.006	20.952	0.071	0.160	0.036	0.127	1.550	1.670
EUR	0.343	0.394	5.147	0.324	0.895	1.738	0.182	0.592	2.738	0.145	11.974	2.172
EER	0.678	0.583	8.884	1.217	2.050	14.418	0.859	3.406	12.735	0.316	0.195	6.599
ROW	3.132	1.289	25.687	0.378	18.280	23.717	1.494	3.667	40.663	0.510	23.840	0.030

effectively econometric residuals—they allow the structure to fit the data exactly. Alternatively, from the perspective of performing theory consistent counterfactual analysis they are idiosyncratic calibration parameters.[39] Table 6 shows the full matrix of total bilateral fixed costs including the residual plus the source and destination charges. So, for example, we might consider that import penetration into EER is difficult, but it is particularly difficult for American firms. On this particular link the total fixed cost of 64.6 is nearly twice as large as the base destination charge to get into EER, 35.1.[40]

6   Counterfactual Simulations

We analyze four scenarios that compare the impacts of tariff and fixed cost reductions:

---

the estimated parameters. We choose the one that minimizes the squared residual bilateral costs.

[39]Hillberry et al. (2005) show the usefulness of framing standard general-equilibrium calibration exercises as the systematic identification of idiosyncratic residual parameters. As in any standard econometric exercise these residual parameters are useful indicators of model fit. In our preferred specification, variation in the estimated f_{rs} explains 14 percent of the variation in bilateral trade flows.

[40]Relatively large fixed costs of entry into aggregated regions like EER may be partially attributed to geographical aggregation. Our aggregation implicitly treats the collection of smaller countries in EER as a large integrated market, so large fixed costs are needed to explain a relatively low volume of trade.

Table 7: Counterfactual Welfare Impacts (% Equivalent Variation)

	Scenario			
	A	B	C	D
	CRTS-Tariff	Tariff	Fix Cost	Both
Region				
CHN	0.3	1.3	2.9	4.8
JPN	0.1	0.3	1.5	1.7
CAN	-0.0	0.1	4.2	4.3
USA	0.0	0.0	0.8	0.9
MEX	0.2	0.7	4.3	5.3
ANZ	0.4	1.4	2.0	4.0
KTW	0.4	1.1	5.3	6.4
ROA	0.3	1.0	4.8	6.0
LAM	0.1	0.4	1.4	2.1
EUR	0.1	0.2	1.2	1.5
EER	-0.1	-0.3	3.8	4.0
ROW	-0.2	-0.7	2.1	1.6

(A)   Armington constant-returns formulation with a 50% reduction in manufacturing tariffs;
(B)   Heterogeneous-firms model with a 50% reduction in manufacturing tariffs;
(C)   Heterogeneous-firms model with a 50% reduction in fixed trade costs; and
(D)   Heterogeneous-firms model with both the tariff and fixed cost cuts.

Scenario (**A**) is a reference case where we assume a standard Armington trade structure and constant-returns production.[41] Table 7 shows welfare changes induced by the tariff cut. Although most regions gain from the tariff cuts, three regions suffer welfare losses (CAN, EER, and ROW).[42] Examining the same tariff cuts in the heterogeneous-firms model, Scenario (**B**), indicates substantially greater gains. Taking the simple-average welfare change across regions the heterogeneous-firms structure indicates welfare gains from liberalization that are nearly four times larger than in the baseline case. The simple-average welfare gain of 0.5% in Scenario (**B**) may not seem particularly impressive, but consider the following statistics from our aggregation of the GTAP data: gross manufacturing output is only 26% of world

---

[41]We simply run the tariff cut on the GE module without making the iterative productivity adjustments. This gives us a perfectly comparable constant-returns benchmark to judge the performance of the new theory.

[42]This is not particularly surprising. At low substitution elasticities (3.8 in this case) the Armington structure implies high optimal tariffs, so the benchmark tariff structure is likely to benefit some regions. See Brown (1987) or Balistreri and Markusen (2008).

gross output, only 17% of manufacturing output is traded to another region, and the simple average benchmark tariff on these flows is only 9.7%. So the typical tariff cut is less than 5%, and applies to less than 5% of gross output. In this context, an average welfare gain of 0.5% seems quite large. With the exception of the `ROW` and `EER` regions, tariff cuts in the heterogeneous-firms structure produce larger net welfare gains than the constant returns benchmark.

