February 2005

Solving the Elasticity Puzzle in International Economics

Kim J. Ruhl*
University of Texas at Austin

ABSTRACT

In models of international trade, the elasticity of substitution between foreign and domestic goods – the Armington elasticity – determines the behavior of trade flows and international prices. International real business cycle models need low elasticities to match the quarterly fluctuations in trade balances and the terms of trade, but static applied general equilibrium models need high elasticities to account for the growth in trade following trade liberalization. To reconcile these contradictory findings, we construct a model in which quarterly fluctuations are caused by temporary shocks, but tariff changes are permanent. In the model, plants do not change export status in response to temporary shocks, while tariff decreases induce some non-exporters to export. In a calibrated model, this entry increases the measured elasticity with respect to a tariff change to 6.2, while the elasticity in response to temporary shocks is 1.4.

 I would like to thank Timothy Kehoe for his advice and encouragement. I also thank V.V. Chari, Patrick Kehoe, Sam Kortum, Erzo G.J. Luttmer, Edward Prescott, and the members of the Macro/Trade Workshop at the University of Minnesota. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

A common feature of many international trade models is national product differentiation: countries produce and trade differentiated goods that are to some extent substitutable for each other. In these models, the elasticity of substitution between home and foreign produced goods – the *Armington elasticity*, after Armington (1969) – is the critical parameter for determining the behavior of trade flows and international prices. The importance of this parameter has manifested in two of the leading branches of international economics: the international business cycle literature that seeks to understand the high frequency fluctuations in macroeconomic aggregates, and the static applied general equilibrium literature that focuses on explaining the patterns of trade and the effects of trade policy. These two disciplines, however, have very different views on the value of the Armington elasticity.

International real business cycle (IRBC) models need small values of the elasticity to generate the volatility of the terms of trade and the negative correlation between the terms of trade and the trade balance that are found in the data. IRBC modelers commonly use Armington elasticities around 1.5, though sensitivity analysis suggests values even lower than this may be appropriate. (See for example, Backus, Kehoe and Kydland (1994), Zimmermann (1997), or Heathcote and Perri (2002).) Not surprisingly, when empirical researchers have estimated the Armington elasticity from high frequency data they find small estimates that range from about 0.2 to 3.5. In contrast, applied general equilibrium (GE) models need large Armington elasticities to explain the growth in trade volumes that result from a change in tariffs. As shown in Yi (2003), these models need elasticities of 12 or 13 to generate the large growth in trade found in the data. Empirical work on trade liberalizations, as well as cross section regressions relating trade patterns to tariff and non-tariff barriers, find Armington elasticities that range from about 6 to 15, similar to the ones needed in applied GE models.

The key to understanding these two different measurements of the Armington elasticity is to realize that the source of variation in the prices and quantities being measured is different. The high frequency variation in the time series data is caused by transitory shocks to supply or demand. These are exactly the transitory shocks used in the business cycle models that need small Armington elasticities. Trade liberalization, however, can be thought of as a permanent change. When agents react differently to temporary and permanent changes, the measured Armington elasticities will differ.

In this paper we show that a model combining elements of the real business cycle and the applied GE frameworks can reproduce both the low elasticities estimated from the time series data and the high elasticities implied by the growth in trade following a decrease in tariffs. The model features an entry cost of exporting and heterogeneous plants, as in Melitz (2003), but adds aggregate productivity shocks, as found in the IRBC models of Backus, Kehoe and Kydland (1992) and Stockman and Tesar (1995). As in Melitz (2003), entry costs interact with plant level heterogeneity to partition plants into exporters and non-exporters. A plant decides whether to become an exporter by comparing the expected future value of exporting to the cost of entry. If the expected gain from exporting is larger than the cost of entry, the plant becomes an exporter. The movement of plants into and out of exporting in response to temporary changes in productivity or permanent changes in tariffs drives the two very different elasticities measured in the model.

Since temporary changes in productivity change expected future profits from exporting little, few plants choose to change their exporting status. Thus, temporary changes in productivity tend to show up as price changes in the goods already being traded, and consumers substitute between goods at the low – "true" – elasticity specified as a parameter in the model. We call this increase (or decrease) in trade in goods that were already being traded an increase (or decrease) on the *intensive margin.* When all trade change occurs on the intensive margin, as it does in most trade models, estimating the elasticity from aggregate trade flows will recover the true elasticity. In our model there is a small number of plants entering and exiting the export market in response to productivity shocks, but the trade from these plants is small compared to aggregate trade. Thus, when we estimate the Armington elasticity in the model in response to productivity shocks, we find small values that are close to the true elasticity specified by the parameters. The estimated elasticity is not exactly equal to the true elasticity because the price indices constructed for use in time series estimation do not incorporate the changing set of goods being traded. As shown in Feenstra (1994), not including these newly traded goods imparts a bias to the indices. We discuss this mismeasurement and its implications in Section 5.

In contrast, a permanent change, such as lower tariffs, raises the profit from exporting in all states of nature and increases the expected future gain from exporting more than a temporary productivity shock. This induces more plants to begin exporting, resulting in large trade flows. We call the increase in trade flows from newly traded goods an increase on the *extensive margin*. The increase in the extensive margin is the key to understanding the large elasticities measured in response to trade liberalization. Following a decrease in tariffs, trade increases for two reasons: consumers buy more of the goods they already import, since the delivered price is now lower, and they buy new imports that were not previously available. The first kind of growth is intensive margin growth, while the second is extensive margin growth. If the change in aggregate trade is mistakenly assumed to be all intensive margin growth, the small changes in tariffs appear to induce large changes in imports on the intensive margin, which implies a large elasticity of substitution. When we make a similar measurement in our model after a decrease in tariffs, we find an Armington elasticity that is more than 3 times the true value. When we shut down the extensive margin in our model, however, our measured elasticity is the same as the true elasticity.

A simple example makes the point clear. Suppose a country is importing 10 dollars worth of goods. In response to a productivity shock, the price of imports relative to the price of the domestically produced goods falls by 10 percent, but no new plants enter the export market. If the true Armington elasticity is 2, then consumers import 12 dollars worth of goods. Measuring the elasticity, we see that trade increased by 20 percent in response to a 10 percent change in prices, and thus the elasticity is correctly measured as 2. Now suppose that a 10 percent tariff is removed, causing a 10 percent reduction in prices and, in addition, new plants enter the export market and export 4 dollars worth of goods. The change in export value is now the 2 dollars from the producers already exporting (intensive margin growth), and the 4 dollars from the new exporters (extensive margin growth). Viewed through a model without an extensive margin, the 60 percent increase in trade is all intensive margin trade, so the elasticity appears to be 6, which is 3 times the true elasticity.

There is growing evidence that the extensive margin, which drives the central result of this paper, is an important facet of the data. Empirically, Hillberry and McDaniel (2002) find evidence of extensive margin growth for the U.S. following the implementation of the North American Free Trade Agreement (NAFTA). Kehoe and Ruhl (2002) document extensive margin growth in a study of six different trade liberalizations and lay out a simple Ricardian model to highlight the forces at work. In a test of several plant-level models of exporter behavior, Bernard, Jensen and Schott (2003) study data on U.S. manufacturing plants and find strong evidence that new plants choose to enter the export market as trade costs fall.[1] In a cross sectional study, Hummels and Klenow (2002) find that larger and richer countries have larger extensive margins. The larger, richer, countries also tend to be the countries with the most liberalized trade policy.

Earlier theoretical work on export entry costs and uncertainty has focused on hysteresis in exporting. Baldwin (1988), Baldwin and Krugman (1989), and Dixit (1989) show how temporary increases in profitability (they considered exchange rates) could increase exports as plants enter the export market, but that high levels of exports would persist even as exchange rates appreciated. The plants that had incurred the sunk cost to export would continue to do so, even faced with a less attractive exchange rate, since they had already incurred the set-up cost.

In more recent work, Melitz (2003) incorporates export entry costs into a model of monopolistic competition, as in Dixit and Stiglitz (1977), in an environment with no aggregate uncertainty. His analysis focuses on the reaction of an industry to changes caused by trade liberalization. The model presented here has a structure of production similar to that in Melitz (2003), but focuses on the effects of industry structure on aggregate trade in the presence of aggregate uncertainty. In Choi (2003) export entry costs are imbedded in a standard international real business cycle model to study the effects of entry and exit on international correlations of consumption and output. The

[1] Bernard, Jensen et al. (2003) test the predictions of the plant-level models of Bernard, Eaton, Jensen and Kortum (2003), Melitz (2003), and Yeaple (2002).

model in Choi (2003) is conceptually similar to the one presented here, in that it involves aggregate uncertainty modeled as shocks to productivity. His analysis, however, is focused on the characteristics of international business cycles, while we are concerned with the different behavior of trade flows in both the short and the long run.

Empirical justification for the export entry costs that are central to this model, and the literature cited above, has come as the result of plant-level dynamic models of export entry decisions. The seminal work in this literature is Roberts and Tybout (1997), who develop an econometric model of a plant's decision to enter the export market. Using panel data on the Colombian manufacturing sector, they reject the null hypothesis that entry costs are unimportant. In a study of German plants, Bernard and Wagner (2001) find evidence of substantial sunk costs in export entry. Using a detailed data set on U.S. manufacturing plants, Bernard and Jensen (2001) find that export entry costs are significant and that plant heterogeneity is important in the export decision.

The next section reviews the evidence on the elasticity of substitution as estimated by different authors using different techniques. Section 3 presents a model in which plants choose whether or not to export in the presence of fixed costs and uncertainty about future productivity. Section 4 discusses computational issues and calibration of the model. Section 5 presents the model results, and Section 6 concludes.

2. Measuring Import Price Elasticity

In this section we review previous estimates of the Armington elasticity. Based on our hypothesis that high frequency changes in profitability lead to different responses than permanent changes, we divide the literature into two subsections. We find that studies that use high frequency data to estimate the Armington elasticity find low values, while studies that use cross sectional data, or data from trade liberalizations, tend to find much higher values.

In Armington (1969), it is posited that goods produced by different countries are intrinsically different goods. The utility that consumers derive from these nationally differentiated goods is represented by a constant elasticity of substitution utility function,

$$U = \left[\omega C_h^{\rho} + (1-\omega) C_f^{\rho} \right]^{\frac{1}{\rho}} \tag{1}$$

where C_k is the consumption of the good produced in country k. Maximizing this function with respect to the standard budget constraint, and rearranging the first order conditions yields

$$\frac{C_h}{C_f} = \left(\frac{P_f}{P_h} \frac{\omega}{1-\omega} \right)^{\sigma}, \tag{2}$$

where P_h is the price of the good produced in the home country and P_f is the price of the good produced in the foreign country. It is easy to see from (2) that $\sigma = 1/(1-\rho)$ is the elasticity of substitution between the goods. When the two goods are differentiated by country of origin, as in this case, this elasticity is commonly referred to as the *Armington elasticity*.

