

THE IMPACT OF TRADE ON INTRA-INDUSTRY REALLOCATIONS AND AGGREGATE INDUSTRY PRODUCTIVITY

BY MARC J. MELITZ¹

This paper develops a dynamic industry model with heterogeneous firms to analyze the intra-industry effects of international trade. The model shows how the exposure to trade will induce only the more productive firms to enter the export market (while some less productive firms continue to produce only for the domestic market) and will simultaneously force the least productive firms to exit. It then shows how further increases in the industry's exposure to trade lead to additional inter-firm reallocations towards more productive firms. The paper also shows how the aggregate industry productivity growth generated by the reallocations contributes to a welfare gain, thus highlighting a benefit from trade that has not been examined theoretically before. The paper adapts Hopenhayn's (1992a) dynamic industry model to monopolistic competition in a general equilibrium setting. In so doing, the paper provides an extension of Krugman's (1980) trade model that incorporates firm level productivity differences. Firms with different productivity levels coexist in an industry because each firm faces initial uncertainty concerning its productivity before making an irreversible investment to enter the industry. Entry into the export market is also costly, but the firm's decision to export occurs after it gains knowledge of its productivity.

KEYWORDS: Intra-industry trade, firm heterogeneity, firm dynamics, selection.

1. INTRODUCTION

RECENT EMPIRICAL RESEARCH using longitudinal plant or firm-level data from several countries has overwhelmingly substantiated the existence of large and persistent productivity differences among establishments in the same narrowly defined industries. Some of these studies have further shown that these productivity differences are strongly correlated with the establishment's export status: relatively more productive establishments are much more likely to export (even within so-called "export sectors," a substantial portion of establishments do not export). Other studies have highlighted the large levels of resource reallocations that occur across establishments in the same industry. Some of these studies have also correlated these reallocations with the establishments' export status.

This paper develops a dynamic industry model with heterogeneous firms to analyze the role of international trade as a catalyst for these inter-firm reallocations within an industry. The model is able to reproduce many of the most salient patterns emphasized by recent micro-level studies related to trade. The model shows how the exposure to trade induces only the more productive firms to export while simultaneously forcing the least productive firms to exit. Both the exit of the least productive firms and the additional export sales gained by

¹Many thanks to Alan Deardorff, Jim Levinsohn, and Elhanan Helpman for helpful comments and discussions. This manuscript has also benefited from comments by the editor and two anonymous referees. Funding from the Alfred P. Sloan Foundation is gratefully acknowledged.

the more productive firms reallocate market shares towards the more productive firms and contribute to an aggregate productivity increase. Profits are also reallocated towards more productive firms. The model is also consistent with the widely reported stories in the business press describing how the exposure to trade enhances the growth opportunities of some firms while simultaneously contributing to the downfall or “downsizing” of other firms in the same industry; similarly, protection from trade is often reported to shelter inefficient firms. Rigorous empirical work has recently corroborated this anecdotal evidence. Bernard and Jensen (1999a) (for the U.S.), Aw, Chung, and Roberts (2000) (for Taiwan), and Clerides, Lack, and Tybout (1998) (for Colombia, Mexico, and Morocco) all find evidence that more productive firms self-select into export markets. Aw, Chung, and Roberts (2000) also find evidence suggesting that exposure to trade forces the least productive firms to exit. Pavcnik (2002) directly looks at the contribution of market share reallocations to sectoral productivity growth following trade liberalization in Chile. She finds that these reallocations significantly contribute to productivity growth in the tradable sectors. In a related study, Bernard and Jensen (1999b) find that within-sector market share reallocations towards more productive exporting plants accounts for 20% of U.S. manufacturing productivity growth.

Clearly, these empirical patterns cannot be motivated without appealing to a model of trade incorporating firm heterogeneity. Towards this goal, this paper embeds firm productivity heterogeneity within Krugman’s model of trade under monopolistic competition and increasing returns. The current model draws heavily from Hopenhayn’s (1992a, 1992b) work to explain the endogenous selection of heterogeneous firms in an industry. Hopenhayn derives the equilibrium distribution of firm productivity from the profit maximizing decisions of initially identical firms who are uncertain of their initial and future productivity.² This paper adapts his model to a monopolistically competitive industry (Hopenhayn only considers competitive firms) in a general equilibrium setting.³ A contribution of this paper is to provide such a general equilibrium model incorporating firm heterogeneity that yet remains highly tractable. This is achieved by integrating firm heterogeneity in a way such that the relevance of the distribution of productivity levels for aggregate outcomes is completely summarized by a single “sufficient” statistic—an average firm productivity level. Once this productivity average is determined, the model with

²One of the most robust empirical patterns emerging from recent industry studies is that new entrants have lower average productivity and higher exit rates than older incumbents. This suggests that uncertainty concerning productivity is an important feature explaining the behavior of prospective and new entrants.

³Montagna (1995) also adapts Hopenhayn’s model to a monopolistic competition environment (in a partial equilibrium setting), but confines the analysis to a static equilibrium with no entry or exit and further constrains the distribution of firm productivity levels to be uniform. Although it is assumed that only the more productive firms earning positive profits remain in the industry in future periods, the present value of these profit flows does not enter into the firms’ entry decision.

productivity heterogeneity yields identical aggregate outcomes as one with representative firms that all share the same average productivity level.

This simplicity does not come without some concessions. The analysis relies on the Dixit and Stiglitz (1977) model of monopolistic competition. Although this modeling approach is quite common in the trade literature, it also exhibits some well-known limitations. In particular, the firms' markups are exogenously fixed by the symmetric elasticity of substitution between varieties. Another concession is the simplification of the firm productivity dynamics modeled by Hopenhayn (1992a). Nevertheless, the current model preserves the initial firm uncertainty over productivity and the forward looking entry decision of firms facing sunk entry costs and expected future probabilities of exit. As in Hopenhayn (1992a), the analysis is restricted to stationary equilibria. Firms correctly anticipate this stable aggregate environment when making all relevant decisions. The analysis then focuses on the long run effects of trade on the relative behavior and performance of firms with different productivity levels.

Another recent paper by Bernard, Eaton, Jenson, and Kortum (2000) (henceforth BEJK) also introduces firm-level heterogeneity into a model of trade by adapting a Ricardian model to firm-specific comparative advantage. Both papers predict the same basic kinds of trade-induced reallocations, although the channels and motivations behind these reallocations vary. In BEJK, firms compete to produce the same variety—including competition between domestic and foreign producers of the same variety. This delivers an endogenous distribution of markups, a feature that is missing in this paper. BEJK also show how their model can be calibrated to provide a good fit to a combination of micro and macro US data patterns. Comparative statics are then obtained by simulating this fitted model. The BEJK model assumes that the total number of world varieties produced and consumed remains exogenously fixed and relies on a specific parameterization of the distribution of productivity levels.

In contrast, the current paper allows the total range of varieties produced to vary with the exposure to trade, and endogenously determines the subset of those varieties that are consumed in a given country. Despite leaving the distribution of firm productivity levels unrestricted, the model remains tractable enough to perform analytical comparisons of steady states that reflect different levels of exposure to trade. Although the current model only considers symmetric countries, it can easily be extended to asymmetric countries by relying on an exogenously fixed relative wage between countries.⁴ In this model, differences in country size—holding the relative wage fixed—only affect the relative number of firms, and not their productivity distribution. This straightforward extension is therefore omitted for expositional simplicity.

⁴This assumption is also made in BEJK. See Helpman, Melitz, and Yeaple (2002) for an extension of the current model that explicitly considers the asymmetric country case when the relative wage is determined via trade in a homogeneous good sector.

One last, but important, innovation in the current paper is to introduce the dynamic forward-looking entry decision of firms facing sunk market entry costs. Firms face such costs, not just for their domestic market, but also for any potential export market.⁵ These costs are in addition to the per-unit trade costs that are typically modeled. Both survey and econometric works have documented the importance of such export market entry costs. Das, Roberts, and Tybout (2001) econometrically estimate these costs average over U.S. \$1 Million for Colombian plants producing industrial chemicals. As will be detailed later, surveys reveal that managers making export related decisions are much more concerned with export costs that are fixed in nature rather than high per-unit costs. Furthermore, Roberts and Tybout (1977a) (for Colombia), Bernard and Jensen (2001) (for the U.S.), and Bernard and Wagner (2001) (for Germany) estimate that the magnitude of sunk export market entry costs is important enough to generate very large hysteresis effects associated with a plant's export market participation.⁶

2. SETUP OF THE MODEL

2.1. Demand

The preferences of a representative consumer are given by a C.E.S. utility function over a continuum of goods indexed by ω :

$$U = \left[\int_{\omega \in \Omega} q(\omega)^\rho d\omega \right]^{1/\rho},$$

where the measure of the set Ω represents the mass of available goods. These goods are substitutes, implying $0 < \rho < 1$ and an elasticity of substitution between any two goods of $\sigma = 1/(1 - \rho) > 1$. As was originally shown by Dixit and Stiglitz (1977), consumer behavior can be modeled by considering the set of varieties consumed as an aggregate good $Q \equiv U$ associated with an aggregate price

$$(1) \quad P = \left[\int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}.$$

⁵Sunk market entry costs also explain the presence of simultaneous entry and exit in the steady state equilibrium.

⁶Sunk export market entry costs also explain the higher survival probabilities of exporting firms—even after controlling for their higher measured productivity. See Bernard and Jensen (1999a, 2002) for evidence on U.S. firms.

These aggregates can then be used to derive the optimal consumption and expenditure decisions for individual varieties using

$$(2) \quad \begin{aligned} q(\omega) &= Q \left[\frac{p(\omega)}{P} \right]^{-\sigma}, \\ r(\omega) &= R \left[\frac{p(\omega)}{P} \right]^{1-\sigma}, \end{aligned}$$

where $R = PQ = \int_{\omega \in \Omega} r(\omega) d\omega$ denotes aggregate expenditure.

2.2. Production

There is a continuum of firms, each choosing to produce a different variety ω . Production requires only one factor, labor, which is inelastically supplied at its aggregate level L , an index of the economy's size. Firm technology is represented by a cost function that exhibits constant marginal cost with a fixed overhead cost. Labor used is thus a linear function of output q : $l = f + q/\varphi$. All firms share the same fixed cost $f > 0$ but have different productivity levels indexed by $\varphi > 0$. For expositional simplicity, higher productivity is modeled as producing a symmetric variety at lower marginal cost. Higher productivity may also be thought of as producing a higher quality variety at equal cost.⁷ Regardless of its productivity, each firm faces a residual demand curve with constant elasticity σ and thus chooses the same profit maximizing markup equal to $\sigma/(\sigma - 1) = 1/\rho$. This yields a pricing rule

$$(3) \quad p(\varphi) = \frac{w}{\rho\varphi},$$

where w is the common wage rate hereafter normalized to one. Firm profit is then

$$\pi(\varphi) = r(\varphi) - l(\varphi) = \frac{r(\varphi)}{\sigma} - f,$$

where $r(\varphi)$ is firm revenue and $r(\varphi)/\sigma$ is variable profit. $r(\varphi)$, and hence $\pi(\varphi)$, also depend on the aggregate price and revenue as shown in (2):

$$(4) \quad r(\varphi) = R(P\rho\varphi)^{\sigma-1},$$

$$(5) \quad \pi(\varphi) = \frac{R}{\sigma}(P\rho\varphi)^{\sigma-1} - f.$$

⁷Given the form of product differentiation, the modeling of either type of productivity difference is isomorphic.