In Scenario (**C**) we examine a 50% cut in the fixed costs associated with non-domestic trade links. This generates important gains across the board. The results are consistent with a recent trade literature focussing on the relative importance of unobserved (non-tariff) barriers and tariffs.[43] In Scenario (**D**) both the tariff and fixed cost reductions are combined. There are considerable increases in welfare under Scenario (**D**) considering that Manufacturing is the only sector being liberalized. The simple-average welfare gain under Scenario (**D**) is nearly 30 times larger than in the Armington reference case. Notice also that fixed-cost (or non-tariff barrier) reductions often complement tariff cuts. Absolute welfare increases of 2% to 6% are considerably larger than most computational estimates of the value of trade liberalization.[44]

As noted above, one of the key critiques of current policy simulation models is that they fail to account for the productivity growth associated with trade liberalization. Table 8 indicates the simulated gains in average productivity across firms active in their respective domestic markets. Consistent with the arguments put forward by the proponents of the heterogeneous-firms model, our simulations show productivity gains due to liberalization. Increased exposure to external markets, whether induced by a reduction in tariff or non-tariff barriers, induces productivity growth.

The other key component of the model is that trade policy affects the extensive margin. The number of foreign varieties increases when trade costs fall. The threshold for import

---

[43]Anderson and van Wincoop (2004)

[44]See Rutherford and Tarr (2002).

28

Table 8: Domestic-firm Productivity Growth (% Change)

	Scenario		
	B	C	D
	Tariff	Fix Cost	Both
Region			
CHN	0.8	1.4	2.5
JPN	0.5	1.4	2.2
CAN	0.7	4.7	5.5
USA	0.3	1.5	1.9
MEX	1.1	4.2	5.7
ANZ	2.1	2.7	6.1
KTW	1.1	3.4	5.0
ROA	1.3	3.2	5.0
LAM	1.1	1.4	3.0
EUR	0.4	1.5	2.1
EER	0.8	2.8	4.0
ROW	0.6	1.5	2.3

penetration falls and more foreign firms find it profitable to enter a given market. In contrast, the effect of changes in trade costs on the number of domestic varieties is not clear. The number of exporting firms increases and the profits of all exporting firms increase, which induces entry of new varieties. The increased activity of these firms, however, bids up the input price. This effect, together with substitution toward imported varieties, induces exit of domestic varieties with low productivity realizations.

On net, however, consumers will likely benefit from lost domestic varieties because factor returns increase and the remaining domestic varieties are less expensive. More productive firms optimally price lower, so eliminating low productivity firms depresses the average price. All of the variety and price effects can be summarized in the solution price index on manufactured goods, P_r. Table 9 presents the percentage change in the price index across the scenarios. Further, we break out the variety effects in Table 10. Although the number of overall varieties falls for many regions, trade growth on the extensive margin combined with lower domestic prices result in lower overall price indexes and welfare gains.

Table 9: Manufacturing Price Index, P_r (% Change)

	Scenario		
	B	C	D
	Tariff	Fix Cost	Both
Region			
CHN	0.7	-0.9	-1.0
JPN	0.1	0.3	-0.2
CAN	-0.6	-4.3	-4.9
USA	-0.2	-2.2	-2.6
MEX	-2.4	-3.8	-6.3
ANZ	0.3	-2.4	-3.3
KTW	-0.1	-1.7	-2.5
ROA	-0.6	-2.1	-3.3
LAM	-1.0	-2.4	-3.7
EUR	0.1	-1.4	-1.7
EER	-1.0	-3.6	-5.0
ROW	-2.3	-3.5	-5.7

7   Conclusion

A broad body of empirical literature documents persistent differences in plant-level productivity. This literature has also shown that the reallocation of production activities, from less- to more-productive plants, is an important part of aggregate productivity growth. These basic characteristics of industrial organization have important implications for international trade and commercial policy. The unifying theory proposed by Melitz (2003) offers insights into these implications. Our contribution is to present a quantitative assessment of the effects of trade cost changes within this structure.

In the case of a 50% reduction in tariffs on traded manufactured goods the simple-average welfare gains are on the order of four times greater in the Melitz structure than in a conventional trade policy model. These gains are complemented, and compounded, when reductions in the inferred fixed costs are also considered. When we add a 50% reduction in cross-border fixed costs to the tariff cuts the welfare gains grow to roughly 30 times what is measured in the constant-returns reference liberalization.