An alternative way of incorporating national product differentiation is to assume that countries produce intermediate goods that are combined to produce an aggregate consumption-investment good. In these models, the feasibility condition is

$$C + X = \left[\omega q_h^{\rho} + (1-\omega) q_f^{\rho} \right]^{\frac{1}{\rho}}, \tag{3}$$

where C is consumption, X is investment, q_h is the intermediate good produced in the home country, and q_f is the intermediate good produced in the foreign country. When using this specification, the constant elasticity function on the right-hand side of (3) is commonly called the *Armington aggregator*. Minimizing the cost of producing one unit of the aggregate good implies the same first order condition as (2).

In trade models that feature imperfect competition and differentiated goods, such as those in Helpman and Krugman (1985), consumers are frequently modeled as having preferences over varieties of goods within an industry also featuring a constant elasticity of substitution,

$$U_j = \left[\omega \int_{\imath \in \mathrm{I}_h} c_{h,j}(\imath)^{\rho} \, d\imath + (1-\omega) \int_{\imath \in \mathrm{I}_f} c_{f,j}(\imath)^{\rho} \, d\imath \right]^{\frac{1}{\rho}}, \tag{4}$$

where $c_{h,j}(\imath)$ is consumption of domestically produced variety, $\imath$, and $c_{f,j}(\imath)$ is consumption of the foreign produced variety in industry j. The set I_k is the set of varieties that the consumer has available for purchase. Maximizing (4) subject to a standard budget constraint yields the familiar condition,

$$\frac{c_{h,j}(\iota)}{c_{f,j}(\iota')} = \left(\frac{p_{f,j}(\iota')}{p_{h,j}(\iota)} \frac{\omega}{1-\omega} \right)^{\sigma}, \tag{5}$$

where $p_{f,j}(\iota')$ and $p_{h,j}(\iota)$ are, respectively, the price of the foreign and home goods. For example, $c_{h,j}(\iota)$ could be a shirt made domestically, while $c_{f,j}(\iota')$ is a shirt made abroad and imported. The empirical literature we survey below uses data collected at a level of aggregation higher than the "variety" level we model in (5). Thus, it is useful to write the first order condition in terms of industry level aggregate quantities and prices,

$$\frac{C_{h,j}}{C_{f,j}} = \left(\frac{P_{f,j}}{P_{h,j}} \frac{\omega}{1-\omega} \right)^{\sigma} \tag{6}$$

where $C_{h,j}$ is the aggregate amount of domestic consumption of goods in industry j and $P_{h,j}$ is an aggregate price index for the domestically produced goods. $C_{f,j}$ and $P_{f,j}$ are similarly defined for the goods imported into the home country.

Note that in contrast to the Armington model, goods in this specification are not necessarily different because they are made in different countries. Here goods are different by their very nature; these are the "differentiated goods" found in models of monopolistic competition such as Dixit and Stiglitz (1977). That the goods are made in different countries matters only to the consumer through the home bias parameter, ω. What is the same about the two models, however, is the implication of the first order conditions, (2) and (6). These equations are the basis for the estimation of Armington elasticities.

High Frequency Estimates

We begin by reviewing the Armington elasticities used in the IRBC literature. Commonly, the parameters of real business cycle models are not estimated, but are calibrated in the tradition of Kydland and Prescott (1982). In calibration, parameters are chosen so that the variables in the model's deterministic equilibrium match the average values of their counterparts in the data. Calibrated models are then used to produce simulated time series, and the second moments of the simulated data are compared to those in the real data. The models are usually subjected to sensitivity analysis, in which a single parameter is varied and the resulting changes to the model's results are analyzed.

The Armington elasticities generally used in the IRBC literature range from 0.5 to 2.0. In Backus, Kehoe et al. (1994) each country produces a tradable intermediate good which is combined using an Armington aggregator, as in (3), to produce an aggregate consumption-investment good. The authors' baseline choice of the Armington elasticity is 1.5, but they perform sensitivity analysis to this parameter. They find that the model with a smaller elasticity (0.5) can better account for the volatility of the terms of trade and the negative correlation between the terms of trade and the trade balance, then can the model with a larger (2.5) elasticity. Heathcote and Perri (2002) use a similar two-intermediate-goods environment to study business cycles under different degrees of financial market completeness. Beginning with a baseline value of 1.0, they find that the volatility of the terms of trade, the cross-country correlation of investment, and the cross-country correlation of consumption and output are closer to those in the data when lower values of the elasticity are used.

In addition to the low elasticities needed for IRBC models to match the features of the high frequency data, low elasticities are also found when they are directly estimated from high frequency data. The estimating equations are derived from the first order conditions, such as those in (2) or (6). Taking the logarithm of (6) yields the basic equation estimated by several authors,

$$\log\left(C_{f,jt}/C_{h,jt}\right) = \sigma_j \log\left(P_{h,jt}/P_{f,jt}\right) + \varepsilon_{jt}, \tag{7}$$

where $C_{f,jt}$ is the real quantity of imports in industry j, $C_{h,jt}$ is the real consumption of domestically produced goods in industry j, and $P_{h,jt}$ and $P_{f,jt}$ are price indices for domestic sales and imports. To estimate this equation, quarterly data on imports and domestic consumption of the industry's good, as well as data on the relative prices is collected. Typically, the price data take the form of a unit price index for imports, and a producer price index for domestic goods. Reinert and Roland-Holst (1992) estimate an equation similar to (7) for 163 industries and find elasticities that range from 0.02 to 3.49, with an average value of 0.91. Blonigen and Wilson (1999) estimate elasticities for 146 sectors, and find an average elasticity of 0.81, with a maximum value of 3.52. Similar estimates are found in Reinert and Shiells (1993) and in the short run elasticities reported in Gallaway, McDaniel and Rivera (2003). The elasticities found by estimating (7) on high frequency time series data are fairly robust. Adjustments have been made for, among other things, serially correlated errors, differing levels of aggregation, and seasonal effects. The elasticities estimated using these expanded techniques still find low values.

What is clear from these two strands of literature is that the high frequency fluctuations in prices and quantities seem to imply that the Armington elasticity is small. We now turn to the estimates of the Armington elasticity that use cross sectional data, or data from trade liberalizations.

Trade Liberalization and Cross Sectional Estimates

In this subsection we consider estimates of the Armington elasticity that are derived from cross sectional studies or the response of trade flows to changes in tariffs. In contrast to the high frequency data used in the studies above, the main sources of variation considered in these studies are permanent in nature. In the long run time series studies of specific trade liberalization episodes, the trade liberalization being analyzed is rarely undertaken as a temporary policy to be reversed later. In the cross section regressions, cross-country variation in trade barriers tends to be permanent. Though per-mile transportation costs may be falling, distances between countries are not; the relative cost of importing a good from two different countries stays about the same. The heterogeneity in bilateral trade policy, which is also an important source of variation, may change, but again, the changes are likely permanent.

The large increase in trade that follows trade liberalizations, such as the NAFTA, point to Armington elasticities much higher than those found in the above section. These elasticities can be computed by comparing the change in imports and domestic consumption over two points in time, typically a long span as liberalization is usually a gradual process, and the change in tariffs. The simple calculation is

$$\sigma = -\frac{\%\,\text{change } \; C_f / C_h}{\text{change } \; \tau}, \tag{8}$$

where τ is the *ad valorem* tariff rate. Calculating the elasticity this way assumes that the change in tariff is the only change in relative prices. Data for the NAFTA countries are presented in Table 1. The trade and production data are for the agriculture, mining, and manufacturing sectors. Domestic consumption is calculated by subtracting total exports from total production. Since trade data are collected as value data (not value added) gross output is the proper measure of production. The imports measured in this table are only imports from the NAFTA partner countries, and the tariff measures are simple averages based on Cox (1994). The elasticities implied by these data range from 7 to 16, which are about 3 to 5 times larger than those found in the high frequency studies. The values from this simple calculation line up well with the estimates derived from econometric models of trade liberalization. Clausing (2001) finds a price elasticity of 9.6 in a study of the Canada-U.S. Free Trade Agreement.[2] Head and Ries (2001) find elasticities that range from 7.9 to 11.4 in a regression relating trade shares to both tariff and non-tariff barriers between Canada and the United States. In a detailed study of U.S. trade that features data on thousands of goods, Romalis (2002) estimates elasticities that range between 4 and 13. These estimates are based on the substantial variation in tariff rates across partners and goods that is a feature of the disaggregated data.

Further evidence of high Armington elasticities can be found in cross sectional studies. In a model of economic geography with explicit consideration of transportation costs, Hummels (2001) uses data from Argentina, Brazil, Chile, New Zealand, Paraguay, and the United States to estimate the elasticity of substitution between varieties in a model with preferences similar to (4). The estimation produces elasticities ranging from 3 to 8. Baier and Bergstrand (2001) estimate a panel model over 16 OECD countries and find that trade flows are about 16 times more responsive to changes in tariffs than changes in the relative prices. In a cross section regression involving 19 countries and 50 goods, Eaton and Kortum (2002) estimate a parameter that corresponds closely to the

[2] In models based on (4), the elasticity of substitution and the price elasticity of demand are equal. In models with a finite number of varieties the price elasticity converges to the elasticity of substitution as the number of varieties approaches infinity.

elasticity of substitution. The authors find an elasticity of about 8, using both ordinary least squares and method-of-moments estimators. In a “calibration-as-estimation” exercise, Anderson, Balistreri, Fox and Hillberry (2002) find that the Armington elasticities needed to match the world bilateral trade pattern are, on average, 17.

In this section we have seen how estimates of the Armington elasticity differ considerably when the underlying source of variation differs. Elasticities measured by studying the high frequency changes in prices and quantities are capturing responses to temporary shocks, and range between 0.2 and 3.5. Estimates based on changes in trade policy, other trade costs, or cross sectional variation are capturing the response of trade flows to more permanent factors, and range between 6 and 15. The large differences in these estimates strongly suggest that trade flows are more sensitive to permanent changes then to temporary shocks.

In the next section a model is specified that produces temporary changes through productivity shocks and permanent changes through trade policy. It is then possible to study the effects of temporary and permanent changes on trade flows.

3. Model

The model is designed to incorporate the major elements of the standard applied GE models into an environment with aggregate uncertainty as in IRBC models. Consumers derive utility from consuming differentiated goods sold in markets characterized by monopolistic competition. The economy departs from the standard framework by requiring plants to pay a one-time entry cost before exporting and by subjecting the economy to aggregate productivity shocks. This structure of production is similar to Melitz (2003), but here the economy is subject to uncertainty about future productivity. These elements create an economy in which plants respond differently to temporary shocks to productivity, than they do to permanent changes in policy.