On the other hand, the ratios of any two firms' outputs and revenues only depend on the ratio of their productivity levels:

$$(6) \quad \frac{q(\varphi_1)}{q(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^\sigma, \quad \frac{r(\varphi_1)}{r(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma-1}.$$

In summary, a more productive firm (higher φ) will be bigger (larger output and revenues), charge a lower price, and earn higher profits than a less productive firm.

2.3. Aggregation

An equilibrium will be characterized by a mass M of firms (and hence M goods) and a distribution $\mu(\varphi)$ of productivity levels over a subset of $(0, \infty)$. In such an equilibrium, the aggregate price P defined in (1) is then given by

$$P = \left[\int_0^\infty p(\varphi)^{1-\sigma} M \mu(\varphi) d\varphi \right]^{\frac{1}{1-\sigma}}.$$

Using the pricing rule (3), this can be written $P = M^{1/(1-\sigma)} p(\tilde{\varphi})$, where

$$(7) \quad \tilde{\varphi} = \left[\int_0^\infty \varphi^{\sigma-1} \mu(\varphi) d\varphi \right]^{\frac{1}{\sigma-1}}.$$

$\tilde{\varphi}$ is a weighted average of the firm productivity levels φ and is independent of the number of firms M .⁸ These weights reflect the relative output shares of firms with different productivity levels.⁹ $\tilde{\varphi}$ also represents aggregate productivity because it completely summarizes the information in the distribution of productivity levels $\mu(\varphi)$ relevant for all aggregate variables (see Appendix):

$$\begin{aligned} P &= M^{\frac{1}{1-\sigma}} p(\tilde{\varphi}), & R &= PQ = Mr(\tilde{\varphi}), \\ Q &= M^{1/\rho} q(\tilde{\varphi}), & \Pi &= M\pi(\tilde{\varphi}), \end{aligned}$$

where $R = \int_0^\infty r(\varphi) M \mu(\varphi) d\varphi$ and $\Pi = \int_0^\infty \pi(\varphi) M \mu(\varphi) d\varphi$ represent aggregate revenue (or expenditure) and profit. Thus, an industry comprised of M firms with any distribution of productivity levels $\mu(\varphi)$ that yields the same average productivity level $\tilde{\varphi}$ will also induce the same aggregate outcome as an industry with M representative firms sharing the same productivity level $\varphi = \tilde{\varphi}$.

⁸Subsequent conditions on the equilibrium $\mu(\varphi)$ must of course ensure that $\tilde{\varphi}$ is finite.

⁹Using $q(\varphi)/q(\tilde{\varphi}) = (\varphi/\tilde{\varphi})^\sigma$ (see (6)), $\tilde{\varphi}$ can be written as $\tilde{\varphi}^{-1} = \int_0^\infty \varphi^{-1} [q(\varphi)/q(\tilde{\varphi})] \mu(\varphi) d\varphi$. $\tilde{\varphi}$ is therefore the weighted harmonic mean of the φ 's where the weights $q(\varphi)/q(\tilde{\varphi})$ index the firms' relative output shares.

This variable will be alternatively referred to as aggregate or average productivity. Further note that $\bar{r} = R/M$ and $\bar{\pi} = \Pi/M$ represent both the average revenue and profit per firm as well as the revenue and profit level of the firm with average productivity level $\varphi = \tilde{\varphi}$.

3. FIRM ENTRY AND EXIT

There is a large (unbounded) pool of prospective entrants into the industry. Prior to entry, firms are identical. To enter, firms must first make an initial investment, modeled as a fixed entry cost $f_e > 0$ (measured in units of labor), which is thereafter sunk. Firms then draw their initial productivity parameter φ from a common distribution $g(\varphi)$.¹⁰ $g(\varphi)$ has positive support over $(0, \infty)$ and has a continuous cumulative distribution $G(\varphi)$.

Upon entry with a low productivity draw, a firm may decide to immediately exit and not produce. If the firm does produce, it then faces a constant (across productivity levels) probability δ in every period of a bad shock that would force it to exit. Although there are some realistic examples of severe shocks that would constrain a firm to exit independently of productivity (such as natural disasters, new regulation, product liability, major changes in consumer tastes), it is also likely that exit may be caused by a series of bad shocks affecting the firm's productivity. This type of firm level process is explicitly modeled by Hopenhayn (1992a, 1992b). The simplification made in this model entails that the shape of the equilibrium distribution of productivity $\mu(\varphi)$ and the ex-ante survival probabilities are exogenously determined by $g(\varphi)$ and δ . On the other hand, the range of productivity levels (for surviving firms), and hence the average productivity level, are endogenously determined.¹¹ Importantly, this simplified industry model will nevertheless generate one of the most robust empirical patterns highlighted by micro-level studies: new entrants (including the firms whose entry is unsuccessful) will have, on average, lower productivity and a higher probability of exit than incumbents.

This paper only considers steady state equilibria in which the aggregate variables remain constant over time. Since each firm's productivity level does not change over time, its optimal per period profit level (excluding f_e) will also remain constant. An entering firm with productivity φ would then immediately exit if this profit level were negative (and hence never produce), or would produce and earn $\pi(\varphi) \geq 0$ in every period until it is hit with the bad shock and is

¹⁰This captures the fact that firms cannot know their own productivity with certainty until they start producing and selling their good. (Recall that productivity differences may reflect cost differences as well as differences in consumer valuations of the good.)

¹¹The increased tractability afforded by this simplification permits the detailed analysis of the impact of trade on this endogenous range of productivity levels and on the distribution of market shares and profits across this range.

forced to exit. Assuming that there is no time discounting,¹² each firm's value function is given by

$$v(\varphi) = \max \left\{ 0, \sum_{t=0}^{\infty} (1-\delta)^t \pi(\varphi) \right\} = \max \left\{ 0, \frac{1}{\delta} \pi(\varphi) \right\},$$

where the dependence of $\pi(\varphi)$ on R and P from (5) is understood. Thus, $\varphi^* = \inf\{\varphi : v(\varphi) > 0\}$ identifies the lowest productivity level (hereafter referred to as the cutoff level) of producing firms. Since $\pi(0) = -f$ is negative, $\pi(\varphi^*)$ must be equal to zero. This will be referred to as the zero cutoff profit condition.

Any entering firm drawing a productivity level $\varphi < \varphi^*$ will immediately exit and never produce. Since subsequent firm exit is assumed to be uncorrelated with productivity, the exit process will not affect the equilibrium productivity distribution $\mu(\varphi)$. This distribution must then be determined by the initial productivity draw, conditional on successful entry. Hence, $\mu(\varphi)$ is the conditional distribution of $g(\varphi)$ on $[\varphi^*, \infty)$:

$$(8) \quad \mu(\varphi) = \begin{cases} \frac{g(\varphi)}{1 - G(\varphi^*)} & \text{if } \varphi \geq \varphi^*, \\ 0 & \text{otherwise,} \end{cases}$$

and $p_{in} \equiv 1 - G(\varphi^*)$ is the ex-ante probability of successful entry.¹³ This defines the aggregate productivity level $\tilde{\varphi}$ as a function of the cutoff level φ^* :¹⁴

$$(9) \quad \tilde{\varphi}(\varphi^*) = \left[\frac{1}{1 - G(\varphi^*)} \int_{\varphi^*}^{\infty} \varphi^{\sigma-1} g(\varphi) d\varphi \right]^{\frac{1}{\sigma-1}}.$$

The assumption of a finite $\tilde{\varphi}$ imposes certain restrictions on the size of the upper tail of the distribution $g(\varphi)$: the $(\sigma - 1)$ th uncentered moment of $g(\varphi)$ must be finite. Equation (8) clearly shows how the shape of the equilibrium distribution of productivity levels is tied to the exogenous ex-ante distribution $g(\varphi)$ while allowing the range of productivity levels (indexed by the cutoff φ^*) to be endogenously determined. Equation (9) then shows how this endogenous range affects the aggregate productivity level.

¹²Again, this is assumed for simplicity. The probability of exit δ introduces an effect similar to time discounting. Modeling an additional time discount factor would not qualitatively change any of the results.

¹³The equilibrium distribution $\mu(\varphi)$ can be determined from the distribution of initial productivity with certainty by applying a law of large numbers to $g(\varphi)$. See Hopenhayn (1992a, note 5) for further details.

¹⁴This dependence of $\tilde{\varphi}$ on φ^* is understood when it is subsequently written without its argument.

3.1. Zero Cutoff Profit Condition

Since the average productivity level $\tilde{\varphi}$ is completely determined by the cutoff productivity level φ^* , the average profit and revenue levels are also tied to the cutoff level φ^* (see (6)):

$$\bar{r} = r(\tilde{\varphi}) = \left[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*} \right]^{\sigma-1} r(\varphi^*), \quad \bar{\pi} = \pi(\tilde{\varphi}) = \left[\frac{\tilde{\varphi}(\varphi^*)}{\varphi^*} \right]^{\sigma-1} \frac{r(\varphi^*)}{\sigma} - f.$$

The zero cutoff profit condition, by pinning down the revenue of the cutoff firm, then implies a relationship between the average profit per firm and the cutoff productivity level:

$$(10) \quad \pi(\varphi^*) = 0 \iff r(\varphi^*) = \sigma f \iff \bar{\pi} = f k(\varphi^*),$$

where $k(\varphi^*) = [\tilde{\varphi}(\varphi^*)/\varphi^*]^{\sigma-1} - 1$.

3.2. Free Entry and the Value of Firms

Since all incumbent firms—other than the cutoff firm—earn positive profits, the average profit level $\bar{\pi}$ must be positive. In fact, the expectation of future positive profits is the only reason that firms consider sinking the investment cost f_e required for entry. Let \bar{v} represent the present value of the average profit flows: $\bar{v} = \sum_{t=0}^{\infty} (1-\delta)^t \bar{\pi} = (1/\delta) \bar{\pi}$. Also \bar{v} is the average value of firms, conditional on successful entry: $\bar{v} = \int_{\varphi^*}^{\infty} v(\varphi) \mu(\varphi) d\varphi$. Further define v_e to be the net value of entry:

$$(11) \quad v_e = p_{in} \bar{v} - f_e = \frac{1 - G(\varphi^*)}{\delta} \bar{\pi} - f_e.$$

If this value were negative, no firm would want to enter. In any equilibrium where entry is unrestricted, this value could further not be positive since the mass of prospective entrants is unbounded.

