Market Size, Trade, and Productivity*

Marc J. Melitz
Harvard University
NBER
and
CEPR

Gianmarco I.P. Ottaviano
University of Bologna
FEEM
and
CEPR

June 6, 2003

PRELIMINARY AND INCOMPLETE DRAFT

COMMENTS AND SUGGESTIONS ARE WELCOME

Abstract

We develop a monopolistically competitive model of trade with firm heterogeneity - in terms of productivity differences - and endogenous differences in the 'toughness' of competition across markets - in terms of the number and average productivity of competing firms. We show that countries with higher 'market potential' attract a larger number of sellers, and hence more product variety. These sellers, on average, are more productive, are larger, and set lower prices than firms selling in smaller markets. Thus, consumers in bigger markets benefit both from lower prices and more product variety. We also describe the effect of market size on the variance of these firm performance measures. We then study the effects of reciprocal, unilateral, and preferential trade liberalization.

Keywords: market structure, market size, productivity heterogeneity, trade liberalization

J.E.L. Classification: F12, R13.

*We are grateful to Richard Baldwin, Alejandro Cunat, Gilles Duranton, Elhanan Helpman, Jacques Thisse, and Tony Venables for helpful discussions.

1 Introduction

We develop a monopolistically competitive model of trade with heterogeneous firms and endogenous differences in the 'toughness' of competition across countries. Firm heterogeneity – in the form of productivity differences – is introduced in a similar way to Melitz (2003): firms face some initial uncertainty concerning their future productivity when making a costly and irreversible investment decision prior to entry. However, we further incorporate endogenous markups using the linear demand system with horizontal product differentiation developed by Ottaviano, Tabuchi, and Thisse (2002). This generates an endogenous distribution of markups across firms that responds to the 'toughness' of competition in a market – the number and average productivity of competing firms in that market. We analyze how these features vary across markets of different size that are not perfectly integrated through trade.

We first introduce a closed economy version of our model. We show that market size induces important changes in the equilibrium distribution of firms and their performance measures. Bigger markets exhibit higher levels of product variety and host more productive firms that set lower markups (hence prices are lower). These firms are bigger (in terms of both output and sales) and earn higher profits (although average markups are lower). Although profits are higher, the average profitability of the industry – measured as the profit to sales ratio – does not vary with market size. Firm survival is also lower in the bigger market: an entrant has a higher probability of failure. Finally, the variance of costs, prices, and markups are lower in bigger markets, while the variance of output and sales are higher. These theoretical results derived for a closed economy match the empirical findings reported by Syverson (2002) and Campbell and Hopenhayn (2002) using firm/plant level data in relatively non-traded industries.

We then present the open economy version of the model. We show that the foregoing results continue to hold in a two-country setup. Specifically, unless trade is perfectly free, the bigger market still exhibits 'tougher' competition than its smaller trading partner. Accordingly, as in the closed economy case, the bigger market still exhibits larger and more productive firms as well as more product variety, lower prices, and lower markups. Again, the larger market exhibits lower variance of costs, prices, and markups but higher variance of output and sales.

When bilateral trade liberalization is considered, our model predicts intra-industry selection and re-allocation effects similar to those emphasized in Melitz (2003) and strongly supported by several recent micro-econometric studies (see, among others, Aw, Chung, and Roberts, 2000; Bernard and

Jensen, 1999; Clerides, Lach, and Tybout, 1998; Pavcnik, 2002; and Tybout, 2002 for a recent survey): trade liberalization increases average productivity by forcing the least productive firms to exit and re-allocating market shares towards more productive firms who export; lower productivity firms only serve their domestic market. Our model also explains other empirical patterns linking the extent of trade barriers to the distribution of productivity, prices, and markups across firms. In an important departure from Melitz (2003), our model exhibits a link between bilateral trade liberalization and reductions in markups, thus highlighting the potential pro-competitive effects often associated with episodes of trade liberalization. Our model also highlights an important feedback mechanism between market size, pro-competitive effects, and firm selection: although bilateral trade liberalization increases average productivity and reduces average markups in both countries, it widens the gap between different-sized countries as competition gets 'tougher' in both countries but relatively more so in the larger country. This market experiences a larger response in all firm performance measures as it attracts a disproportionately larger number of firms. Welfare rises in both countries, but proportionately more in the bigger market.