The economy consists of two countries, denoted as home (h) and foreign (f). Each country is endowed with $\overline{L}_k$ units of labor, which is the only factor of production. Each country produces two types of goods: a non-traded, homogeneous good, q, which is sold in a competitive market, and a continuum of differentiated goods, indexed by ι, which are produced by monopolistically competitive plants. A differentiated good produced in country h and consumed in country f is denoted by $c_h^f(\iota)$. For clarity, only the home country variables and maximization problems are described. The foreign country faces analogous problems.

At each date t, one of H possible events, η_t, occurs. Each event is associated with a vector of economy-wide productivity shocks, $z_t = \left(z_h(\eta_t), z_f(\eta_t)\right)$, and the initial event η_0 is given as an initial condition. We assume η follows a stationary first-order Markov process with transition matrix Λ. An element of Λ, the probability of event η' happening tomorrow, given event η happened today, is $\lambda_{\eta\eta'} = \mathrm{pr}\left(\eta_{t+1} = \eta' \middle| \eta_t = \eta\right)$. Define a state, s, as a history of events, $s = \left(\eta_0, \eta_1, ..., \eta_{t(s)}\right)$, where the function $t(\bullet)$ returns the date in which state s occurs. The probability of being in state s is then defined as $\lambda_s = \lambda_{\eta_0\eta_1}\lambda_{\eta_1\eta_2}\cdots\lambda_{\eta_{t(s)-1}\eta_{t(s)}}$.

A key feature of this economy is that plants are heterogeneous in productivity and in the entry cost that must be paid in order to export. A plant is indexed by its idiosyncratic productivity, $\phi \in \mathcal{F}$, its entry cost, $\kappa \in \mathcal{K}$, and whether or not it is an exporter. A plant is an exporter only if it has paid the cost κ. The distribution of plants that begin the period as exporters is represented by a measure over (ϕ, κ), $\mu_{hx}(\phi, \kappa)$, which has support $\mathcal{F} \times \mathcal{K}$. Plants that begin the period as non-exporters are tracked by a similar measure, $\mu_{hd}(\phi, \kappa)$. The aggregate state variables in this economy are $\eta, \mu_{hd}, \mu_{hx}, \mu_{fd}, \mu_{fx}$. Denoting the vector of plant distributions as $\mu = \left(\mu_{hd}, \mu_{hx}, \mu_{fd}, \mu_{fx}\right)$, the aggregate state variables can be represented by (η, μ). The agents in the economy take as given the laws of motion for μ,

$$\mu'_{hd}(\phi, \kappa) = \mathrm{M}_{hd}(\phi, \kappa, \eta, \mu)$$

$$\mu'_{hx}(\phi, \kappa) = \mathrm{M}_{hx}(\phi, \kappa, \eta, \mu),$$

with similar laws of motion for the distributions in the foreign country. For notational simplicity, define $\mathrm{M}(\ )$ as the law of motion over the distribution of all plants,

$$\mu' = \mathrm{M}(\eta, \mu). \tag{9}$$

Households

The economy is populated by a unit measure of identical households. It is assumed that each household owns an equal share in all plants in operation. Given these assumptions, and the homotheticity of preferences, the households can be represented by a stand-in household. Households, as owners of the plants, decide when to pay the entry cost for a plant to become an exporter. This decision amounts to an investment choice, since households are trading off consumption today, in the form of the entry costs, for more consumption tomorrow, in the form of higher profits. Rather than have the households solve this problem explicitly, we choose to set up the plants' problems such that plants' decisions are aligned with the households' interests. We discuss this further when we construct the plant's choice problem. Given this specification, the stand-in household in each country chooses consumption to maximize utility, and inelastically supplies labor to plants. The stand-in household's problem can be written as

$$\max_{c_h^h(\iota),c_f^h(\iota),q} \gamma \log\left(C_h\right)+\left(1-\gamma\right)\log\left(q_h\right), \tag{10}$$

where C_h is an aggregate of the differentiated goods that are combined using a constant elasticity of substitution form similar to (4),

$$C_h=\left[\omega \int_{\iota\in \mathrm{I}_h^h(\mu)} c_h^h\left(\iota\right)^{\rho} d\iota+\left(1-\omega\right)\int_{\iota\in \mathrm{I}_f^h(\mu)} c_f^h\left(\iota\right)^{\rho} d\iota\right]^{\frac{1}{\rho}}.$$

The household's maximization is subject to the budget constraint,

$$\int_{\iota\in \mathrm{I}_h^h(\mu)} p_h^h\left(\iota;\eta,\mu\right)c_h^h\left(\iota\right)d\iota+\int_{\iota\in \mathrm{I}_f^h(\mu)} p_f^h\left(\iota;\eta,\mu\right)\left(1+\tau\right)c_f^h\left(\iota\right)d\iota+p_{hq}\left(\eta,\mu\right)q_h = \overline{L}_h+\Pi_h\left(\eta,\mu\right),$$

where $c_k^j\left(\iota\right)$ is the differentiated good ι made in country k and consumed in country j. Similarly, $p_k^j\left(\iota\right)$ is the price of differentiated good ι made in country k and consumed in country j. The price of the non-traded good in the home country is p_{hq}. Imports are subject to an ad valorem tariff, τ, which is modeled as an iceberg transportation cost for simplicity. Aggregate profits, Π_h, are returned to the consumer, and the wage is normalized to 1.

The solution to the consumer's problem features the familiar demand function,

$$\tilde{c}_h^h\left(p_h^h(\iota);\iota,\eta,\mu\right)=\left(\frac{P_h(\eta,\mu)}{p_h^h(\iota;\eta,\mu)}\right)^{\frac{1}{1-\rho}}C_h(\eta,\mu), \tag{11}$$

for a differentiated good produced in the home country , and

$$\tilde{c}_f^h\left(p_f^h(\iota);\iota,\eta,\mu\right)=\left(\frac{P_h(\eta,\mu)}{p_f^h(\iota;\eta,\mu)(1+\tau)}\right)^{\frac{1}{1-\rho}}C_h(\eta,\mu) \tag{12}$$

for a good produced in the foreign country. The aggregate price index for differentiated goods in this economy has the standard form

$$P_h(\eta,\mu)=\left[\omega^{\frac{1}{1-\rho}}\int_{\iota\in I_h^h(\mu)}p_h^h(\iota;\eta,\mu)^{\frac{\rho}{\rho-1}}d\iota+(1-\omega)^{\frac{1}{1-\rho}}(1+\tau)^{\frac{\rho}{\rho-1}}\int_{\iota\in I_f^h(\mu)}p_f^h(\iota;\eta,\mu)^{\frac{\rho}{\rho-1}}d\iota\right]^{\frac{\rho-1}{\rho}}.$$

Differentiated Goods Plants

There is a continuum of differentiated goods plants indexed by their idiosyncratic productivities, ϕ, and entry costs, κ. A plant's marginal cost of production consists of two parts: the idiosyncratic, non-stochastic productivity, ϕ, and the economy wide, stochastic shock to productivity, $z_h(\eta)$. The production function for a plant of type (ϕ,κ) in aggregate state (η,μ) is

$$y(\phi,\kappa,\eta,\mu)=z_h(\eta)\phi l. \tag{13}$$

When a plant chooses to begin exporting, it must pay the entry cost, κ. This cost must be paid before the realization of η and can not be recovered afterward. After paying this entry cost, the plant faces no further costs associated with exporting.

An incumbent plant enters the period as either an exporter or a non-exporter. After productivity is revealed, plants choose how much labor to hire and how much to produce. After production, a mass of entrants, ν, arrives who have not paid the fixed cost to export. The joint distribution of idiosyncratic productivity and entry costs over these entrants has p.d.f. Φ. At the end of the period, non-exporters decide whether to continue as non-exporters, or to begin exporting and pay κ. In addition, plants face an exogenous probability of death, δ. Exporting decisions, along with the exogenous death of plants, determine the next period's distributions of exporters and non-exporters, $\mu_{hd}{}'$ and $\mu_{hx}{}'$. The timing of decisions, and the evolution of the distributions over plants is displayed in Figure 1.

The plant faces two problems. The first is a static maximization of period profits. The second is the dynamic decision of exporter status. We turn to the static problem first. Plants are monopolistic competitors who choose prices to maximize profits, taken as given the aggregate price index and the wage. The plant realizes, however, that the consumer's demand is downward sloping, and thus the demand functions defined in (11) and (12) appear in the plant's problem. For clarity in the plant's dynamic problem, it is useful to define the value of maximized profits from selling domestically for a plant of type (ϕ,κ) with aggregate state (η,μ) as

$$\begin{aligned}\pi_{hd}(\phi,\kappa,\eta,\mu) &= \max_{p_h^h,\, l} \tilde{c}_h^h\left(p_h^h;\phi,\kappa,\eta,\mu\right)p_h^h - l \\ \text{s.t.} \quad z_h(\eta)\phi l &= \tilde{c}_h^h\left(p_h^h;\phi,\kappa,\eta,\mu\right)\end{aligned} \tag{14}$$

and maximized profits from exporting as

$$\begin{aligned}\pi_{hx}(\phi,\kappa,\eta,\mu) &= \max_{p_h^f,\, l} \tilde{c}_h^f\left(p_h^f;\phi,\kappa,\eta,\mu\right)p_h^f - l \\ \text{s.t.} \quad z_h(\eta)\phi l &= \tilde{c}_h^f\left(p_h^f;\phi,\kappa,\eta,\mu\right).\end{aligned} \tag{15}$$

Optimization implies that plants set prices as a constant markup over marginal costs, and produces the price functions

$$p_h^h(\phi,\kappa,\eta,\mu) = p_h^f(\phi,\kappa,\eta,\mu) = \frac{1}{\rho\phi z_h(\eta)}, \tag{16}$$

as well as labor demand functions, $l_h^h(\phi,\kappa,\eta,\mu)$ and $l_h^f(\phi,\kappa,\eta,\mu)$. Having defined the maximized values from the static problem, the plant's dynamic problem is reduced to choosing only exporting decisions. An exporter's value function is defined by

$$\begin{aligned}&V_{hx}(\phi,\kappa,\eta,\mu) = \\ &\quad \xi(\eta,\mu)\left(\pi_{hd}(\phi,\kappa,\eta,\mu) + \pi_{hx}(\phi,\kappa,\eta,\mu)\right) + (1-\delta)\beta\sum_{\eta'} V_{hx}(\phi,\kappa,\eta',\mu')\lambda_{\eta\eta'} \\ &\quad s.t. \qquad \mu' = \mathrm{M}(\eta,\mu)\end{aligned} \tag{17}$$

where $\mathrm{M}(\ )$ is the law of motion for the aggregate state variable μ. The term $(1-\delta)$ reflects the probability of exogenous death that the plant faces. A non-exporter's problem can be written as

$$\begin{aligned} &V_{hd}(\phi,\kappa,\eta,\mu) = \xi(\eta,\mu)\pi_d(\phi,\kappa,\eta,\mu) \\ &+(1-\delta)\beta \max\left\{\sum_{\eta'} V_{hx}(\phi,\kappa,\eta',\mu')\lambda_{\eta\eta'} - \frac{\xi(\eta,\mu)\kappa}{(1-\delta)\beta},\ \sum_{\eta'} V_{hd}(\phi,\kappa,\eta',\mu')\lambda_{\eta\eta'}\right\} \\ &\qquad s.t. \qquad \mu' = \mathrm{M}(\eta,\mu) \end{aligned} \tag{18}$$

The two terms in the maximization correspond, respectively, to the expected future profit from becoming an exporter, and the expected future profit from continuing as a non-exporter. This choice is the crucial one for the results presented here. The small, temporary productivity shocks change the future expected profits from exporting little, and thus few plants are willing to sink the cost of becoming exporters. A permanent change, such as a tariff decrease, increases a plant's profit in every realization of η, and thus has a larger effect on the expected future profits from exporting. This larger effect induces more plants to enter the export market, and increases the amount of goods being traded. Thus, in this model, permanent changes in tariffs have larger impacts on trade than do temporary shocks to productivity. Solving the non-exporter's problem yields the decision rule over next period's export status, $d_h(\phi,\kappa,\eta,\mu)$, which is equal to 1 if the plant chooses to export next period and is equal to 0 if the plant chooses to continue as a non-exporter in the next period.