4. EQUILIBRIUM IN A CLOSED ECONOMY

The free entry (FE) and zero cutoff profit (ZCP) conditions represent two different relationships linking the average profit level $\bar{\pi}$ with the cutoff productivity level φ^* (see (10) and (11)):

$$(12) \quad \begin{aligned} \bar{\pi} &= f k(\varphi^*) && \text{(ZCP),} \\ \bar{\pi} &= \frac{\delta f_e}{1 - G(\varphi^*)} && \text{(FE).} \end{aligned}$$

In (φ, π) space, the FE curve is increasing and is cut by the ZCP curve only once from above (see Appendix for proof). This ensures the existence and

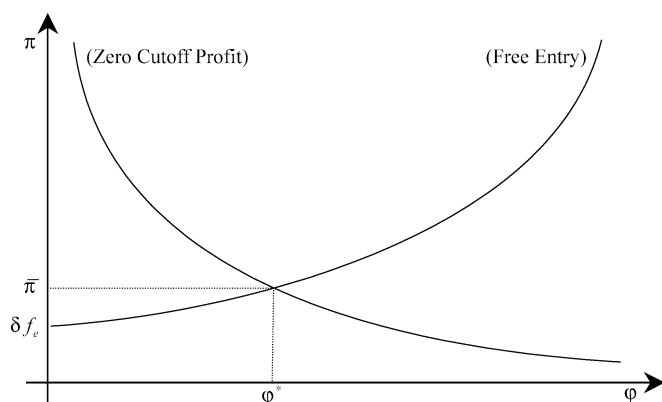


FIGURE 1.—Determination of the equilibrium cutoff φ^* and average profit $\bar{\pi}$.

uniqueness of the equilibrium φ^* and $\bar{\pi}$, which is graphically represented in Figure 1.¹⁵

In a stationary equilibrium, the aggregate variables must also remain constant over time. This requires a mass M_e of new entrants in every period, such that the mass of successful entrants, $p_{in}M_e$, exactly replaces the mass δM of incumbents who are hit with the bad shock and exit: $p_{in}M_e = \delta M$. The equilibrium distribution of productivity $\mu(\varphi)$ is not affected by this simultaneous entry and exit since the successful entrants and failing incumbents have the same distribution of productivity levels. The labor used by these new entrants for investment purposes must, of course, be reflected in the accounting for aggregate labor L , and affects the aggregate labor available for production: $L = L_p + L_e$ where L_p and L_e represent, respectively, the aggregate labor used for production and investment (by new entrants). Aggregate payments to production workers L_p must match the difference between aggregate revenue and profit: $L_p = R - \Pi$ (this is also the labor market clearing condition for production workers). The market clearing condition for investment workers requires $L_e = M_e f_e$. Using the aggregate stability condition, $p_{in}M_e = \delta M$, and the free entry condition, $\bar{\pi} = \delta f_e / [1 - G(\varphi^*)]$, L_e can be written:

$$L_e = M_e f_e = \frac{\delta M}{p_{in}} f_e = M \bar{\pi} = \Pi.$$

Thus, aggregate revenue $R = L_p + \Pi = L_p + L_e$ must also equal the total payments to labor L and is therefore exogenously fixed by this index of country

¹⁵The ZCP curve need not be decreasing everywhere as represented in the graph. However, it will monotonically decrease from infinity to zero for $\varphi^* \in (0, +\infty)$ as shown in the graph if $g(\varphi)$ belongs to one of several common families of distributions: lognormal, exponential, gamma, Weibul, or truncations on $(0, +\infty)$ of the normal, logistic, extreme value, or Laplace distributions. (A sufficient condition is that $g(\varphi)\varphi/[1 - G(\varphi)]$ be increasing to infinity on $(0, +\infty)$.)

size.¹⁶ The mass of producing firms in any period can then be determined from the average profit level using

$$(13) \quad M = \frac{R}{\bar{r}} = \frac{L}{\sigma(\bar{\pi} + f)}.$$

This, in turn, determines the equilibrium price index $P = M^{1/(1-\sigma)} p(\tilde{\varphi}) = M^{1/(1-\sigma)} / \rho \tilde{\varphi}$, which completes the characterization of the unique stationary equilibrium in the closed economy.

4.1. Analysis of the Equilibrium

All the firm-level variables—the productivity cutoff φ^* and average $\tilde{\varphi}$, and the average firm profit $\bar{\pi}$ and revenue \bar{r} —are independent of the country size L . As indicated by (13), the mass of firms increases proportionally with country size, although the distribution of firm productivity levels $\mu(\varphi)$ remains unchanged. Welfare per worker, given by

$$(14) \quad W = P^{-1} = M^{\frac{1}{\sigma-1}} \rho \tilde{\varphi},$$

is higher in a larger country due only to increased product variety. This influence of country size on the determination of aggregate variables is identical to that derived by Krugman (1980) with representative firms. Once $\tilde{\varphi}$ and $\bar{\pi}$ are determined, the aggregate outcome predicted by this model is identical to one generated by an economy with representative firms who share the same productivity level $\tilde{\varphi}$ and profit level $\bar{\pi}$. On the other hand, this model with heterogeneous firms explains how the aggregate productivity level $\tilde{\varphi}$ and the average firm profit level $\bar{\pi}$ are endogenously determined and how both can change in response to various shocks. In particular, a country's production technology (referenced by the distribution $g(\varphi)$) need not change in order to induce changes in aggregate productivity. In the following sections, I argue that the exposure of a country to trade creates precisely the type of shock that induces reallocations between firms and generates increases in aggregate productivity. These results cannot be explained by representative firm models where the aggregate productivity level is exogenously given as the productivity level common to all firms. Changes in aggregate productivity can then only result from changes in firm level technology and not from reallocations.

¹⁶It is important to emphasize that this result is not a direct consequence of aggregation and market clearing conditions: it is a property of the model's stationary equilibrium. Aggregate income need not necessarily equal the payments to all workers, since there may be some investment income derived from the financing of new entrants. Each new entrant raises the capital f_e , which provides a random return of $\pi(\varphi)$ (if $\varphi \geq \varphi^*$) or zero (if $\varphi < \varphi^*$) in every period. In equilibrium, the aggregate return Π equals the aggregate investment cost L_e in every period—so there is no net investment income (this would not be the case with a positive time discount factor).

5. OVERVIEW AND ASSUMPTIONS OF THE OPEN ECONOMY MODEL

I now examine the impact of trade in a world (or trade bloc) that is composed of countries whose economies are of the type that was previously described. When there are no additional costs associated with trade, then trade allows the individual countries to replicate the outcome of the integrated world economy. Trade then provides the same opportunities to an open economy as would an increase in country size to a closed economy. As was previously discussed, an increase in country size has no effect on firm level outcomes. The transition to trade will thus not affect any of the firm level variables: The same number of firms in each country produce at the same output levels and earn the same profits as they did in the closed economy. All firms in a given country divide their sales between domestic and foreign consumers, based on the size of their country relative to the integrated world economy. Thus, in the absence of any costs to trade, the existence of firm heterogeneity does not affect the impact of trade. This impact is identical to the one described by Krugman (1980) with representative firms: Although firms are not affected by the transition to trade, consumers enjoy welfare gains driven by the increase in product variety.¹⁷

On the other hand, there is mounting evidence that firms wishing to export not only face per-unit costs (such as transport costs and tariffs), but also—critically—face some fixed costs that do not vary with export volume. Interviews with managers making export decisions confirm that firms in differentiated product industries face significant fixed costs associated with the entry into export markets (see Roberts and Tybout (1977b)): A firm must find and inform foreign buyers about its product and learn about the foreign market. It must then research the foreign regulatory environment and adapt its product to ensure that it conforms to foreign standards (which include testing, packaging, and labeling requirements). An exporting firm must also set up new distribution channels in the foreign country and conform to all the shipping rules specified by the foreign customs agency. Although some of these costs cannot be avoided, others are often manipulated by governments in order to erect

¹⁷The irrelevance of firm heterogeneity for the impact of trade is not just a consequence of negligible trade costs. The assumption of an exogenously fixed elasticity of substitution between varieties also plays a significant role in this result. The presence of heterogeneity (even in the absence of trade costs) plays a significant role in determining the impact of trade once this assumption is dropped. In a separate appendix (available upon request to the author), the current model is modified by allowing the elasticity of substitution to endogenously increase with product variety. This link between trade and the elasticity of substitution was studied by Krugman (1979) with representative firms. In the context of the current model, the appendix shows how the size of the economy then affects the aggregate productivity level and the skewness of market shares and profits across firms with different productivity levels. Larger economies have higher aggregate productivity levels—even though they have the same firm level technology index by $g(\varphi)$. Therefore, even in the absence of trade costs, trade increases the size of the “world” economy and induces reallocations of market shares and profits towards more productive firms and generates an aggregate productivity gain.

non-tariff barriers to trade. Regardless of their origin, these costs are most appropriately modeled as independent of the firm's export volume decision.¹⁸

When there is uncertainty concerning the export market, the timing and sunk nature of the costs become quite relevant for the export decision (most of the previously mentioned costs must be sunk prior to entry into the export market). The strong and robust empirical correlations at the firm level between export status and productivity suggest that the export market entry decision occurs after the firm gains knowledge of its productivity, and hence that uncertainty concerning the export markets is not predominantly about productivity (as is the uncertainty prior to entry into the industry). I therefore assume that a firm who wishes to export must make an initial fixed investment, but that this investment decision occurs after the firm's productivity is revealed. For simplicity, I do not model any additional uncertainty concerning the export markets. The per-unit trade costs are modeled in the standard iceberg formulation, whereby $\tau > 1$ units of a good must be shipped in order for 1 unit to arrive at destination.

Although the size of a country relative to the rest of the world (which constitutes its trading partners) is left unrestricted, I do assume that the world (or trading group) is comprised of some number of identical countries. This assumption is made in order to ensure factor price equalization across countries and hence focus the analysis on firm selection effects that are independent of wage differences.¹⁹ In this model with trade costs, size differences across countries will induce differences in equilibrium wage levels. These wage differences then generate further firm selection effects and aggregate productivity differences across countries.²⁰ I therefore assume that the economy under study can trade with $n \geq 1$ other countries (the world is then comprised of $n + 1 \geq 2$ countries). Firms can export their products to any country, although entry into each of these export markets requires a fixed investment cost of $f_{ex} > 0$ (measured in units of labor). Regardless of export status, a firm still incurs the same overhead production cost f .

6. EQUILIBRIUM IN THE OPEN ECONOMY

The symmetry assumption ensures that all countries share the same wage, which is still normalized to one, and also share the same aggregate vari-

¹⁸The modeling of a fixed export cost is not new. Bernard and Jensen (1999a), Clerides, Lack, and Tybout (1998), Roberts and Tybout (1977a), and Roberts, Sullivan, and Tybout (1995) all introduce a fixed export cost into the theoretical sections of their work in order to explain the self-selection of firms into the export market. However, these analyses are restricted to a partial equilibrium setting in which the distribution of firm productivity levels is fixed.

¹⁹As was previously mentioned, another way to abstract from endogenous relative wage movements when countries are asymmetric is to introduce a freely traded homogeneous good sector. See Helpman, Melitz, and Yeaple (2002) for an example incorporating this extension.