In stark contrast to the case of bilateral liberalization, unilateral and preferential trade liberalizations are not welfare improving for all countries. When liberalization is unilateral, product variety falls in the liberalizing country and conversely rises in the other country: fewer firms enter the liberalized market and more firms choose to locate in the other country. Average productivity then decreases in the liberalizing country while it increases for its trading partner. Markups and prices are then higher in the liberalizing country and lower for its trading partner. As a result, welfare decreases in the liberalizing country and increases in the relatively more protected one. This result highlights an important difference between market size and 'market potential'. The latter characterizes a country's access to all national markets. This difference is most apparent in the case of countries with identical size. In this case, firms face higher trade costs when shipping their products from the liberalized country. Conversely, export costs are relatively lower for firms in the more protected country. Therefore, while both countries provide domestic firms with the same access to local customers, the more protected country provides domestic firms with better access to foreign customers. By transforming this relatively protected country into an attractive 'export base', unilateral trade liberalization favors that country and harms the liberalizing one.

In order to investigate the effects of preferential trade agreements, we extend our model to a three country version. We then show that the effects of preferential liberalization are opposite to the case of unilateral liberalization: welfare improves in the liberalizing countries and deteriorates in the protectionist country. Product variety increases in the liberalizing countries but decreases in the other – as firms are attracted to the former countries. In these liberalizing countries, average productivity increases, leading to tougher competition with lower markups, and hence lower prices. These effects are all reversed in the other country. The liberalizing countries thus become better 'export bases': they get improved access to each other's market while maintaining the same ease of access to the excluded country's market.

In all these cases, our model remains highly tractable and can easily be extended to a very general framework with multiple asymmetric countries integrated to different extents through asymmetric trade costs. We therefore believe that this model provides a useful tool that is particularly well suited for the analysis of various liberalization and regional integration scenarios in the presence of firm heterogeneity.

The paper is organized in four additional sections after the introduction. The first presents and solves the closed economy model. The second derives the two-country model and studies the effects of international market size differences. The third investigates the impacts of trade liberalization considering both bilateral and unilateral experiments. This includes a three-country version of the model that highlights the effects of preferential trade agreements.

2 Closed Economy

Consider an economy with L consumers, each supplying one unit of labor.

2.1 Preferences and Demand

Preferences are defined over a continuum of differentiated varieties indexed by $i \in \Omega$, and a homogenous good chosen as numeraire. All consumers share the same utility function given by

$$U = q_0^c + \alpha \int_{i \in \Omega} q_i^c di - \frac{1}{2}\gamma \int_{i \in \Omega} (q_i^c)^2 \, di - \frac{1}{2}\eta \left(\int_{i \in \Omega} q_i^c di \right)^2, \qquad (1)$$

where q_0^c and q_i^c represent the individual consumption levels of the numeraire good and each variety i. The demand parameters α, η, and γ are all positive. The parameters α and η index the substitution pattern between the differentiated varieties and the numeraire: increases in α and decreases in η both shift out the demand for the differentiated varieties relative to the numeraire. The parameter γ indexes the degree of product differentiation between the varieties. In the limit when $\gamma = 0$, consumers only care about their total consumption level over all varieties, $Q^c = \int_{i \in \Omega} q_i^c di$. The varieties are then perfect substitutes. The degree of product differentiation increases with γ as consumers give increasing weight to the distribution of consumption levels across varieties.

The marginal utilities for all goods are bounded, and a consumer may thus not have positive demand for any particular good. We assume that consumers have positive demands for the numeraire good ($q_0^c > 0$). The inverse demand for each variety i is then given by

$$p_i = \alpha - \gamma q_i^c - \eta Q^c, \tag{2}$$

whenever $q_i^c > 0$. This will be the case so long as

$$p_i \leq \frac{1}{\eta N + \gamma} \left(\gamma \alpha + \eta N \bar{p} \right), \tag{3}$$