Notice that in (17) and (18), current profits are valued at ξ, the multiplier on the consumer's budget constraint. This is because, with complete asset markets, plants discount future profits at rate $(1+r_{\eta\eta'})^{-1}$, which in equilibrium is

$$\frac{1}{1+r_{\eta\eta'}} = \beta\frac{\xi(\eta',\mu')}{\xi(\eta,\mu)}.$$

So, when plants value current profits at ξ and discount future profits by the household's discount factor, β, the plants' programming problems in (17) and (18) is equivalent to the one in which the stochastic discount factor is explicitly specified.

The non-traded good, q, is produced by a constant returns to scale plant and is sold in a competitive market. The plant's problem is

$$\begin{aligned}&\max_{l}\ p_{hq}(\eta,\mu)q_h - l\\ &\text{s.t.}\ \ q_h = z_h(\eta)l\end{aligned}\ , \tag{19}$$

where the price of good q_h is p_{hq}. The model is closed by assuming that trade is balanced each period,

$$\int_{\iota\in \mathrm{I}_h^f(\mu)} p_h^f(\iota;\eta,\mu)c_h^f(\iota;\eta,\mu)d\iota \ - \int_{\iota\in \mathrm{I}_f^h(\mu)} p_f^h(\iota;\eta,\mu)c_f^h(\iota;\eta,\mu)d\iota = 0\,. \tag{20}$$

Equilibrium Defined

Equilibrium is defined recursively. For simplicity, only the home country equilibrium objects are enumerated; the agents in the foreign country solve problems analogous to those in the home country, and have the corresponding decision rules.

A *recursive equilibrium* in this economy is value functions $V_{hx}(\phi,\kappa,\eta,\mu)$ and $V_{hd}(\phi,\kappa,\eta,\mu)$ for each plant type $(\phi,\kappa)\in\mathcal{F}\times\mathcal{K}$, decision rules, $d_h(\phi,\kappa,\eta,\mu)$, $l_h^h(\phi,\kappa,\eta,\mu)$, $l_h^f(\phi,\kappa,\eta,\mu)$, $p_h^h(\phi,\kappa,\eta,\mu)$, and $p_h^f(\phi,\kappa,\eta,\mu)$ for each plant type $(\phi,\kappa)\in\mathcal{F}\times\mathcal{K}$, the sets of goods available for consumption $\mathrm{I}_h^h(\mu)$ and $\mathrm{I}_f^h(\mu)$, decision rules $c_h^h(\iota,\eta,\mu)$ for each $\iota\in\mathrm{I}_h^h(\mu)$, $c_h^f(\iota,\eta,\mu)$ for each $\iota\in\mathrm{I}_f^h(\mu)$, and $q_h(\eta,\mu)$ for the household, the price function, $p_{hq}(\eta,\mu)$, and the laws of motion for the distribution of plants, $\mathrm{M}_{hd}(\eta,\mu)$ and $\mathrm{M}_{hx}(\eta,\mu)$, such that these functions satisfy:

1. the household's problem, (10),
2. the plant's problems for each type $(\phi,\kappa)\in\mathcal{F}\times\mathcal{K}$, (17) and (18),
3. the consistency of aggregate and individual decisions,

   $\mathrm{M}_{hx}(\phi,\kappa)=$
   $\left[\mu_{hx}(\phi,\kappa,\eta,\mu)+\left(\nu\,\Phi(\phi,\kappa)+\mu_{hd}(\phi,\kappa,\eta,\mu)\right)d_h(\phi,\kappa,\eta,\mu)\right](1-\delta)$

   and $\mathrm{M}_{hd}(\phi,\kappa)=\left(1-d_h(\phi,\kappa,\eta,\mu)\right)\left[\mu_{hd}(\phi,\kappa)+\nu\,\Phi(\phi,\kappa)\right](1-\delta)$,

   for all $(\phi,\kappa)\in\mathcal{F}\times\mathcal{K}$,

4. the goods and labor market clearing conditions,
5. the balanced trade condition, (20).

Analysis of Equilibrium

The equilibrium in this economy is characterized by a cut-off rule for each type of entry cost, in each country. The rule is a function from the state space to the space of productivity types. The cut-off rule for entry cost κ when the aggregate state is (η, μ) is denoted $\hat{\phi}_\kappa(\eta, \mu)$. This is the productivity level at which the non-exporting plant with idiosyncratic productivity $\hat{\phi}_\kappa(\eta, \mu)$ and entry cost κ is exactly indifferent between continuing as a non-exporter, and entering the export market when the aggregate state is (η, μ). Thus, the cutoff productivity satisfies:

$$\xi(\eta,\mu)\kappa = (1-\delta)\beta\left[\sum_{\eta'} V_{hx}\left(\hat{\phi}_\kappa(\eta,\mu),\kappa,\eta',\mu'\right)\lambda_{\eta\eta'} - \sum_{\eta'} V_{hd}\left(\hat{\phi}_\kappa(\eta,\mu),\kappa,\eta',\mu'\right)\lambda_{\eta\eta'}\right]. \tag{21}$$

The right-hand-side of (21) is the discounted expected future gain from exporting: it is the difference in future profits if the plant exports rather than only selling domestically. The left-hand-side of (21) is the cost of entering the export market. Plants with productivity below $\hat{\phi}_\kappa(\eta, \mu)$ sell only to the domestic market, while plants with productivity above $\hat{\phi}_\kappa(\eta, \mu)$ sell to both the domestic and the exports markets. As the aggregate state of the economy changes, the expected future profits of the plants change, which shifts the cutoff plants, implying the entry and exit of plants in the export market.

Plants moving into and out of exporting play the crucial role in explaining the different responses of aggregate exports to changes in trade policy and productivity. If the shocks to productivity are small or not very persistent, there will be little movement of plants into or out of exporting. Thus, productivity shocks induce incumbent exporting plants to change prices, and consumers react to the change in prices by substituting according to the elasticity of substitution implied by ρ. This is growth on the intensive margin. With only a few plants entering the export market, extensive margin growth is small. If transitory shocks are the dominant source of the variation measured in time series regressions such as Reinert and Roland-Holst (1992) or Blonigen and Wilson (1999), they would operate mostly on the intensive margin, and the low elasticities estimated in these studies would reflect a low true elasticity of substitution.

A permanent change, such as the lowering of tariffs that accompanies a free trade agreement, increases the profit from exporting for all realizations of future productivity, and thus has a larger impact on the expected future profits from exporting. The larger increase in expected future profits induces some plants that were only selling in the domestic market to enter the export market. These newly traded goods create growth on the extensive margin. These extensive margin effects increase trade flows by more than that implied by the fall in tariffs and the elasticity of substitution. This mechanism contributes to the seemingly large estimated elasticities found when regressing trade volumes on tariff changes or transportation costs as in Clausing (2001) or Hummels (2001).

In Figure 2 we plot the steady state expected gain from exporting for different productivity types and the value of a particular entry cost. The intersection of these two lines defines $\hat{\phi}_{\kappa,ss}$: the steady state cutoff productivity for plants with entry cost κ. This is the graphical interpretation of (21). All the plants with productivity less than $\hat{\phi}_{\kappa,ss}$ do not export in the steady state, while plants with productivity greater than $\hat{\phi}_{\kappa,ss}$ do. To see how the model responds to temporary shocks compared to permanent changes, consider two scenarios. In the first scenario, the economy is subject to productivity shocks that can take the value of either plus 2 percent or minus 2 percent of the mean productivity. The process governing these shocks has a one period autocorrelation of 0.90. Figure 2 shows how the expected future value of exporting shifts up when the economy is subjected to a 2 percent positive shock in each country. This shift lowers the cutoff productivity to $\hat{\phi}_{\kappa,\mathrm{t}}$ and the plants between $\hat{\phi}_{\kappa,\mathrm{t}}$ and $\hat{\phi}_{\kappa,ss}$ enter the export market. A typical conditional distribution over plant productivity types is shown in Figure 3. As the cutoff productivity moves to the left, the amount of new plants entering can be inferred from the slope of the distribution. If the slope is steep, the mass of plants entering from even a small change in the cutoff productivity can be large. The slope of this distribution near the marginal exporter is a function of the model's parameters, which will be determined in the calibration.

Now consider the same economy, starting again from the steady state, and this time subjecting the economy to a *permanent* 2 percent positive shock in each country. As can be seen in Figure 2, the permanent 2 percent shock shifts the expected future value of exporting up by much more than the persistent, but not permanent, shock. The cutoff plant is now $\hat{\phi}_{\kappa,\mathrm{p}}$, and the plants between $\hat{\phi}_{\kappa,\mathrm{p}}$ and $\hat{\phi}_{\kappa,ss}$ begin exporting. The larger shift in expected future profits from exporting from the permanent shock leads to a greater number of plants beginning to export. In Figure 3 we can see how the large shift in the cutoff productivity leads to a greater number of new entrants than in the case of the temporary shock.

It is the difference in the number of new exporters under the two scenarios that changes the implied Armington elasticity. If no new plants enter the export market in response to the productivity shock, consumers only substitute between the goods already being imported and domestically produced goods at the rate σ. This is trade growth on the intensive margin. In this case, the measured Armington elasticity is exactly $1/(1-\rho)$. If plants enter the export market in response to a change in future profits, trade flows increase from the trade of the new exporters as well as the increase in trade of the continuing exporters. If newly traded goods are not accounted for, this extensive margin growth shows up in the trade aggregates as intensive margin growth, and the response of imports appears very large. The large response in trade to the change in tariffs results in a large estimate of the Armington elasticity.