Table 10: Changes in the Number of Operating Firms, N_{rs} (% Change)

	Scenario		
	B	C	D
	Tariff	Fix Cost	Both
Imported Varieties			
(extensive margin):			
CHN	28.8	196.8	275.4
JPN	54.7	212.2	378.1
CAN	11.0	157.9	190.1
USA	8.9	155.2	178.0
MEX	207.4	158.4	716.9
ANZ	25.9	174.9	230.4
KTW	22.1	206.0	266.0
ROA	20.8	190.2	246.2
LAM	27.2	155.4	225.2
EUR	17.4	178.0	225.2
EER	20.4	168.1	225.1
ROW	12.4	147.9	186.5
Domestic Varieties:			
CHN	-2.1	-4.8	-7.3
JPN	-2.1	-5.8	-8.9
CAN	-2.3	-16.1	-19.6
USA	-1.1	-6.6	-7.8
MEX	-4.5	-14.7	-20.4
ANZ	-6.9	-13.0	-20.3
KTW	-3.8	-11.7	-17.0
ROA	-4.4	-11.0	-16.9
LAM	-3.9	-6.0	-11.1
EUR	-1.5	-6.4	-8.2
EER	-3.2	-9.2	-12.6
ROW	-2.7	-5.2	-8.8
Total Varieties:			
CHN	-1.9	-3.4	-5.4
JPN	8.3	33.9	61.7
CAN	-2.2	-14.4	-17.6
USA	0.0	11.1	12.5
MEX	26.6	10.7	87.8
ANZ	-6.8	-12.4	-19.5
KTW	-2.6	-1.8	-4.2
ROA	-3.9	-7.3	-12.2
LAM	-3.8	-5.5	-10.3
EUR	0.8	15.3	19.2
EER	-3.0	-8.1	-11.1
ROW	-2.6	-4.7	-8.1

Some truth in advertising is in order for our results. First, we employ a novel method for measuring unobserved fixed costs. We depart from the econometric literature by employing a nonlinear estimation that includes *extensive-form* conditions as side constraints. Our focus is on arriving at fitted values, while maintaining complete consistency between the econometric and simulation models. Our estimation method is also a stark departure from traditional calibration methods used to fit simulation models; we do not allow preference-bias parameters to drive trade. The onus of explaining the observed pattern of trade is on the theory and the standard parameters that appear in the theory, not on added preference-bias parameters. The very large fixed costs that we estimate are open to criticism, and we view them as crude indicators of how big the barriers may be. It is generally accepted by economists that unobserved trade costs are an important component of the world trade equilibrium. We follow one of the few paths available, which is to accept the structure fully and use it to inform unobservables from the observables.

The second major caveat that we place on our results involves the data. We accept the GTAP data as given and further aggregate it. This is useful in terms of reducing computational complexity and in allowing us to efficiently summarize reports. The GTAP data are balanced; they have already been fitted to a set of fundamental accounting identities. The data are consistent with general-equilibrium adding-up restrictions, but the original fitting procedure weakens the validity of any statistical inference that one might draw from our estimation.

The usual aggregation biases abound in our data, and we have additional concerns given the theory's focus on firm-level behavior. Our aggregate manufacturing sector is not a satisfying definition of an industry or product. Regional aggregation is also problematic. The aggregate rest-of-world region is actually numerous small disjoint markets rather than a large integrated market. We probably overstate the fixed costs of entering the aggregate regions because large fixed costs are necessary for explaining the relative lack of trade with artifi-

cially large regions.

We thus present our estimates conditional on the particular aggregation of the data, the assumed structure, and our maintained hypotheses about key structural parameters. We see important extensions in the area of regional and industry disaggregation. We are somewhat unique in our development of an econometric method that facilitates directly, and fully consistent, welfare analysis of policy. Others may find this departure from standard regression analysis useful and relevant. We are firmly within the empirical-trade tradition, which places theory, not established statistical methods, as the foundation for analysis. Given the rich nature of contemporary theory we hope our empirical welfare analysis encourages others to continue developing the literature in this direction.

References

Alvarez, Fernando, and Robert E. Lucas (2007) 'General equilibrium analysis of the Eaton-Kortum model of international trade.' *Journal of Monetary Economics* 54, 1726–1768

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

— (2004) 'Trade costs.' *Journal of Economic Literature* 42(3), 691–751

Anderson, Kym, William J. Martin, and Dominique van der Mensbrugghe (2005) 'Global impacts of the Doha scenarios on poverty.' *World Bank Policy Research Working Paper Series 3735*

Aw, Bee Yan, Xiaomin Chen, and Mark J. Roberts (2001) 'Firm-level evidence on productivity differentials and turnover in Taiwanese manufacturing.' *Journal of Development Economics* 66, 51–86

Balistreri, Edward J., and James R. Markusen (2008) 'Sub-national differentiation and the role of the firm in optimal international pricing.' *Economic Modelling* In Press, Corrected Proof, –

Balistreri, Edward J., and Russell H. Hillberry (2007) 'Structural estimation and the border puzzle.' *Journal of International Economics* 72(2), 451–463

— (2008) 'The gravity model: An illustration of structural estimation as calibration.' *Economic Inquiry* 46(4), 511–527