²⁰In these asymmetric equilibria with fixed export costs, large countries enjoy higher aggregate productivity, welfare, and wages relative to smaller countries.

ables. Each firm's pricing rule in its domestic market is given, as before, by $p_d(\varphi) = w/\rho\varphi = 1/\rho\varphi$. Firms who export will set higher prices in the foreign markets that reflect the increased marginal cost τ of serving these markets: $p_x(\varphi) = \tau/\rho\varphi = \tau p_d(\varphi)$. Thus, the revenues earned from domestic sales and export sales to any given country are, respectively, $r_d(\varphi) = R(P\rho\varphi)^{\sigma-1}$ and $r_x(\varphi) = \tau^{1-\sigma}r_d(\varphi)$, where R and P denote the aggregate expenditure and price index in every country. The balance of payments condition implies that R also represents the aggregate revenue of firms in any country, and hence aggregate income. The combined revenue of a firm, $r(\varphi)$, thus depends on its export status:

$$(15) \quad r(\varphi) = \begin{cases} r_d(\varphi) & \text{if the firm does not export,} \\ r_d(\varphi) + nr_x(\varphi) = (1 + n\tau^{1-\sigma})r_d(\varphi) & \text{if the firm exports to all countries.} \end{cases}$$

If some firms do not export, then there no longer exists an integrated world market for all goods. Even though the symmetry assumption ensures that all the characteristics of the goods available in every country are similar, the actual bundle of goods available will be different across countries: consumers in each country have access to goods (produced by the nonexporting firms) that are not available to consumers in any other country.

6.1. Firm Entry, Exit, and Export Status

All the exogenous factors affecting firm entry, exit, and productivity levels remain unchanged by trade. Prior to entry, firms face the same ex-ante distribution of productivity levels $g(\varphi)$ and probability δ of the bad shock. In a stationary equilibrium, any incumbent firm with productivity φ earns variable profits $r_x(\varphi)/\sigma$ in every period from its export sales to any given country. Since the export cost is assumed equal across countries, a firm will either export to all countries in every period or never export.²¹ Given that the export decision occurs after firms know their productivity φ , and since there is no additional export market uncertainty, firms are indifferent between paying the one time investment cost f_{ex} , or paying the amortized per-period portion of this cost $f_x = \delta f_{ex}$ in every period (as before, there is no additional time discounting other than the probability of the exit inducing shock δ). This per-period representation of the export cost is henceforth adopted for notational simplicity. In the stationary equilibrium, the aggregate labor resources used in every period

²¹The restriction that export costs are equal across countries can be relaxed. Some firms then export to some countries but not others—depending on these cost differences. This extension would also generate an increasing relationship between a firm's productivity and the number of its export destinations.

to cover the export costs do not depend on this choice of representation.²² The per-period profit flow of any exporting firm then reflects the per-period fixed cost f_x , which is incurred per export country.

Since no firm will ever export and not also produce for its domestic market,²³ each firm's profit can be separated into portions earned from domestic sales, $\pi_d(\varphi)$, and export sales per country, $\pi_x(\varphi)$, by accounting for the entire overhead production cost in domestic profit:

$$(16) \quad \pi_d(\varphi) = \frac{r_d(\varphi)}{\sigma} - f, \quad \pi_x(\varphi) = \frac{r_x(\varphi)}{\sigma} - f_x.$$

A firm who produces for its domestic market exports to all n countries if $\pi_x(\varphi) \geq 0$. Each firm's combined profit can then be written: $\pi(\varphi) = \pi_d(\varphi) + \max\{0, n\pi_x(\varphi)\}$. Similarly to the closed economy case, firm value is given by $v(\varphi) = \max\{0, \pi(\varphi)/\delta\}$, and $\varphi^* = \inf\{\varphi : v(\varphi) > 0\}$ identifies the cutoff productivity level for successful entry. Additionally, $\varphi_x^* = \inf\{\varphi : \varphi \geq \varphi^* \text{ and } \pi_x(\varphi) > 0\}$ now represents the cutoff productivity level for exporting firms. If $\varphi_x^* = \varphi^*$, then all firms in the industry export. In this case, the cutoff firm (with productivity level $\varphi^* = \varphi_x^*$) earns zero total profit ($\pi(\varphi^*) = \pi_d(\varphi^*) + n\pi_x(\varphi^*) = 0$) and nonnegative export profit ($\pi_x(\varphi^*) \geq 0$). If $\varphi_x^* > \varphi^*$, then some firms (with productivity levels between φ^* and φ_x^*) produce exclusively for their domestic market. These firms do not export as their export profits would be negative. They earn nonnegative profits exclusively from their domestic sales. The firms with productivity levels above φ_x^* earn positive profits from both their domestic and export sales. By their definition, the cutoff levels must then satisfy $\pi_d(\varphi^*) = 0$ and $\pi_x(\varphi_x^*) = 0$.

This partitioning of firms by export status will occur if and only if $\tau^{\sigma-1}f_x > f$: the trade costs relative to the overhead production cost must be above a threshold level. Note that, when there are no fixed export costs ($f_x = 0$), no level of variable cost $\tau > 1$ can induce this partitioning. However, a large enough fixed export cost $f_x > f$ will induce partitioning even when there are no variable trade costs. As the partitioning of firms by export status (within sectors) is empirically ubiquitous, I will henceforth assume that the combination of fixed and variable trade costs are high enough to generate partitioning, and therefore that $\tau^{\sigma-1}f_x > f$. Although the equilibrium where all firms export will not be

²²In one case, only the new entrants who export expend resources to cover the full investment cost f_{ex} . In the other case, all exporting firms expend resources to cover the smaller amortized portion of the cost $f_x = \delta f_{ex}$. In equilibrium, the ratio of new exporters to all exporters is δ (see Appendix), so the same aggregate labor resources are expended in either case.

²³A firm would earn strictly higher profits by also producing for its domestic market since the associated variable profit $r_d(\varphi)/\sigma$ is always positive and the overhead production cost f is already incurred.

formally derived, it exhibits several similar properties to the equilibrium with partitioning that will be highlighted.²⁴

Once again, the equilibrium distribution of productivity levels for incumbent firms, $\mu(\varphi)$, is determined by the ex-ante distribution of productivity levels, conditional on successful entry: $\mu(\varphi) = g(\varphi)/[1 - G(\varphi^*)] \forall \varphi \geq \varphi^*$. The ex-ante probability of successful entry is still identified by $p_{in} = 1 - G(\varphi^*)$. Furthermore, $p_x = [1 - G(\varphi_x^*)]/[1 - G(\varphi^*)]$ now represents the ex-ante probability that one of these successful firms will export. The ex-post fraction of firms that export must then also be represented by p_x . Let M denote the equilibrium mass of incumbent firms in any country. $M_x = p_x M$ then represents the mass of exporting firms while $M_t = M + nM_x$ represents the total mass of varieties available to consumers in any country (or alternatively, the total mass of firms competing in any country).

6.2. Aggregation

Using the same weighted average function defined in (9), let $\tilde{\varphi} = \tilde{\varphi}(\varphi^*)$ and $\tilde{\varphi}_x = \tilde{\varphi}(\varphi_x^*)$ denote the average productivity levels of, respectively, all firms and exporting firms only. The average productivity across all firms, $\tilde{\varphi}$, is based only on domestic market share differences between firms (as reflected by differences in the firms' productivity levels). If some firms do not export, then this average will not reflect the additional export shares of the more productive firms. Furthermore, neither $\tilde{\varphi}$ nor $\tilde{\varphi}_x$ reflect the proportion τ of output units that are "lost" in export transit. Let $\tilde{\varphi}_t$ be the weighted productivity average that reflects the combined market share of all firms and the output shrinkage linked to exporting. Again, using the weighted average function (9), this combined average productivity can be written:

$$\tilde{\varphi}_t = \left\{ \frac{1}{M_t} [M \tilde{\varphi}^{\sigma-1} + nM_x (\tau^{-1} \tilde{\varphi}_x)^{\sigma-1}] \right\}^{\frac{1}{\sigma-1}}.$$

By symmetry, $\tilde{\varphi}_t$ is also the weighted average productivity of all firms (domestic and foreign) competing in a single country (where the productivity of exporters is adjusted by the trade cost τ). As was the case in the closed economy, this productivity average plays an important role as it once again completely summarizes the effects of the distribution of productivity levels $\mu(\varphi)$ on the aggregate outcome. Thus, the aggregate price index P , expenditure level R ,

²⁴Even when there is no partitioning of firms by export status, the opening of the economy to trade will still induce reallocations and distributional changes among the heterogeneous firms—so long as the fixed export costs are positive. In the absence of such costs (given any level of per-unit costs τ), opening to trade will not induce any distributional changes among firms, and heterogeneity will not play an important role.

and welfare per worker W in any country can then be written as functions of only the productivity average $\tilde{\varphi}_t$ and the number of varieties consumed M_t :²⁵

$$(17) \quad \begin{aligned} P &= M_t^{\frac{1}{1-\sigma}} p(\tilde{\varphi}_t) = M_t^{\frac{1}{1-\sigma}} \frac{1}{\rho \tilde{\varphi}_t}, & R &= M_t r_d(\tilde{\varphi}_t), \\ W &= \frac{R}{L} M_t^{\frac{1}{\sigma-1}} \rho \tilde{\varphi}_t. \end{aligned}$$

By construction, the productivity averages $\tilde{\varphi}$ and $\tilde{\varphi}_x$ can also be used to express the average profit and revenue levels across different groups of firms: $r_d(\tilde{\varphi})$ and $\pi_d(\tilde{\varphi})$ represent the average revenue and profit earned by domestic firms from sales in their own country. Similarly, $r_x(\tilde{\varphi}_x)$ and $\pi_x(\tilde{\varphi}_x)$ represent the average export revenue and profit (to any given country) across all domestic firms who export. The overall average—across all domestic firms—of combined revenue, \bar{r} , and profit, $\bar{\pi}$ (earned from both domestic and export sales), are then given by

$$(18) \quad \bar{r} = r_d(\tilde{\varphi}) + p_x n r_x(\tilde{\varphi}_x), \quad \bar{\pi} = \pi_d(\tilde{\varphi}) + p_x n \pi_x(\tilde{\varphi}_x).$$

6.3. Equilibrium Conditions

As in the closed economy equilibrium, the zero cutoff profit condition will imply a relationship between the average profit per firm $\bar{\pi}$ and the cutoff productivity level φ^* (see (10)):

$$\begin{aligned} \pi_d(\varphi^*) &= 0 &\iff & \pi_d(\tilde{\varphi}) = f k(\varphi^*), \\ \pi_x(\varphi_x^*) &= 0 &\iff & \pi_x(\tilde{\varphi}_x) = f_x k(\varphi_x^*), \end{aligned}$$

where $k(\varphi) = [\tilde{\varphi}(\varphi)/\varphi]^{1-\sigma} - 1$ as was previously defined. The zero cutoff profit condition also implies that φ_x^* can be written as a function of φ^* :

$$(19) \quad \frac{r_x(\varphi_x^*)}{r_d(\varphi^*)} = \tau^{1-\sigma} \left(\frac{\varphi_x^*}{\varphi^*} \right)^{\sigma-1} = \frac{f_x}{f} \iff \varphi_x^* = \varphi^* \tau \left(\frac{f_x}{f} \right)^{\frac{1}{\sigma-1}}.$$

Using (18), $\bar{\pi}$ can therefore be expressed as a function of the cutoff level φ^* :

$$(20) \quad \begin{aligned} \bar{\pi} &= \pi_d(\tilde{\varphi}) + p_x n \pi_x(\tilde{\varphi}_x) \\ &= f k(\varphi^*) + p_x n f_x k(\varphi_x^*) \quad (\text{ZCP}), \end{aligned}$$

where φ_x^* , and hence p_x , are implicitly defined as functions of φ^* using (19). Equation (20) thus identifies the new zero cutoff profit condition for the open economy.