4. Calibration and Computation

In order to solve the plants' problems in (14) and (15), the plants need to know the aggregate price index, P_h, and the consumption of the composite good, C_h. These values depend on the state variables for this economy, which include the distributions over plants, μ_{hd} and μ_{hx}, and their counterparts for the plants in the foreign country. Due to the high dimensionality of these objects, standard computational techniques are not applicable. The approach taken here is to specify a rule that the plants use to predict the necessary aggregate variables. In the rule used below, the aggregate variables in the plant's problems are conditional only on the state of aggregate technology. The values of these variables are chosen so that they are the conditional mean values of the stochastic process over their counterparts in equilibrium. This approach is similar to that used in Krusell and Smith (1998) and Golosov and Lucas (2003).

Algorithmically, the model is solved as follows:

1. Choose values for $P_h^n(\eta_i), P_f^n(\eta_i), C_h^n(\eta_i), C_f^n(\eta_i),\ \ i = 0,...,\mathrm{H}$.
2. Given these values, solve for the exporter and non-exporter value functions and export decision rules.
3. For a given $\mu_{hd}, \mu_{hx}, \mu_{fd}, \mu_{fx}$, and η simulate the economy for 2000 periods using the policy functions found in 2., but computing equilibrium prices in each period. This produces a series of data: $\{P_{ht}, P_{ft}, C_{ht}, C_{ft}\}_{t=0,...,2000}$.
4. To ensure no dependence on initial conditions, throw away first 500 observations and calculate the means of $\{P_{ht}, P_{ft}, C_{ht}, C_{ft}\}_{t=500,...,2000}$, conditional on η_t.
5. If the mean values of the price levels and aggregate quantities are the same as the guess, $P_h^n(\eta_i), P_f^n(\eta_i), C_h^n(\eta_i), C_f^n(\eta_i),\ \ i = 0,...,\mathrm{H}$, stop. If not, adjust the prices and quantities to form the next guess, $P_h^{n+1}(\eta_i), P_f^{n+1}(\eta_i), C_h^{n+1}(\eta_i), C_f^{n+1}(\eta_i),\ \ i = 0,...,\mathrm{H}$, and repeat steps 2 through 5 until the guess and the observed conditional means derived from the guess are the same.

A second computational difficulty involves the continuum of plants. Since the functional equations in (17) and (18) do not have analytic solutions, these equations need to be solved computationally for each type. To facilitate this we use a finite number of plant types, and thus have to solve for only a finite number of value and policy functions. A drawback to this method is that the discrete jump in plant types does not guarantee a marginal plant type. Since all plants of the same type behave the same, the decision of one type of plant to enter the export market implies that the entire mass of plants of this type enter, which may have an effect on the aggregate variables. In particular, it may be the case that a type of plant would want to enter the export market for a given set of aggregate prices, but having entered, the resulting change in aggregate prices would be such that the plant would now choose to exit. This behavior would cause the algorithm outlined above to cycle, rather than converge. We minimize this problem by choosing a large number of types, and forcing any vacillating plant to not enter. As the number of types grows large, the error induced by a plant like this becomes very small.[3]

Calibration

The model is calibrated to match the United States and a symmetric partner country in 1987, which is before both the Canada-U.S. Free Trade Agreement and the NAFTA. We choose parameters such that the model's deterministic steady state displays the key aggregate and plant level patterns from the data. We set the ad valorem tariff rate to be 15 percent, and choose ρ to be 0.50, which implies a value of 2.00 for the true elasticity of substitution. This choice of elasticity is in the range found in the time series estimations, and close to those used in the international real business cycle literature.

The model period is one quarter, so setting β equal to 0.99 implies and annual real interest rate of about 4.00 percent. We consider the traded goods sector in the model to be manufacturing, even though there is trade in primaries and services. We exclude primaries and services since the plant level data needed to calibrate the model is only available for manufacturing plants. This definition of the tradable goods sector implies that γ, the share of expenditures on manufactures, is 0.19. The parameter governing the exogenous death of plants is set so that 2.35 percent of jobs are lost to exiting plants per year. This is the average found in Davis, Haltiwanger and Schuh (1996) for the years 1973-1998. In the baseline calibration, the home bias parameter, ω, is set to 0.50, which implies no intrinsic reason for a home consumer to prefer goods made in the home country.

The parameters ν, the mass of entrants each period, $\bar{\kappa}$, the largest value of the entry cost, and $\bar{\phi}$, the largest value of the idiosyncratic productivity, along with the

[3] In the calibrated model, we use 24,000 plant types, which would result in *less than* 0.004 percent of plant types acting non-optimally.

distribution over entrants' productivities and entry costs, jointly determine the exporting and production structure of the traded goods sector.

The idiosyncratic productivity can take values in the interval $(0, \bar{\phi}]$ and the entry cost can take values in the interval $\left[0, \bar{\kappa}\right]$. The computational difficulty described above requires that we use a large number of discrete plant types. We choose to evenly divide the interval over productivities into 800 different types, and the interval over entry costs into 30 different types, for a total of 24,000 plant types. The shape of the distribution over entrants, Φ, is chosen so that the shape of the equilibrium distribution over plant sizes matches that for U.S. manufacturing plants. This requires a joint distribution over $\left(\phi, \kappa\right)$. To minimize the number of free parameters, the distribution over ϕ is assumed to be independent of the distribution over κ. The conditional measures have the forms:

$$\begin{aligned} \Phi\left(\phi \middle| \kappa\right) &= \frac{1}{\phi^{\theta_\phi}} \\ \Phi\left(\kappa \middle| \phi\right) &= \frac{1}{\left(\bar{\kappa} - \kappa\right)^{\theta_\kappa}} \end{aligned} \tag{22}$$

which are then normalized to be (discrete) probability distributions. The distribution over idiosyncratic productivity puts less weight on higher productivities. The distribution over entry costs puts less weight on lower fixed costs.

Given the functional form of Φ, the parameters θ_ϕ, θ_κ, ν, $\bar{\phi}$, and $\bar{\kappa}$ determine the amount of output that is exported, the average plant size, the largest plant in existence, the distribution of all plants by size, and the distribution of exporting plants by size. The average manufacturing plant size is about 65 employees, and the largest plant in operation has about 6,000 employees. Fitting the two plant size distributions using so few parameters is difficult. Figure 4 and Figure 5 display the relative success of the calibration. Figure 4 shows the distribution of all manufacturing plants, taken from the 1987 Census of Manufacturing, and the distribution of all plants in the calibrated model. Figure 5 shows the distribution of only manufacturing plants that export, also taken from the Census of Manufacturing, and the size distribution of exporting plants from the model. Both figures indicate that the model fits the data fairly well considering the small number of free parameters.

The final parameters to specify are the ones governing the exogenous productivity shocks. Aggregate productivity in each country is either $1-\varepsilon$, which we denote the low shock, or $1+\varepsilon$, which we denote the high shock. Each country has an identical symmetric transition matrix over its shock,

$$\Lambda_i = \begin{bmatrix} \bar{\lambda} & 1-\bar{\lambda} \\ 1-\bar{\lambda} & \bar{\lambda} \end{bmatrix} \tag{23}$$

which requires only one parameter, $\bar{\lambda}$. Assuming that the shocks to productivity in the two countries are independent, the process requires calibrating only these two parameters. We choose $\bar{\lambda}$ and ε so that the volatility and persistence of the Hodrick-Prescott filtered aggregate Solow residuals in the model match those in the U.S. data. The persistence of the Solow residuals is measured as the one period autocorrelation, and we impose that one country's productivity does not have any "spillover" effects on the other country's productivity. We choose to match a one period autocorrelation of 0.900, similar to the value of 0.906 used in Backus, Kehoe et al. (1992) and 0.910 in Heathcote and Perri (2003). The volatility of the Solow residuals is measured as the percent standard deviation of the Hodrick-Prescott filtered data. We set the standard deviation to 1.00 for the baseline calibration. Table 2 summarizes the calibration.

5. Model Results

In Section 3 we saw that *qualitatively* the model has different implications for temporary and permanent changes to exporting profitability, and that these different responses could lead to different estimates of the Armington elasticity. We now use the calibrated model to see if the exporting entry costs can *quantitatively* account for the different estimates of the Armington elasticity found in the literature.

High Frequency Estimates

To see if the model can account for the low estimates derived from the high-frequency data, we use the model to generate simulated time series data on the prices and quantities of domestic goods and imports. To accurately test the model we must construct measures from the simulated data that are consistent with the methods used in the empirical works cited above. To do so, we construct a Laspeyres price index for imports, as in Reinert and Roland-Holst (1992), which weights period t prices by the quantities imported in a chosen base period,

$$\hat{P}_{f,t}^{h} = \frac{\sum_{i \in \mathrm{I}_{f,0}^{h}} p_{f,t}^{h}(i) c_{f,0}^{h}(i)}{\sum_{i \in \mathrm{I}_{f,0}^{h}} p_{f,0}^{h}(i) c_{f,0}^{h}(i)}. \tag{24}$$

where $\mathrm{I}_{f,0}^{h}$ is the set of goods traded in the base period, $t = 0$.[4] A similar Laspeyres index for domestic consumption is constructed, and these indices are used to deflate expenditures on imports and domestically produced goods. The data are used to estimate the Armington elasticity using a specification similar to that found in the empirical literature,

$$\log\left(c_{f,t}^{h} / c_{h,t}^{h}\right) = \sigma \log\left(p_{h,t}^{h} / p_{f,t}^{h}\right) + \varepsilon_t. \tag{25}$$

The results of this estimation are presented in the column labeled "Laspeyres Prices" in Table 3. Using prices measured as in the data, the estimated Armington elasticity is 1.4, and the parameter is precisely estimated. This estimate fits well into the range of values estimated from time series data. As surveyed above, the time series estimates range from 0.2 to 3.5, with an average of about 1.0, a little lower than the elasticity estimated from the model. Our estimate is also close to the values used in the international business cycle literature; these studies commonly use Armington elasticities between 1.0 and 2.0.

In interpreting the results, it is important to notice that the estimated elasticity is about 30 percent lower than the true elasticity. In equilibrium, the consumer's first order conditions must hold with equality, which implies that the regression in (25) should fit exactly, and the elasticity of substitution should be 2.0. The equation estimated with the Laspeyres price indices, however, does not fit exactly because the price measures are not accounting for the changing set of goods being traded.[5] Not taking into account the changing set of goods results in significant measurement error and a lower estimated value for the Armington elasticity.

---

[4] In both the computational model and the data, there are only a finite number of goods being traded, so the formulas expressed in this section are also written in terms of a finite number of goods. The counterparts to these formulas for the model presented in Section 3, which features a continuum of goods, are the same except the summations are replaced by integrals.

[5] I am grateful to Erzo G.J. Luttmer for pointing out this feature of the model.