Bartelsman, Eric J., and Mark Doms (2000) 'Understanding productivity: Lessons from longitudinal microdata.' *Journal of Economic Literature* 38(3), 569–594

Behrens, Kristian, Giordano Mion, Yasusada Murata, and Jens Sudekum (2008) 'Trade, wages, and productivity.' *IZA Discussion Paper No. 3682*

Bernard, Andrew, and J. Bradford Jensen (1999) 'Exceptional exporter performance: Cause, effect, or both?' *Journal of International Economics* 47(1), 1–25

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott (2006) 'Trade costs, firms and productivity.' *Journal of Monetary Economics* 53(5), 917–937

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum (2003) 'Plants and productivity in international trade.' *American Economic Review* 93, 1268–1290

Bernard, Andrew B., Stephen Redding, and Peter K. Schott (2007) 'Comparative advantage and heterogeneous firms.' *Review of Economic Studies* 74, 31–66

Broda, Christian, and David E. Weinstein (2006) 'Globalization and the gains from variety.' *Quarterly Journal of Economics* 121(2), 541–585

Brown, Drusilla K. (1987) 'Tariffs, the terms of trade, and national product differentiation.' *Journal of Policy Modeling* 9(3), 503–526

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

Dimaranan, Betina V. (Editor) (2006) *Global Trade, Assistance, and Production: The GTAP 6 Data Base* (West Lafayette: Center for Global Trade Analysis, Purdue University)

Eaton, Jonathan, and Samuel Kortum (2003) 'Technology, geography and trade.' *Econometrica* 70(5), 1741–1780

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2004) 'Dissecting trade: Firms, industries and export destinations.' *American Economic Review* 94(2), 150–154

Evenett, Simon, and Anthony Venables (2002) 'Export growth in developing countries: Market entry and bilateral trade flows.' *mimeo, University of St. Gallen*

Foster, Lucia, John Haltiwanger, and C.J. Krizan (2001) 'Aggregate productivity growth: Lessons from the microeconomic evidence.' In *New Directions in Productivity Analysis,* ed. Edward Dean, Michael Harper, and Charles Hulten (University of Chicago Press)

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein (2008) 'Estimating trade flows: Trading partners and trading volumes.' *The Quarterly Journal of Economics* 123(2), 441–487

Hillberry, Russell, and David Hummels (2008) 'Trade responses to geographic frictions: A decomposition using micro-data.' *European Economic Review* 52(3), 527–550

Hillberry, Russell H., Michael A. Anderson, Edward J. Balistreri, and Alan K. Fox (2005) 'Taste parameters as model residuals: Assessing the 'fit' of an Armington trade model.' *Review of International Economics* 13(5), 973–984

Hummels, David (2001) 'Toward a geography of trade costs.' *mimeo*, Purdue University

Hummels, David, and Peter Klenow (2005) 'The variety and quality of a nation's exports.' *American Economic Review* 95(3), 704–723

Kehoe, Timothy J. (2005) 'An evaluation of the performance of applied general equilibrium models of the impact of NAFTA.' In *Frontiers in Applied General Equilibrium Modeling: Essays in Honor of Herbert Scarf,* ed. Timothy J. Kehoe, T.N. Srinivasan, and John Whalley (Cambridge University Press) pp. 341–377

Kehoe, Timothy J., and Kim J. Ruhl (2002) 'How important is the new goods margin in international trade?' *mimeo, University of Minnesota*

Markusen, J.R. (2002) *Multinational Firms and the Theory of International Trade* (Cambridge: MIT Press)

Mathiesen, Lars (1987) 'An algorithm based on a sequence of linear complementarity problems applied to a Walrasian equilibrium model: An example.' *Mathematical Programming* 37, 1–18

Mayer, Thierry, and Soledad Zignago (2006) 'Notes on CEPII's distance measures.' *CEPII explanatory note.* http://www.cepii.fr/distance/noticedist_en.pdf

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

Pavcnik, Nina (2002) 'Trade liberalization, exit, and productivity improvement: Evidence from Chilean plants.' *Review of Economic Studies* 69(1), 245–276

Roberts, Mark J., and James Tybout (1997) 'The decision to export in Columbia: An empirical model of entry with sunk costs.' *American Economic Review* 87(4), 545–564

Rutherford, Thomas F., and David G. Tarr (2002) 'Trade liberalization, product variety and growth in a small open economy: a quantitative assessment.' *Journal of International Economics* 56(2), 247–272

Su, Che-Lin, and Kenneth L. Judd (2008) 'Constrained optimization approaches to estimation of structural models.' *SSRN eLibrary*