²⁵In other words, the aggregate equilibrium in any country is identical to one with M_t representative firms that all share the same productivity level $\tilde{\varphi}_t$.

As before, $\bar{v} = \sum_{t=0}^{\infty} (1 - \delta)^t \bar{\pi} = \bar{\pi} / \delta$ represents the present value of the average profit flows and $v_e = p_{in} \bar{v} - f_e$ yields the net value of entry. The free entry condition thus remains unchanged: $v_e = 0$ if and only if $\bar{\pi} = \delta f_e / p_{in}$. Regardless of profit differences across firms (based on export status), the expected value of future profits, in equilibrium, must equal the fixed investment cost.

6.4. Determination of the Equilibrium

As in the closed economy case, the free entry condition and the new zero cut-off profit condition identify a unique φ^* and $\bar{\pi}$: the new ZCP curve still cuts the FE curve only once from above (see Appendix for proof). The equilibrium φ^* , in turn, determines the export productivity cutoff φ_x^* as well as the average productivity levels $\tilde{\varphi}$, $\tilde{\varphi}_x$, $\tilde{\varphi}_t$, and the ex-ante successful entry and export probabilities p_{in} and p_x . As was the case in the closed economy equilibrium, the free entry condition and the aggregate stability condition, $p_{in} M_e = \delta M$, ensure that the aggregate payment to the investment workers L_e equals the aggregate profit level Π . Thus, aggregate revenue R remains exogenously fixed by the size of the labor force: $R = L$. Once again, the average firm revenue is determined by the ZCP and FE conditions: $\bar{r} = r_d(\tilde{\varphi}) + p_x n r_x(\tilde{\varphi}_x) = \sigma(\bar{\pi} + f + p_x n f_x)$. This pins down the equilibrium mass of incumbent firms,

$$(21) \quad M = \frac{R}{\bar{r}} = \frac{L}{\sigma(\bar{\pi} + f + p_x n f_x)}.$$

In turn, this determines the mass of variety available in every country, $M_t = (1 + n p_x) M$, and their price index $P = M_t^{1/(1-\sigma)} / \rho \tilde{\varphi}_t$ (see (17)). Almost all of these equilibrium conditions also apply to the case where all firms export. The only difference is that $\varphi_x^* = \varphi^*$ (and hence $p_x = 1$) and (19) no longer holds.

7. THE IMPACT OF TRADE

The result that the modeling of fixed export costs explains the partitioning of firms by export status and productivity level is not exactly earth-shattering. This can be explained quite easily within a simple partial equilibrium model with a fixed distribution of firm productivity levels. On the other hand, such a model would be ill-suited to address several important questions concerning the impact of trade in the presence of export market entry costs and firm heterogeneity: What happens to the range of firm productivity levels? Do all firms benefit from trade or does the impact depend on a firm's productivity? How is aggregate productivity and welfare affected? The current model is much better suited to address these questions, which are answered in the following sections. The current section analyzes the effects of trade by contrasting the closed and open economy equilibria. The following section then studies the impact of incremental trade liberalization, once the economy is open. All of the following

analyses rely on comparisons of steady state equilibria and should therefore be interpreted as capturing the long run consequences of trade.

Let φ_a^* and $\bar{\varphi}_a$ denote the cutoff and average productivity levels in autarky. I use the notation of the previous section for all variables and functions pertaining to the new open economy equilibrium. As was previously mentioned, the FE condition is identical in both the closed and open economy. Inspection of the new ZCP condition in the open economy (20) relative to the one in the closed economy (12) immediately reveals that the ZCP curve shifts up: the exposure to trade induces an increase in the cutoff productivity level ($\varphi^* > \varphi_a^*$) and in the average profit per firm. The least productive firms with productivity levels between φ_a^* and φ^* can no longer earn positive profits in the new trade equilibrium and therefore exit. Another selection process also occurs since only the firms with productivity levels above φ_x^* enter the export markets. This export market selection effect and the domestic market selection effect (of firms out of the industry) both reallocate market shares towards more efficient firms and contribute to an aggregate productivity gain.²⁶

Inspection of the equations for the equilibrium number of firms ((13) and (21)) reveals that $M < M_a$ where M_a represents the number of firms in autarky.²⁷ Although the number of firms in a country decreases after the transition to trade, consumers in the country still typically enjoy greater product variety ($M_t = (1 + np_x)M > M_a$). That is, the decrease in the number of domestic firms following the transition to trade is typically dominated by the number of new foreign exporters. It is nevertheless possible, when the export costs are high, that these foreign firms replace a larger number of domestic firms (if the latter are sufficiently less productive). Although product variety then impacts negatively on welfare, this effect is dominated by the positive contribution of the aggregate productivity gain. Trade—even though it is costly—always generates a welfare gain (see Appendix for proof).

7.1. The Reallocation of Market Shares and Profits Across Firms

I now examine the effects of trade on firms with different productivity levels. To do this, I contrast the performance of a firm with productivity $\varphi \geq \varphi_a^*$ before and after the transition to trade. Let $r_a(\varphi) > 0$ and $\pi_a(\varphi) \geq 0$ denote the firm's revenue and profit in autarky. Recall that, in both the closed and open economy equilibria, the aggregate revenue of domestic firms is exogenously given by the country's size ($R = L$). Hence, $r_a(\varphi)/R$ and $r(\varphi)/R$ represent the firm's market share (within the domestic industry) in autarky and in the equilibrium with trade. Additionally, in this equilibrium with trade, $r_d(\varphi)/R$ represents the

²⁶Because $\bar{\varphi}_t$ factors in the output lost in export transit (from τ), it is possible for $\bar{\varphi}_t$ to be lower than $\bar{\varphi}_a$ when τ is high and f_x is low. It is shown in the Appendix that any productivity average that is based on a firm's output "at the factory gate" must be higher in the open economy.

²⁷Recall that the average profit $\bar{\pi}$ must be higher in the open economy equilibrium.

firm's share of its domestic market (since R also represents aggregate consumer expenditure in the country). The impact of trade on this firm's market share can be evaluated using the following inequalities (see Appendix):

$$r_d(\varphi) < r_a(\varphi) < r_d(\varphi) + nr_x(\varphi) \quad \forall \varphi \geq \varphi^*.$$

The first part of the inequality indicates that all firms incur a loss in domestic sales in the open economy. A firm who does not export then also incurs a total revenue loss. The second part of the inequality indicates that a firm who exports more than makes up for its loss of domestic sales with export sales and increases its total revenues. Thus, a firm who exports increases its share of industry revenues while a firm who does not export loses market share. (The market share of the least productive firms in the autarky equilibrium—with productivity between φ_a^* and φ^* —drops to zero as these firms exit.)

Now consider the change in profit earned by a firm with productivity φ . If the firm does not export in the open economy, it must incur a profit loss, since its revenue, and hence variable profit, is now lower. The direction of the profit change for an exporting firm is not immediately clear since it involves a trade-off between the increase in total revenue (and hence variable profit) and the increase in fixed cost due to the additional export cost. For such a firm ($\varphi \geq \varphi_x^*$), this profit change can be written:²⁸

$$\begin{aligned} \Delta\pi(\varphi) &= \pi(\varphi) - \pi_a(\varphi) = \frac{1}{\sigma}([r_d(\varphi) + nr_x(\varphi)] - r_a(\varphi)) - nf_x \\ &= \varphi^{\sigma-1} f \left[\frac{1 + n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} - \frac{1}{(\varphi_a^*)^{\sigma-1}} \right] - nf_x, \end{aligned}$$

where the term in the bracket must be positive since $r_d(\varphi) + nr_x(\varphi) > r_a(\varphi)$ for all $\varphi > \varphi^*$. The profit change, $\Delta\pi(\varphi)$, is thus an increasing function of the firm's productivity level φ . In addition, this change must be negative for the exporting firm with the cutoff productivity level φ_x^* .²⁹ Therefore, firms are partitioned by productivity into groups that gain and lose profits. Only a subset of the more productive firms who export gain from trade. Among firms in this group, the profit gain increases with productivity. Figure 2 graphically represents the changes in revenue and profits driven by trade. The exposure to trade thus generates a type of Darwinian evolution within an industry that was described in the introduction: the most efficient firms thrive and grow—they export and increase both their market share and profits. Some less efficient firms still export and increase their market share but incur a profit loss. Some even less efficient firms remain in the industry but do not export and incur losses of both market share and profit. Finally, the least efficient firms are driven out of the industry.

²⁸ Using $r_d(\varphi) = (\varphi/\varphi^*)^{\sigma-1}\sigma f$ and $r_a(\varphi) = (\varphi/\varphi_a^*)^{\sigma-1}\sigma f$.

²⁹ Since $\pi_x(\varphi_x^*) = 0$ and $r_d(\varphi_x^*) < r_a(\varphi_x^*)$.

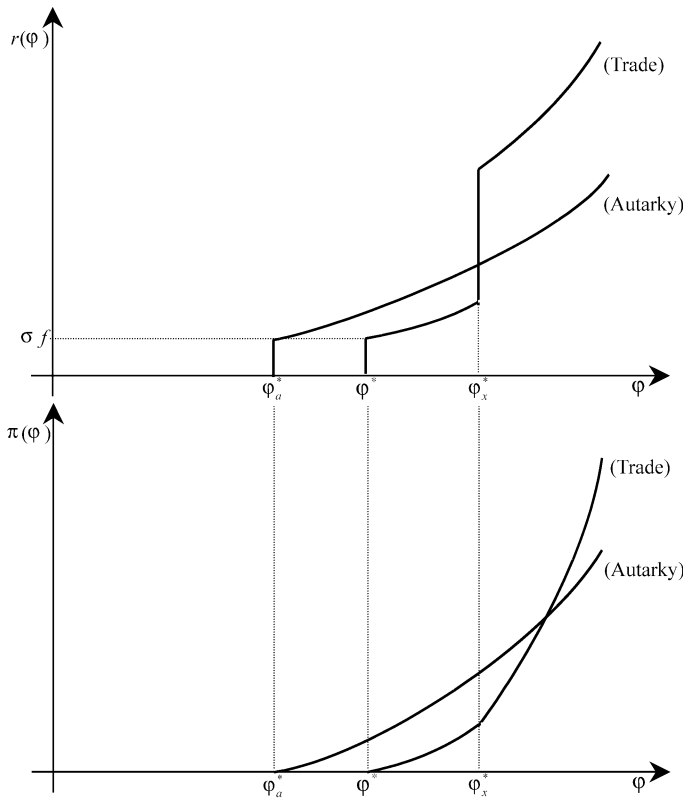


FIGURE 2.—The reallocation of market shares and profits.