Decomposing the price index exposes the source of the measurement error. Using the expression for prices in (16), the price index in (24) can be written as

$$\hat{P}^h_{f,t} = \frac{w_{f,t}}{z_{f,t}} \left[\frac{\sum_{i \in \mathrm{I}^h_{f,0}} \frac{1}{\phi_i \rho} c^h_{f,0}(i)}{\sum_{i \in \mathrm{I}^h_{f,0}} \frac{w_{f,0}}{\phi_i \rho z_{f,0}} c^h_{f,0}(i)} \right]. \tag{26}$$

This splits the index into two parts. The first term on the right-hand side is the ratio of the wage and the aggregate productivity. This term is common to all plants, so we denote it the *aggregate component*. The second term depends on the set of goods being imported, so we denote it the *composition component.* As is clear from (24), the set of goods over which the index is computed is held fixed through time. As a practical matter, goods that become traded subsequent to the base year must be ignored: the quantities imported of "newly traded" goods are by definition zero in the base year. This formulation implies that any change in the composition of goods being imported will not be reflected in the price index. In fact, the composition component is constant across time. Thus, this index changes only with the aggregate component, and does not take into account the changing set of goods being imported.

In contrast to the Laspeyres index, the constant elasticity of substitution (CES) price index makes adjustments for the set of goods being traded. This index is exactly the cost of purchasing one unit of the composite import good,

$$P^h_{f,t} = \left[\sum_{i \in \mathrm{I}^h_{f,t}} \left(p^h_{f,t}(i) \right)^{\frac{\rho}{\rho-1}} \right]^{\frac{\rho-1}{\rho}}. \tag{27}$$

The summation in this index runs over all the goods being imported, incorporating the changing composition of goods. The domestic price index can similarly be defined. If we run the regression specified in (25) using the CES indices that allow for the changing types of goods being imported, we find the estimated elasticity is exactly 2.0. The regression results are in the column labeled "CES Prices" in Table 3. The regression's fit is perfect because the regression equation is exactly the first order condition from the consumer's problem. This equation must hold every period in equilibrium, so the implied elasticity from this regression is the one implied by the first order conditions, and the R-squared is 1.0.

To see why the mismeasurement implies lower estimates, substitute the equilibrium prices into (27) and arrange terms to form

$$P_{f,t}^{h} = \frac{w_{f,t}}{z_{f,t}} \left[\sum_{i \in \mathrm{I}_{f,t}^{h}} \frac{1}{\rho \phi_i}^{\frac{\rho}{\rho-1}} \right]^{\frac{\rho-1}{\rho}} . \tag{28}$$

As with the Laspeyres index, the CES price index can be split into an aggregate component and a composition component. The composition component in the CES index changes as the set of goods being imported changes. As more (fewer) goods are imported, this term decreases (increases), lowering (raising) the price index. Now compare the Laspeyres index in (26) to the CES index in (28). The indices have identical aggregate components, but the composition components differ. In the Laspeyres index, the composition component is constant, so the index moves only with the aggregate component, while the composition component in the CES index changes with the set of goods being imported. The divergence of the composition component in these two indices is the source of the measurement error. This kind of measurement error has been noted by Helkie and Hooper (1988), Krugman (1989), Feenstra (1994) and Feenstra and Shiells (1997) in the context of income elasticity estimates. They argued that the growing number of goods being imported by the U.S. is not reflected in the import price index, so the new imports lead to an increase in the value of imports but not in the price index. The rising value of imports is then attributed to the growth in U.S. income, implying a large value for the income elasticity of imports. In our model, goods both enter and exit the import market, and both situations lead to bias in the measurement of prices.

To illustrate the bias when goods exit the import market, consider an economy where the home country and foreign country are in high productivity states. If the home country's productivity falls, income in the home country falls, and the profit of a plant exporting to the home country decreases. These lower profits will make some of the newly born plants choose not to export – though they would have exported if productivity had stayed high in the home country. Plants of these same types that were already alive and exporting die off at the rate δ and are not replaced, so the number of goods being imported by the home country falls. As the number of goods being imported into the home country falls, the composition component of the CES index rises, leading to an increase in the CES index. This can be seen in Figure 6, which plots the CES and Laspeyres indices from a subsample of the simulation. The Laspeyres index, whose composition component is constant, changes only as the aggregate component changes. The simple correlation coefficient between these two indices, for the entire simulation, is 0.80. The divergence between these two measures of prices is what drives the lower estimated elasticity.

Trade Liberalization

To see if the model can generate the large response to changes in tariffs, we consider the complete removal of the 15 percent tariff in the baseline model. To capture this idea, we compute the equilibrium under the 15 percent tariff and then compute the equilibrium in the absence of tariffs. By proceeding in this way, we avoid having to compute the transition path from the high tariff equilibrium to the low tariff equilibrium. The transition path between tariff policies is certainly interesting from the point of view of welfare analysis, but is left for future research. The results of the decrease in tariffs are dramatic. As can be seen in the first column of Table 4, the decline in tariffs leads to an 87.1 percent increase in exports. The ratio of imports to domestic consumption grows by 93.0 percent as a result of the tariff decrease. Measuring the Armington elasticity with respect to the 15 percent tariff change yields an elasticity of 6.2, which is more than 4 times the elasticity estimated from the time series data. This value is in the middle of the range of values reported in Hummels (2001), whose estimates range from about 3 to 8, with an average of 5.6. Our estimate also falls into the range estimated in Romalis (2002), which finds elasticities typically between 4 and 13. The estimates in Clausing (2001) and Head and Ries (2001), who study the Canada-U.S. Free Trade Agreement, range from 8 to 11. Though the value of 6.2 found here is lower than their findings, our simple model has abstracted from other possibly important elements such as returns to scale, intermediate good trade, or capital accumulation.

The four-fold difference between the time series estimate of the Armington elasticity and the elasticity implied by trade liberalization is driven by the growth in trade on the extensive margin. Table 4 displays the results of trade liberalization in two different models: the baseline model specified in Sections 3 and 4, and a model identical to the baseline model except that the entry costs have been set to 0 for all plants. The model without entry costs has the obvious counterfactual implication that all plants export. Given this characteristic of the model, we set the home bias parameter, ω, to 0.8 so that the country still only exports 7.5 percent of output when tariffs are 15 percent. The rest of the model is calibrated as in Section 4, although we no longer have to choose the maximum value of the entry cost or the shape parameter for the distribution over entry costs. The effects of trade liberalization in the two models are drastically different. Eliminating a 15 percent tariff in the baseline model increases exports by 87.1 percent, while exports increase by only 30.5 percent in the model without entry costs. The extra 57.1 percentage points of growth in exports are due to the large number of new exporters that enter the market following the decrease in tariffs. The model without entry costs has no extensive margin; all the plants are already exporting. The increase in trade is driven by the lower delivered price that results from the elimination of the tariff. The baseline model, however, has a substantial extensive margin; almost 38 percent more plants export following the elimination of the tariff. The extra trade generated by the new exporters, along with the intensive margin trade, drive up the ratio of imports to domestic consumption by 93.0 percent; a value 2.9 times larger than the change in the model without entry costs. This increase implies an Armington elasticity of 6.2 compared to the measured elasticity of 2.1 in the model without entry costs.

6. Conclusion

Models in which countries trade differentiated goods depend crucially on the Armington elasticity to determine both the short run and the long run behavior of trade flows and prices. Models that were built to explain high frequency fluctuations, such as those in the IRBC literature, need small values of the Armington elasticity, while applied GE models need large elasticities to match the increase in trade following trade liberalization. Econometric estimates support this dichotomy: estimates derived from high-frequency time series data are low, and range between 0.2 and 3.5, while estimates gleaned from trade liberalizations range from 6 to 15. This paper reconciles these two observations by recognizing that the sources of variation in the two approaches are different. The high-frequency variation in the time series studies is caused by small and persistent, but not permanent, shocks to productivity or demand. Trade liberalization, however, can be thought of as a permanent change. When agents react differently to temporary and permanent changes, the measured elasticities will differ.

We build a model that incorporates elements from the international real business cycle literature and the applied GE models that are commonly used to study trade policy. The key feature of the model is that plants face a cost of entering the export market, as well as uncertainty about future profits from exporting. Temporary shocks will induce few plants to change export status and incumbent exporters respond to the shocks by changing prices. Measuring the elasticity with respect to these changes will recover the low elasticity implied by the model's parameters. A permanent tariff change increases the future value of exporting in all states, and induces a larger number of plants to begin exporting. When the trade from these newly imported goods is not properly accounted for, the response of trade to small changes in tariffs looks large, implying a high Armington elasticity. The calibrated model is capable of producing time series estimates of the Armington elasticity of about 1.4, while the elasticity in response to a reduction in tariffs is 6.2. The initial success of this model adds to the growing evidence that studying plant level choices can lead to a greater understanding of international trade.

References

Anderson, M. A., E. J. Balistreri, A. K. Fox and R. H. Hillberry (2002), "Taste Parameters as Model Residuals: Assessing the 'Fit' of an Armington Trade Model," U.S. International Trade Commission and Washington and Lee University.

Armington, P. S. (1969), "A Theory of Demand for Products Distinguished by Place of Production," *International Monetary Fund Staff Papers*, 16(1), 159-178.

Backus, D. K., P. J. Kehoe and F. E. Kydland (1992), "International Real Business Cycles," *Journal of Political Economy*, 100(4), 745-775.

Backus, D. K., P. J. Kehoe and F. E. Kydland (1994), "Dynamics of the Trade Balance and the Terms of Trade: The J-Curve?" *American Economic Review*, 84(1), 84-103.

Baier, S. L. and J. H. Bergstrand (2001), "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity," *Journal of International Economics*, 53(1), 1-27.

Baldwin, R. and P. Krugman (1989), "Persistent Trade Effects of Large Exchange Rate Shocks," *Quarterly Journal of Economics*, 104(4), 635-654.

Baldwin, R. E. (1988), "Hysteresis in Import Prices: The Beachhead Effect," *American Economic Review*, 78(4), 773-785.

Bernard, A. B., J. Eaton, J. B. Jensen and S. Kortum (2003), "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268-1290.

Bernard, A. B. and J. B. Jensen (2001), "Why Some Firms Export," *8349*.

Bernard, A. B., J. B. Jensen and K. P. Schott (2003), "Falling Trade Costs, Heterogeneous Firms, and Industry Dynamics," Tuck School of Business, Institute for International Economics, and Yale School of Management.

Bernard, A. B. and J. Wagner (2001), "Export Entry and Exit by German Firms," *Weltwirtschaftliches Archiv/Review of World Economics*, 137(1), 105-123.

Blonigen, B. A. and W. W. Wilson (1999), "Explaining Armington: What Determines Substitutability between Home and Foreign Goods?" *Canadian Journal of Economics*, 32(1), 1-21.

Choi, H. (2003), "Export Penetration Costs and International Business Cycles," University of Auckland.

Clausing, K. A. (2001), "Trade Creation and Trade Diversion in the Canada-United States Free Trade Agreement," *Canadian Journal of Economics*, 34(3), 677-696.

Cox, D. J. (1994), "Some Applied General Equilibrium Estimates of the Impact of a North American Free Trade Agreement on Canada," *Modeling Trade Policy: Applied General Equilibrium Assessments in North American Free Trade*, 100-123.

Davis, S. J., J. C. Haltiwanger and S. Schuh (1996), *Job Creation and Destruction*. Cambridge, Mass., MIT Press.

Dixit, A. K. (1989), "Hysteresis, Import Penetration, and Exchange Rate Pass-Through," *Quarterly Journal of Economics*, 104(2), 205-228.

Dixit, A. K. and J. E. Stiglitz (1977), "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297-308.

Eaton, J. and S. Kortum (2002), "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741-1779.

Feenstra, R. C. (1994), "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84(1), 157-177.

Feenstra, R. C. and C. R. Shiells (1997), "Bias in U.S. Import Prices and Demand," in T. F. Bresnahan and R. J. Gordon, eds., *The Economics of New Goods*, National Bureau of Economic Research Studies in Income and Wealth, vol. 58. Chicago and London: University of Chicago Press, 249-273.

Gallaway, M. P., C. A. McDaniel and S. A. Rivera (2003), "Short-Run and Long-Run Industry-Level Estimates of U.S. Armington Elasticities," *North American Journal of Economics and Finance*, 14(1), 49-68.

Golosov, M. and R. E. Lucas, Jr. (2003), "Menu Costs and Phillips Curves," University of Minnesota and University of Chicago.

Head, K. and J. Ries (2001), "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade," *American Economic Review*, 91(4), 858-876.

Heathcote, J. and F. Perri (2002), "Financial Autarky and International Business Cycles," *Journal of Monetary Economics*, 49(3), 601-627.

Heathcote, J. and F. Perri (2003), "Financial Globalization and Real Regionalization," Georgetown University and New York University Stern School of Business.

Helkie, W. L. and P. Hooper (1988), "An Empirical Analysis of the External Deficit, 1980-86," in R. C. e. Bryant, G. e. Holtham and P. e. Hooper, eds., *External Deficits and the Dollar: The Pit and the Pendulum*, Washington, D.C.: Brookings Institution, 10-56.

Helpman, E. and P. R. Krugman (1985), *Market Structure and Foreign Trade : Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge, Mass., MIT Press.

Hillberry, R. H. and C. A. McDaniel (2002), "A Decomposition of North American Trade Growth since NAFTA," *U.S. International Trade Commission 2002-12-A*.

Hummels, D. (2001), "Toward a Geography of Trade Costs," Purdue University.

Hummels, D. and P. J. Klenow (2002), "The Variety and Quality of a Nation's Exports," Purdue University and Stanford University.

Kehoe, T. J. and K. J. Ruhl (2002), "How Important is the New Goods Margin in International Trade?" *Federal Reserve Bank of Minneapolis Staff Report 324*.

Krugman, P. (1989), "Differences in Income Elasticities and Trends in Real Exchange Rates," *European Economic Review*, 33(5), 1031-1046.

Krusell, P. and A. A. Smith, Jr. (1998), "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106(5), 867-896.

Kydland, F. E. and E. C. Prescott (1982), "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345-1370.

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

Reinert, K. A. and D. W. Roland-Holst (1992), "Armington Elasticities for United States Manufacturing Sectors," *Journal of Policy Modeling*, 14(5), 631-639.

Reinert, K. A. and C. R. Shiells (1993), "Armington Models and Terms-of-Trade Effects: Some Econometric Evidence for North America," *Canadian Journal of Economics*, 26(2), 299-316.

Roberts, M. J. and J. R. Tybout (1997), "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87(4), 545-564.

Romalis, J. (2002), "NAFTA's and CUSFTA's Impact on North American Trade," University of Chicago GSB.

Stockman, A. C. and L. L. Tesar (1995), "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements," *American Economic Review*, 85(1), 168-185.

Yeaple, S. R. (2002), "A Simple Model of Firm Heterogeneity, International Trade and Wages," University of Pennsylvania.

Yi, K.-M. (2003), "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy*, 111(1), 52-102.

Zimmermann, C. (1997), "International Real Business Cycles among Heterogeneous Countries," *European Economic Review*, 41(2), 319-356.

Table 1
North American Free Trade Agreement (1993-2001)

	Imports/Domestic Consumption† (% Change)	**Tariffs** (% Change)	**Implied σ**
United States	81.6	5.0	16.2
Canada	36.1	5.0	7.2
Mexico	117.1	13.0	9.0

† Domestic consumption is gross output in agriculture, mining and manufacturing minus total exports in these sectors. Imports are imports of agriculture, mining and manufacturing from the NAFTA partner countries.

Table 2
Baseline Calibration: $\tau = 0.15$, $\rho = 0.50$

Parameter	**Data**
β	Annual real interest rate (4.0%)
δ	Annual rate of jobs lost from plant closure as fraction of total employment (2.4%)
γ	Share of manufacturing in total production (18.7 %)
$\overline{\phi}$	Employment in largest plant (6,000 employees)
$\overline{\kappa}$	Fraction of output that is exported (7.5%)
ν	Average plant size (65 employees)
θ_{ϕ}	Size distribution of all plants (see Figure 4)
θ_{κ}	Size distribution of exporting plants (see Figure 5)
ε	Standard deviation of Solow Residuals (1.0%)
$\overline{\lambda}$	Autocorrelation of Solow Residuals (0.90)

Table 3
Estimates from Simulated Time Series Data

	CES Prices	**Laspeyres Prices**
Intercept	5.54e-11	-0.015
(standard error)	(1.43e-12)	(6.36e-04)
σ	2.00	1.39
(standard error)	(6.04e-13)	(0.06)
R-squared	1.00	0.30

Table 4
Response to Permanent Removal of 15% Tariff

Variable	**Entry Costs Model** (% change)	**No Entry Costs Model** (% change)
Exports	87.1	30.5
Imports/Dom. Consumption	93.0	32.2
Exporting Plants	37.7	0.0
Implied σ	6.2	2.1

Figure 1

$\mu_d(\phi,\kappa)$ measure of non-exporters of type (ϕ,κ)

$\mu_x(\phi,\kappa)$ measure of exporters of type (ϕ,κ)