7.2. Why Does Trade Force the Least Productive Firms to Exit?

There are two potential channels through which trade can affect the distribution of surviving firms. The first to come to mind is the increase in product market competition associated with trade: firms face an increasing number of competitors; furthermore the new foreign competitors, on average, are more productive than the domestic firms. However, this channel is not operative in the current model due to the peculiar and restrictive property of monopolistic competition under C.E.S. preferences: the price elasticity of demand for any variety does not respond to changes in the number or prices of competing varieties. Thus, in the current model, all the effects of trade on the distribution of firms are channeled through a second mechanism operating through the domestic factor market where firms compete for a common source of labor: when entry into new export markets is costly, exposure to trade offers new profit opportunities only to the more productive firms who can “afford” to cover the entry cost. This also induces more entry as prospective firms respond

to the higher potential returns associated with a good productivity draw. The increased labor demand by the more productive firms and new entrants bids up the real wage and forces the least productive firms to exit.

The current model thus highlights a potentially important channel for the redistributive effects of trade within industries that operates through the exposure to export markets. Recent work by Bernard and Jensen (1999b) suggests that this channel substantially contributes to U.S. productivity increases within manufacturing industries. Nevertheless, the model should also be interpreted with caution as it precludes another potentially important channel for the effects of trade, which operates through increases in import competition.

8. THE IMPACT OF TRADE LIBERALIZATION

The preceding analysis compared the equilibrium outcomes of an economy undergoing a massive change in trade regime from autarky to trade. Very few, if any, of the world's current economies can be considered to operate in an autarky environment. It is therefore reasonable to ask whether an *increase* in the exposure of an economy to trade will induce the same effects as were previously described for the transition of an economy from autarky. The current model is well-suited to address several different mechanisms that would produce an increase in trade exposure and plausibly correspond to observed decreases in trade costs over time or some specific policies to liberalize trade. The effects of three such mechanisms are investigated: an increase in the number of available trading partners (resulting, for example, from the incorporation of additional countries into a trade bloc) and a decrease in either the fixed or variable trade costs (resulting either from decreases in real cost levels or from multilateral agreements to reduce tariffs or nontariff barriers to trade). These three scenarios involve comparative statics of the open economy equilibrium with respect to n , τ , and f_x . The main impact of the transition from autarky to trade was an increase in aggregate productivity and welfare generated by a reallocation of market shares towards more productive firms (where the least productive firms are forced to exit). I will show that increases in the exposure to trade occurring through any of these mechanisms will generate very similar results: in all cases, the exposure to trade will force the least productive firms to exit and will reallocate market shares from less productive to more productive firms. The increased exposure to trade will also always deliver welfare gains.³⁰

8.1. *Increase in the Number of Trading Partners*

I first investigate the effects of an increase in n . Throughout the comparative static analyses, I use the notation of the open economy equilibrium to describe

³⁰Formal derivations of all the comparative statics are relegated to the Appendix.

the old equilibrium with n countries. I then add primes (') to all variables and functions when they pertain to the new equilibrium with $n' > n$ countries.

Inspection of equations (20) and (19) defining the new zero cutoff profit condition (as a function of the domestic cutoff φ^*) reveals that the ZCP curve will shift up and therefore that both cutoff productivity levels increase with n : $\varphi^{*'} > \varphi^*$ and $\varphi_x^{*'} > \varphi_x^*$. The increase in the number of trading partners thus forces the least productive firms to exit. As was the case with the transition from autarky, the increased exposure to trade forces all firms to relinquish a portion of their share of their domestic market: $r_d'(\varphi) < r_d(\varphi)$, $\forall \varphi \geq \varphi^*$. The less productive firms who do not export (with $\varphi < \varphi_x^{*'}$) thus incur a revenue and profit loss—and the least productive among them exit.³¹ Again, as was the case with the transition from autarky, the firms who export (with $\varphi \geq \varphi_x^{*'}$) more than make up for their loss of domestic sales with their sales to the new export markets and increase their combined revenues: $r_d'(\varphi) + n'r_x'(\varphi) > r_d(\varphi) + nr_x(\varphi)$. Some of these firms nevertheless incur a decrease in profits due to the new fixed export costs, but the most productive firms among this group also enjoy an increase in profits (which is increasing with the firms' productivity level). Thus, both market shares and profits are reallocated towards the more efficient firms. As was the case for the transition from autarky, this reallocation of market shares generates an aggregate productivity gain and an increase in welfare.³²

8.2. Decrease in Trade Costs

A decrease in the variable trade cost τ will induce almost identical effects to those just described for the increase in trading partners. The decrease from τ to $\tau' < \tau$ (again I use primes to reference all variables and functions in the new equilibrium) will shift up the ZCP curve and induce an increase in the cutoff productivity level $\varphi^{*'} > \varphi^*$. The only difference is that the new export cutoff productivity level $\varphi_x^{*'}$ will now be below φ_x^* . As before, the increased exposure to trade forces the least productive firms to exit, but now also generates entry of new firms into the export market (who did not export with the higher τ). The direction of the reallocation of market shares and profits will be identical to those previously described: all firms lose a portion of their domestic sales,

³¹There is a transitional issue associated with the exporting status of firms with productivity levels between φ_x^* and $\varphi_x^{*'}$. The loss of export sales to any given country—from $r_x(\varphi)$ down to $r_x'(\varphi)$ —is such that firms entering with productivity levels between φ_x^* and $\varphi_x^{*'}$ will not export as the lower variable profit $r_x'(\varphi)/\sigma$ no longer covers the amortized portion of the entry cost f_x . On the other hand, incumbent firms with productivity levels in this range have already incurred the sunk export entry cost and have no reason to exit the export markets until they are hit with the bad shock and exit the industry. Eventually, all these incumbent firms exit and no firm with a productivity level in that range will export once the new steady state equilibrium is attained.

³²As pointed out in footnote 26, the productivity average must be based on a firm's output "at the factory gate."

so that the firms who do not export incur both a market share and profit loss. The more productive firms who export more than make up for their loss of domestic sales with increased export sales, and the most productive firms among this group also increase their profits. As before, the exit of the least productive firms and the market share increase of the most productive firms both contribute to an aggregate productivity gain and an increase in welfare.³³

A decrease in the fixed export market entry cost f_x induces similar changes in the cutoff levels as the decrease in τ . The increased exposure to trade forces the least productive firms to exit (φ^* rises) and generates entry of new firms into the export market (φ_x^* decreases). These selection effects both contribute to an aggregate productivity increase if the new exporters are more productive than the average productivity level. Although the less productive firms who do not export incur both a market share and profit loss, the market share and profit reallocations towards the more productive firms, in this case, will not be similar to those for the previous two cases: the decrease in f_x will not increase the combined market share or profit of any firm that already exported prior to the change in f_x —only new exporters increase their combined sales. However, as in the previous two cases, welfare is higher in the new steady state equilibrium. Both types of trade cost decreases described above also help to explain another empirical feature, reported by Roberts, Sullivan, and Tybout (1995), that some export booms are driven by the entry of new firms into the export markets.³⁴

9. CONCLUSION

This paper has described and analyzed a new transmission channel for the impact of trade on industry structure and performance. Since this channel works through intra-industry reallocations across firms, it can only be studied within an industry model that incorporates firm level heterogeneity. Recent empirical work has highlighted the importance of this channel for understanding and explaining the effects of trade on firm and industry performance.

The paper shows how the existence of export market entry costs drastically affects how the impact of trade is distributed across different types of firms. The induced reallocations between these different firms generate changes in a country's aggregate environment that cannot be explained by a model based on representative firms. On one hand, the paper shows that the existence of such costs to trade does not affect the welfare-enhancing properties of trade: one of the most robust results of this paper is that increases in a country's exposure to trade lead to welfare gains. On the other hand, the paper shows how the export costs significantly alter the distribution of the gains from trade

³³See footnote 32.

³⁴Over half of the substantial export growth in Colombian and Mexican manufacturing sectors was generated by the entry of firms into the export markets.

across firms. In fact, only a portion of the firms—the more efficient ones—reap benefits from trade in the form of gains in market share and profit. Less efficient firms lose both. The exposure to trade, or increases in this exposure, force the least efficient firms out of the industry. These trade-induced reallocations towards more efficient firms explain why trade may generate aggregate productivity gains without necessarily improving the productive efficiency of individual firms.

Although this model mainly highlights the long-run benefits associated with the trade-induced reallocations within an industry, the reallocation of these resources also obviously entails some short-run costs. It is therefore important to have a model that can predict the impact of trade policy on inter-firm reallocations in order to design accompanying policies that would address issues related to the transition towards a new regime. These policies could help palliate the transitional costs while taking care not to hinder the reallocation process. Of course, the model also clearly indicates that policies that hinder the reallocation process or otherwise interfere with the flexibility of the factor markets may delay or even prevent a country from reaping the full benefits from trade.

Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138, U.S.A.; CEPR; and NBER.

Manuscript received April, 2002; final revision received April, 2003.

APPENDIX A: AGGREGATION CONDITIONS IN THE CLOSED ECONOMY

Using the definition of $\tilde{\varphi}$ in (7), the aggregation conditions relating the aggregate variables to the number of firms M and aggregate productivity level $\tilde{\varphi}$ are derived:

$$\begin{aligned} Q &= \left[\int_0^\infty q(\varphi)^\rho M \mu(\varphi) d\varphi \right]^{1/\rho} && \text{(by definition of } Q \equiv U) \\ &= \left[\int_0^\infty q(\tilde{\varphi})^\rho \left(\frac{\varphi}{\tilde{\varphi}} \right)^{\sigma\rho} M \mu(\varphi) d\varphi \right]^{1/\rho} \\ &= M^{1/\rho} q(\tilde{\varphi}), \end{aligned}$$

and using the definition of R and Π as aggregate revenue and profit,

$$\begin{aligned} R &= \int_0^\infty r(\varphi) M \mu(\varphi) d\varphi = \int_0^\infty r(\tilde{\varphi}) \left(\frac{\varphi}{\tilde{\varphi}} \right)^{\sigma-1} M \mu(\varphi) d\varphi = M r(\tilde{\varphi}), \\ \Pi &= \int_0^\infty \pi(\varphi) M \mu(\varphi) d\varphi = \frac{1}{\sigma} \int_0^\infty r(\varphi) M \mu(\varphi) d\varphi - M f = M \left[\frac{r(\tilde{\varphi})}{\sigma} - f \right] = M \pi(\tilde{\varphi}). \end{aligned}$$

APPENDIX B: CLOSED ECONOMY EQUILIBRIUM

B.1. Existence and Uniqueness of the Equilibrium Cutoff Level φ^*

Following is a proof that the FE and ZCP conditions in (12) identify a unique cutoff level φ^* and that the ZCP curve cuts the FE curve from above in (φ, π) space. I do this by showing that

$[1 - G(\varphi)]k(\varphi)$ is monotonically decreasing from infinity to zero on $(0, \infty)$. (This is a sufficient condition for both properties.) Recall that $k(\varphi) = [\tilde{\varphi}(\varphi)/\varphi]^{\sigma-1} - 1$ where

$$(B.1) \quad \tilde{\varphi}(\varphi)^{\sigma-1} = \frac{1}{1 - G(\varphi)} \int_{\varphi}^{\infty} \xi^{\sigma-1} g(\xi) d\xi$$

as defined in (9). Thus,

$$\begin{aligned} k'(\varphi) &= \frac{g(\varphi)}{1 - G(\varphi)} \left[\left(\frac{\tilde{\varphi}(\varphi)}{\varphi} \right)^{\sigma-1} - 1 \right] - \left(\frac{\tilde{\varphi}(\varphi)}{\varphi} \right)^{\sigma-1} \frac{\sigma-1}{\varphi} \\ &= \frac{k(\varphi)g(\varphi)}{1 - G(\varphi)} - \frac{(\sigma-1)[k(\varphi)+1]}{\varphi}. \end{aligned}$$

Define

$$(B.2) \quad j(\varphi) = [1 - G(\varphi)]k(\varphi).$$

Its derivative and elasticity are given by

$$(B.3) \quad j'(\varphi) = -\frac{1}{\varphi}(\sigma-1)[1 - G(\varphi)][k(\varphi)+1] < 0,$$

$$(B.4) \quad \frac{j'(\varphi)\varphi}{j(\varphi)} = -(\sigma-1) \left(1 + \frac{1}{k(\varphi)} \right) < -(\sigma-1).$$

Since $j(\varphi)$ is nonnegative and its elasticity with respect to φ is negative and bounded away from zero, $j(\varphi)$ must be decreasing to zero as φ goes to infinity. Furthermore, $\lim_{\varphi \rightarrow 0} j(\varphi) = \infty$ since $\lim_{\varphi \rightarrow 0} k(\varphi) = \infty$. Therefore, $j(\varphi) = [1 - G(\varphi)]k(\varphi)$ decreases from infinity to zero on $(0, \infty)$.

APPENDIX C: OPEN ECONOMY EQUILIBRIUM

C.1. Aggregate Labor Resources Used to Cover the Export Costs

It was asserted in footnote 22 that the ratio of new exporters to all exporters was δ , and hence that the aggregate labor resources used to cover the export cost did not depend on its representation as either a one time sunk entry cost or a per-period fixed cost. As before, let M_e denote the mass of all new entrants. The ratio of new exporters to all exporters is then $p_x p_{in} M_e / p_x M$. This ratio must be equal to δ as the aggregate stability condition for the equilibrium ensures that $p_{in} M_e = \delta M$.

C.2. Existence and Uniqueness of the Equilibrium Cutoff Level φ^*

Following is a proof that the FE condition and the new ZCP condition in (20) identify a unique cutoff level φ^* and that this new ZCP curve cuts the FE curve from above in (φ, π) space. These conditions imply $\delta f_e / [1 - G(\varphi^*)] = f k(\varphi^*) + p_x n f_x k(\varphi^*)$, or

$$(C.1) \quad f j(\varphi^*) + n f_x j(\varphi^*) = \delta f_e,$$

where $\varphi_x^* = \tau(f_x/f)^{1/(\sigma-1)} \varphi^*$ is implicitly defined as a function of φ^* (see (19)). Since $j(\varphi)$ is decreasing from infinity to zero on $(0, \infty)$, the left-hand side in (C.1) must also monotonically decrease from infinity to zero on $(0, \infty)$. Therefore, (C.1) identifies a unique cutoff level φ^* and the new ZCP curve must cut the FE curve from above.

APPENDIX D: THE IMPACT OF TRADE

D.1. Welfare

Using (14), welfare per worker in autarky can be written as a function of the cutoff productivity level:³⁵

$$W_a = M_a^{\frac{1}{\sigma-1}} \rho \tilde{\varphi}_a = \rho \left(\frac{L}{\sigma f} \right)^{\frac{1}{\sigma-1}} \varphi_a^*.$$

Similarly, welfare in the open economy can also be written as a function of the new cutoff productivity level (see (17)):³⁶

$$(D.1) \quad W = M_t^{\frac{1}{\sigma-1}} \rho \tilde{\varphi}_t = \rho \left(\frac{L}{\sigma f} \right)^{\frac{1}{\sigma-1}} \varphi^*.$$

Since $\varphi^* > \varphi_a^*$, welfare in the open economy must be higher than in autarky: $W > W_a$.

D.2. Reallocations

PROOF THAT $r_d(\varphi) < r_a(\varphi) < r_d(\varphi) + nr_x(\varphi) = (1 + n\tau^{1-\sigma})r_d(\varphi)$: Recall that $r_a(\varphi) = (\varphi/\varphi_a^*)^{\sigma-1} \sigma f$ ($\forall \varphi \geq \varphi_a^*$) in autarky and that $r_d(\varphi) = (\varphi/\varphi^*)^{\sigma-1} \sigma f$ ($\forall \varphi \geq \varphi^*$) in the open economy equilibrium. This immediately yields $r_d(\varphi) < r_a(\varphi)$ since $\varphi^* > \varphi_a^*$. The second inequality is a direct consequence of another comparative static involving τ . It is shown in a following section that $(1 + n\tau^{1-\sigma})r_d(\varphi)$ decreases as τ increases. Since the autarky equilibrium is obtained as the limiting equilibrium as τ increases to infinity, $r_a(\varphi) = \lim_{\tau \rightarrow +\infty} r_d(\varphi) = \lim_{\tau \rightarrow +\infty} [(1 + n\tau^{1-\sigma})r_d(\varphi)]$. Therefore, $r_a(\varphi) < (1 + n\tau^{1-\sigma})r_d(\varphi)$ for any finite τ .

D.3. Aggregate Productivity

It was pointed out in the paper that aggregate productivity $\tilde{\varphi}_t$ in the open economy may not be higher than $\tilde{\varphi}_a$ due to the effect of the output loss incurred in export transit. It was then claimed that a productivity average based on a measure of output “at the factory gate” would always be higher in the open economy. Define

$$(D.2) \quad \Phi = h^{-1} \left(\frac{1}{R} \int_0^\infty r(\varphi) h(\varphi) g(\varphi) d\varphi \right)$$

as such an average where $h(\cdot)$ is any increasing function. The only condition imposed on this average involves the use of the firms’ combined revenues as weights.³⁷ Let $\Phi_a = h^{-1}((1/R) \int_0^\infty r_a(\varphi) \times h(\varphi) g(\varphi) d\varphi)$ represent this productivity average in autarky. Then Φ must be greater than Φ_a —for any increasing function $h(\cdot)$ —as the distribution $r(\varphi)g(\varphi)/R$ first order stochastically dominates the distribution $r_a(\varphi)g(\varphi)/R$: $\int_0^\varphi r(\xi)g(\xi)d\xi \leq \int_0^\varphi r_a(\xi)g(\xi)d\xi \forall \varphi$ (and the inequality is strict $\forall \varphi > \varphi_a^*$).³⁸

APPENDIX E: THE IMPACT OF TRADE LIBERALIZATION

E.1. Changes in the Cutoff Levels

These comparative statics are all derived from the equilibrium condition for the cutoff levels (C.1) and the implicit definition of φ_x^* as a function of φ^* in (19).

³⁵Using the relationship $(\tilde{\varphi}_a/\varphi_a^*)^{\sigma-1} = r(\tilde{\varphi}_a)/r(\varphi_a^*) = (R/M_a)/\sigma f = (L/M_a)/\sigma f$.

³⁶Using $(\tilde{\varphi}_t/\varphi^*)^{\sigma-1} = r_d(\tilde{\varphi}_t)/r_d(\varphi^*) = (R/M_t)/\sigma f = (L/M_t)/\sigma f$.

³⁷This is the standard way of computing industry productivity averages in empirical work.

³⁸This result is a direct consequence of the market share reallocation result.

Increase in n : Differentiating (C.1) with respect to n and using $\partial\varphi_x^*/\partial n = (\varphi_x^*/\varphi^*)\partial\varphi^*/\partial n$ from (19) yields

$$(E.1) \quad \frac{\partial\varphi^*}{\partial n} = \frac{-f_x\varphi^*j'(\varphi_x^*)}{f\varphi^*j'(\varphi^*) + nf_x\varphi_x^*j'(\varphi_x^*)}.$$

Hence $\partial\varphi^*/\partial n > 0$ and $\partial\varphi_x^*/\partial n > 0$ since $j'(\varphi) < 0 \forall \varphi$ (see (B.4)).

Decrease in τ : Differentiating (C.1) with respect to τ and using $\partial\varphi_x^*/\partial\tau = \varphi_x^*/\tau + (\varphi_x^*/\varphi^*)\partial\varphi^*/\partial\tau$ from (19) yields

$$(E.2) \quad \frac{\partial\varphi^*}{\partial\tau} = -\frac{\varphi^*}{\tau} \frac{nf_xj'(\varphi_x^*)\varphi_x^*}{f\varphi^*j'(\varphi^*) + nf_x\varphi_x^*j'(\varphi_x^*)} < 0$$

since $j'(\varphi) < 0 \forall \varphi$, and

$$\frac{\partial\varphi_x^*}{\partial\tau} = -\frac{fj'(\varphi^*)}{nf_xj'(\varphi_x^*)} \frac{\partial\varphi^*}{\partial\tau} > 0.$$

Decrease in f_x : Differentiating (C.1) with respect to f_x and using $\partial\varphi_x^*/\partial f_x = (\varphi_x^*/\varphi^*)\partial\varphi^*/\partial f_x + [1/(\sigma-1)]\varphi_x^*/f_x$ from (19) and $j'(\varphi_x^*)\varphi_x^* = -(\sigma-1)[j(\varphi_x^*) + 1 - G(\varphi_x^*)]$ from (B.2) and (B.4) yields

$$\frac{\partial\varphi^*}{\partial f_x} = \frac{n[1 - G(\varphi_x^*)]}{fj'(\varphi^*) + nf_xj'(\varphi_x^*)(\varphi_x^*/\varphi^*)} < 0$$

since $j'(\varphi) < 0 \forall \varphi$, and

$$\frac{\partial\varphi_x^*}{\partial f_x} = \frac{-1}{nf_xj'(\varphi_x^*)} \left[nj(\varphi_x^*) + fj'(\varphi^*) \frac{\partial\varphi^*}{\partial f_x} \right] > 0.$$

Welfare: Recall from (D.1) that welfare per worker is given by $W = \rho(L/\sigma f)^{1/(\sigma-1)}\varphi^*$. Welfare must therefore rise with increases in n and decreases in f_x or τ since all of these changes induce an increase in the cutoff productivity level φ^* .