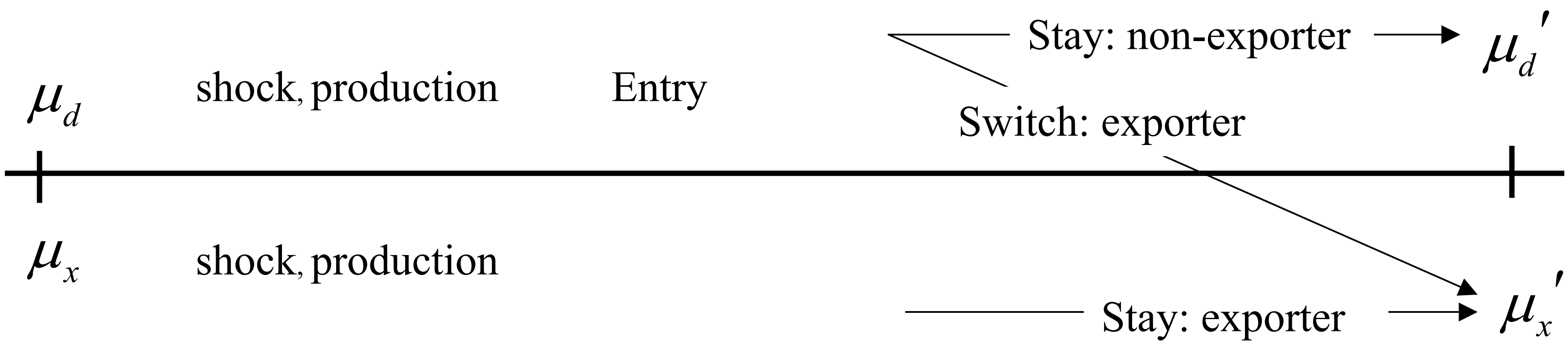


Figure 2

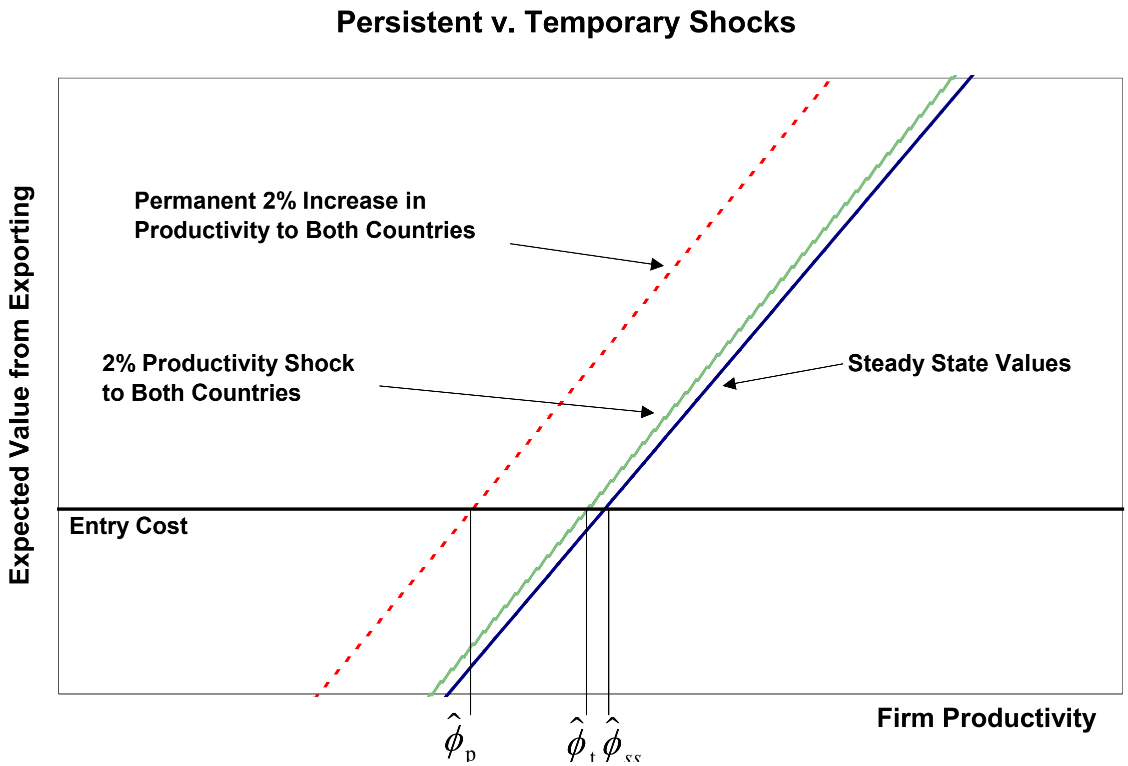


Figure 3

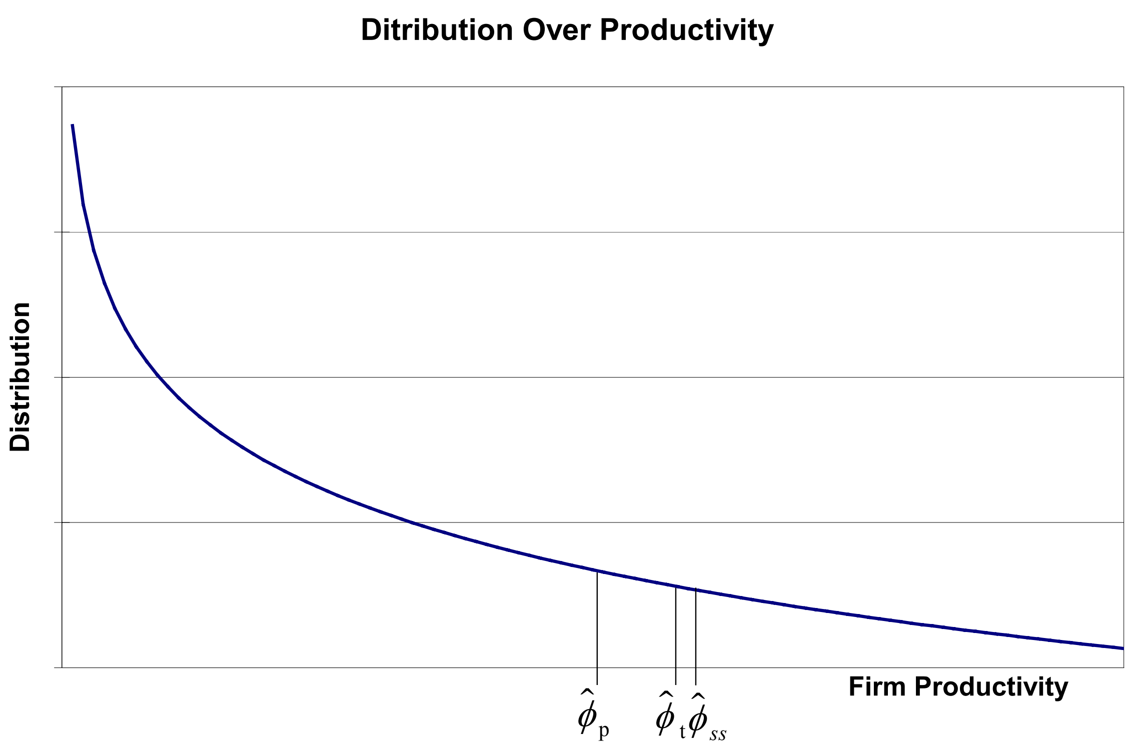


Figure 4

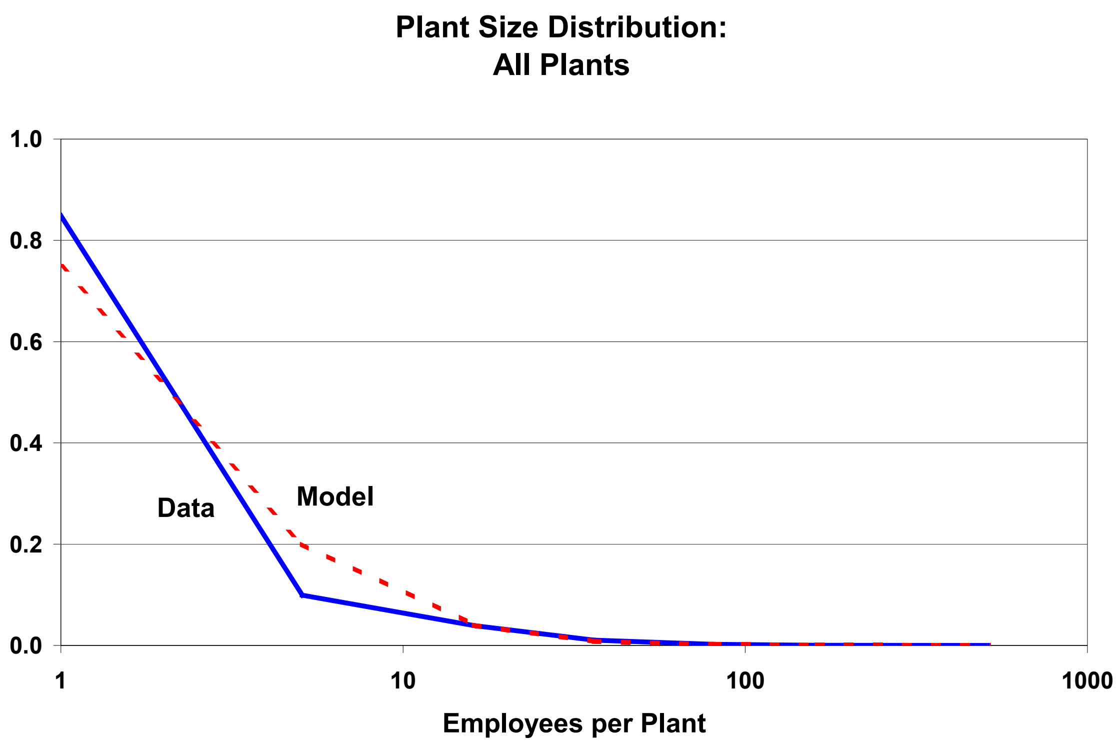


Figure 5

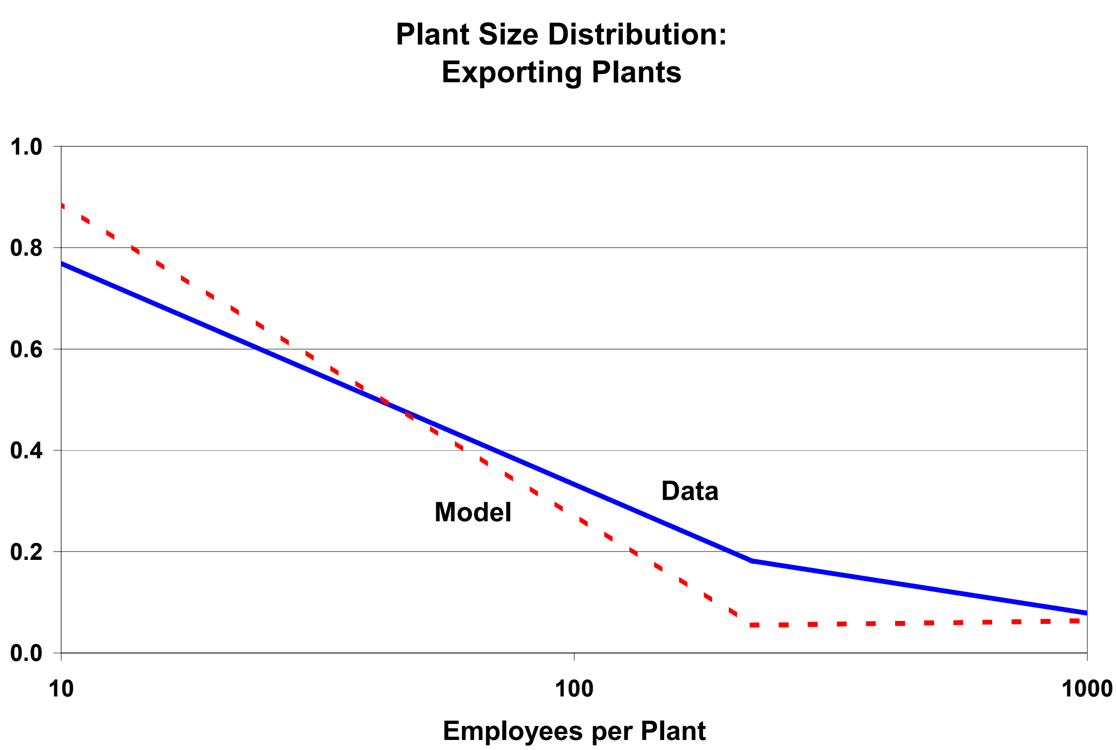
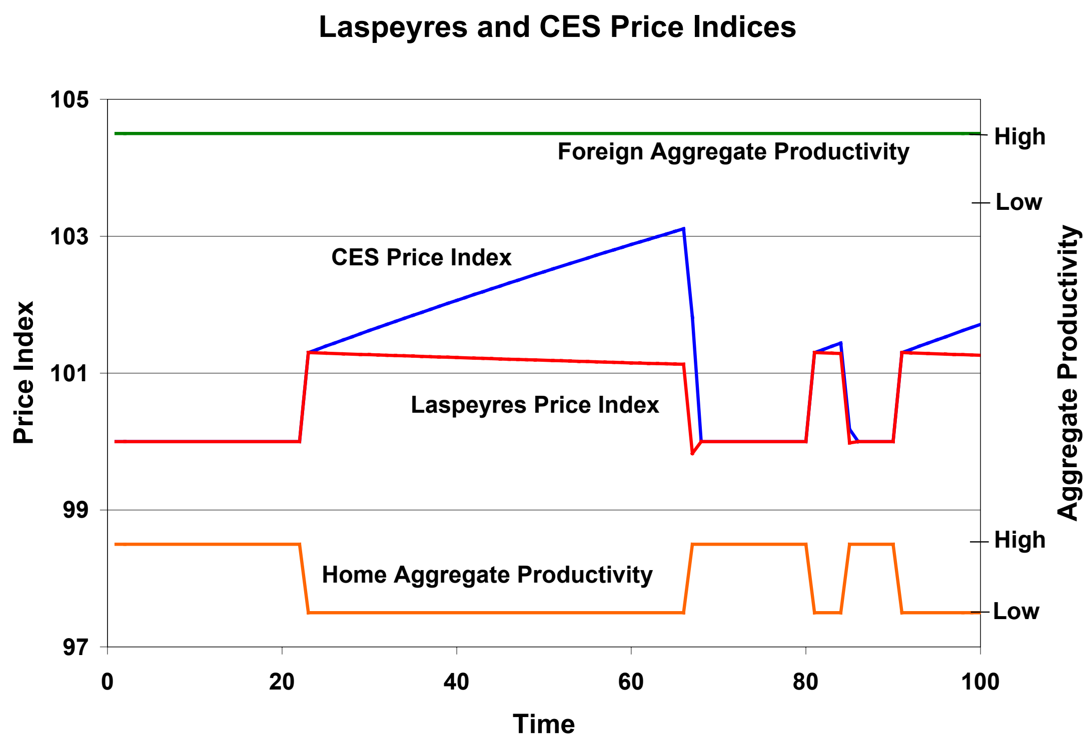


Figure 6


Laspeyres and CES Price Indices
105
103
101
99
97
0
20
40
60
80
100
Time
Price Index
Aggregate Productivity
High
Low
High
Low
Foreign Aggregate Productivity
CES Price Index
Laspeyres Price Index
Home Aggregate Productivity