E.2. Reallocations of Market Shares

Recall that $r_d(\varphi) = (\varphi/\varphi^*)^{\sigma-1}\sigma f$ ($\forall \varphi \geq \varphi^*$) in the new open economy equilibrium. $r_d(\varphi)$ therefore decreases with increases in n and decreases in f_x or τ since all of these changes induce an increase in the cutoff productivity level φ^* . Thus $r'_d(\varphi) < r_d(\varphi) \forall \varphi \geq \varphi^*$ whenever $n' > n$, $\tau' < \tau$, or $f'_x < f_x$ (since $\varphi^* > \varphi$).

The direction of the change in combined domestic and export sales, $r_d(\varphi) + nr_x(\varphi) = (1 + n\tau^{1-\sigma})r_d(\varphi)$, will depend on the direction of the change in $(1 + n\tau^{1-\sigma})/(\varphi^*)^{\sigma-1}$. It is therefore clear that a firm's combined sales will decrease in the same proportion as its domestic sales when f_x decreases since $1 + n\tau^{1-\sigma}$ will remain constant. On the other hand, it is now shown that these combined sales will *increase* when n increases or τ decreases as $(1 + n\tau^{1-\sigma})/(\varphi^*)^{\sigma-1}$ will then increase.

Increase in n : From (E.1),

$$\begin{aligned} \frac{\partial\varphi^*}{\partial n} \frac{1}{\varphi^*} &= -\left[\frac{f}{f_x} \frac{\varphi^*j'(\varphi^*)}{j(\varphi_x^*)} + n \frac{\varphi_x^*j'(\varphi_x^*)}{j(\varphi_x^*)} \right]^{-1} \\ &= -\left[\tau^{\sigma-1} \frac{(\varphi^*)^{\sigma-1}j(\varphi^*)}{(\varphi_x^*)^{\sigma-1}j(\varphi_x^*)} \frac{\varphi^*j'(\varphi^*)}{j(\varphi^*)} + n \frac{\varphi_x^*j'(\varphi_x^*)}{j(\varphi_x^*)} \right]^{-1} \quad (\text{using (19)}) \\ &< [(\sigma-1)(\tau^{\sigma-1} + n)]^{-1} \end{aligned}$$

since $-\varphi j'(\varphi)/j(\varphi) > \sigma - 1 \forall \varphi$ (see (B.4)) and $(\varphi^*)^{\sigma-1}j(\varphi^*)/[(\varphi_x^*)^{\sigma-1}j(\varphi_x^*)] > 1$.³⁹ Hence,

$$\frac{\partial \left[\frac{1+n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} \right]}{\partial n} = \frac{1+n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} \left[\frac{1}{\tau^{\sigma-1}+n} - (\sigma-1) \frac{\partial \varphi^*}{\partial n} \frac{1}{\varphi^*} \right] > 0.$$

Decrease in τ : From (E.2),

$$\begin{aligned} -\frac{\partial \varphi^*}{\partial \tau} \frac{\tau}{\varphi^*} &= \left[\frac{f}{nf_x} \frac{\varphi^{*'}(\varphi^*)}{\varphi_x^{*'}(\varphi_x^*)} + 1 \right]^{-1} \\ &= \left[\frac{f}{nf_x} \frac{[1-G(\varphi^*)][k(\varphi^*)+1]}{[1-G(\varphi_x^*)][k(\varphi_x^*)+1]} + 1 \right]^{-1} && \text{(using (B.3))} \\ &= \left[\frac{f}{nf_x} \left(\frac{\varphi_x^*}{\varphi^*} \right)^{\sigma-1} \frac{\int_{\varphi_x^*}^{\infty} \xi^{\sigma-1} g(\xi) d\xi}{\int_{\varphi_x^*}^{\infty} \xi^{\sigma-1} g(\xi) d\xi} + 1 \right]^{-1} && \text{(using (B.1))} \\ &= \left[\frac{\tau^{\sigma-1}}{n} \frac{\int_{\varphi_x^*}^{\infty} \xi^{\sigma-1} g(\xi) d\xi}{\int_{\varphi_x^*}^{\infty} \xi^{\sigma-1} g(\xi) d\xi} + 1 \right]^{-1} && \text{(using (19))} \\ &< \left[\frac{\tau^{\sigma-1}}{n} + 1 \right]^{-1} \end{aligned}$$

since $\int_{\varphi^*}^{\infty} \xi^{\sigma-1} g(\xi) d\xi / [\int_{\varphi_x^*}^{\infty} \xi^{\sigma-1} g(\xi) d\xi] > 1$ as $\varphi^* < \varphi_x^*$. Hence,

$$\begin{aligned} \frac{\partial \left[\frac{1+n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} \right]}{\partial \tau} &= \frac{1+n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} \left[\frac{(1-\sigma)n\tau^{1-\sigma}}{1+n\tau^{1-\sigma}} - (\sigma-1) \frac{\partial \varphi^*}{\partial \tau} \frac{\tau}{\varphi^*} \right] \\ &= \frac{1+n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} (\sigma-1) \left[-\frac{\partial \varphi^*}{\partial \tau} \frac{\tau}{\varphi^*} - \left(\frac{\tau^{\sigma-1}}{n} + 1 \right)^{-1} \right] \\ &< 0. \end{aligned}$$

E.3. Reallocations of Profits

Increase in n : All surviving firms who do not export (with $\varphi < \varphi_x^{*}$) must incur a profit loss since their profits from domestic sales decrease ($r'_d(\varphi) < r_d(\varphi)$) and those who would have exported previously (with the lower n) further lose any profits from exporting. Similarly, the firm with productivity level $\varphi = \varphi_x^{*}$ also incurs a profit loss (although the firm exports, it gains zero additional profits from doing so and still incurs the loss in domestic profits). The profit change for all exporting firms (with $\varphi \geq \varphi_x^{*}$) can be written:

$$\begin{aligned} \text{(E.3)} \quad \Delta \pi(\varphi) &= \pi'(\varphi) - \pi(\varphi) \\ &= \frac{1}{\sigma} [r'(\varphi) - r(\varphi)] - (n' - n)f_x \\ &= \varphi^{\sigma-1} f \left[\frac{1+n'\tau^{1-\sigma}}{(\varphi^{*'})^{\sigma-1}} - \frac{1+n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} \right] - (n' - n)f_x. \end{aligned}$$

This profit change increases without bound with φ and will be positive for all φ above a cutoff level $\varphi^{\dagger} > \varphi_x^{*}$.⁴⁰

³⁹Note that $\varphi^{\sigma-1}j(\varphi)$ must be a decreasing function of φ since its elasticity with respect to φ is $(\sigma-1) + \varphi j'(\varphi)/j(\varphi) < 0$.

⁴⁰The term in the bracket in (E.3) must be positive as $1+n\tau^{1-\sigma}/(\varphi^*)^{\sigma-1}$ increases with n .

Decrease in τ : As was the case with the increase in n , the least productive firms who do not export (with $\varphi < \varphi_x^*$) incur both a revenue and profit loss. There now exists a new category of firms with intermediate productivity levels ($\varphi_x^{*'} \leq \varphi < \varphi_x^*$) who enter the export markets as a consequence of the decrease in τ . The new export sales generate an increase in revenue for all these firms, but only a portion of these firms (with productivity $\varphi > \varphi^\dagger$ where $\varphi_x^{*'} < \varphi^\dagger < \varphi_x^*$) also increase their profits. Firms with productivity levels $\varphi \geq \varphi_x^*$ who export both before and after the change in τ enjoy a profit increase that is proportional to their combined revenue increase (their fixed costs do not change) and is increasing in their productivity level φ :

$$\begin{aligned}\Delta\pi(\varphi) &= \frac{1}{\sigma}[r'(\varphi) - r(\varphi)] \\ &= \varphi^{\sigma-1} f \left[\frac{1 + n(\tau')^{1-\sigma}}{(\varphi^{*'})^{\sigma-1}} - \frac{1 + n\tau^{1-\sigma}}{(\varphi^*)^{\sigma-1}} \right],\end{aligned}$$

where the term in the bracket must be positive.

E.4. Changes in Aggregate Productivity

Any productivity average based on (D.2) must increase when n increases or τ decreases as the new distribution of firm revenues $r'(\varphi)g(\varphi)/R$ first order stochastically dominates the old one $r(\varphi)g(\varphi)/R$: $\int_0^\varphi r'(\xi)g(\xi)d\xi \leq \int_0^\varphi r(\xi)g(\xi)d\xi \forall \varphi$.⁴¹ Note that this property does not hold when f_x decreases as the revenues of the most productive firms are not higher with the lower f_x . Nevertheless, the productivity average Φ will rise when f_x decreases so long as the new exporters are more productive than the average ($\varphi_x^* > \Phi$).

REFERENCES

- AW, B. Y., S. CHUNG, AND M. J. ROBERTS (2000): "Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China)," *World Bank Economic Review*, 14, 65–90.
- BERNARD, A. B., J. EATON, J. B. JENSON, AND S. KORTUM (2000): "Plants and Productivity in International Trade," NBER Working Paper No. 7688, forthcoming in the *American Economic Review*.
- BERNARD, A. B., AND J. B. JENSEN (1999a): "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*, 47, 1–25.
- (1999b): "Exporting and Productivity," NBER Working Paper No. 7135.
- (2001): "Why Some Firms Export," NBER Working Paper No. 8349.
- (2002): "The Deaths of Manufacturing Plants," NBER Working Paper No. 9026.
- BERNARD, A. B., AND J. WAGNER (2001): "Export Entry and Exit by German Firms," *Weltwirtschaftliches Archiv—Review of World Economics*, 137, 105–123.
- CLERIDES, S. K., S. LACK, AND J. R. TYBOUT (1998): "Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco," *The Quarterly Journal of Economics*, 113, 903–947.
- DAS, S., M. J. ROBERTS, AND J. R. TYBOUT (2001): "Market Entry Costs, Producer Heterogeneity and Export Dynamics," Mimeo, Pennsylvania State University.
- DIXIT, A., AND J. STIGLITZ (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297–308.
- HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2002): "Export Versus FDI," NBER Working Paper No. 9439.

⁴¹ Again, this is a direct consequence of the market share reallocation results.

- HOPENHAYN, H. (1992a): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127–1150.
- (1992b): "Exit, Selection, and the Value of Firms," *Journal of Economic Dynamics and Control*, 16, 621–653.
- KRUGMAN, P. R. (1979): "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9, 469–479.
- (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70, 950–959.
- MONTAGNA, C. (1995): "Monopolistic Competition with Firm-Specific Costs," *Oxford Economic Papers*, 47, 318–328.
- PAVCNIK, N. (2002): "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants," *The Review of Economic Studies*, 69, 245–276.
- ROBERTS, M. J., T. SULLIVAN, AND J. R. TYBOUT (1995): "Micro-Foundations of Export Booms," Mimeo, World Bank.
- ROBERTS, M. J., AND J. R. TYBOUT (1997a): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87, 545–564.
- (1997b): *What Makes Exports Boom?* Washington, D.C.: The World Bank.