where the measure of varieties N and their average price $\bar{p}$ are defined over the set of varieties Ω^* with prices satisfying (3): $\bar{p} = (1/N) \int_{i \in \Omega^*} p_i di$. Note that any price above α must violate this condition since the marginal utility in (2) is bounded above by α; hence $\bar{p} \leq \alpha$ (the inequality must be strict when there is any price heterogeneity). These conditions lead to a linear market demand system for all consumed varieties:

$$q_i = L q_i^c = \frac{\alpha L}{\eta N + \gamma} - \frac{L}{\gamma} p_i + \frac{\eta N}{\eta N + \gamma} \frac{L}{\gamma} \bar{p}, \ \forall i \in \Omega^*. \tag{4}$$

where q_i is the market demand for variety i. In contrast to the case of C.E.S. demand, the price elasticity of demand is not uniquely determined by the level of product differentiation γ. Increases in the 'toughness' of competition, induced either by a lower average price $\bar{p}$ or more product variety N, lead to increases in the price elasticity of demand.

Welfare can be evaluated using the indirect utility function associated with (1):

$$U = I^c + \frac{1}{2} \left(\eta + \frac{\gamma}{N} \right)^{-1} (\alpha - \bar{p})^2 + \frac{1}{2} \frac{N}{\gamma} \sigma_p^2, \tag{5}$$

where I^c is the consumer's income and $\sigma_p^2 = (1/N) \int_{i \in \Omega^*} (p_i - \bar{p})^2 \, di$ represents the variance of prices. To ensure positive demand levels for the numeraire, we assume that $I^c > \int_{i \in \Omega^*} p_i q_i^c di = \bar{p} Q^c - N \sigma_p^2 / \gamma$. Welfare naturally rises with decreases in average prices $\bar{p}$. It also rises with increases in the variance of prices σ_p^2 (holding the mean price $\bar{p}$ constant), as consumers then re-optimize their purchases by shifting expenditures towards lower priced varieties. Finally, the demand system exhibits 'love of variety': holding the distribution of prices constant (namely holding the mean $\bar{p}$ and variance σ_p^2 of prices constant), welfare rises with increases in product variety N.

2.2 Production and Firm Behavior

Labor is the only factor of production and its market is perfectly competitive. The numeraire good is produced under constant returns to scale at unit cost and its market is also perfectly competitive. These assumptions imply a unit wage. Entry in the differentiated product sector is costly as each firm incurs product development and production startup costs. Subsequent production exhibits constant returns to scale at marginal cost c (equal to unit labor requirement). Research and development yield uncertain outcomes for c, and firms learn about this cost level only after making the irreversible investment f_E required for entry. We model this as a draw from a common (and known) distribution $G(c)$ with support on $[0, c_M]$. Since the entry cost is sunk, firms that are able to cover their marginal cost survive and produce. All other firms exit the industry. Surviving firms maximize their profits using the residual demand function (4). In so doing, given the continuum of competitors, a firm takes the average price level $\bar{p}$ and number of firms N as given. This is the monopolistic competition outcome.

The profit maximizing price $p(c)$ and output level $q(c)$ of a firm with cost c must then satisfy

$$q(c) = \frac{L}{\gamma}\left[p(c) - c\right]. \tag{6}$$

The profit maximizing price $p(c)$ may be above the threshold in (3), in which case the firm decides to exit. Let c_D reference the cost of the firm who is just indifferent about remaining in the industry. This firm earns zero profit as its price is driven down to its marginal cost, $p(c_D) = c_D$, and its demand level $q(c_D)$ is driven to zero. We assume that c_M is high enough to be above c_D, so that some firms with cost draws between these two levels choose to exit. All firms with cost $c < c_D$ earn positive profits (gross of the entry cost) and remain in the industry. The threshold cost c_D summarizes the effects of both the average price and number of firms on the performance measures of all firms. Let $r(c) = p(c)q(c)$, $\pi(c) = r(c) - c$, $\mu(c) = p(c) - c$ denote the revenue, profit, and (absolute) markup of a firm with cost c. All these performance measures can then be written as functions of c and c_D only:

$$\begin{aligned} p(c) &= \frac{1}{2}\left(c_D + c\right), & (7)\\ \mu(c) &= \frac{1}{2}\left(c_D - c\right), & (8)\\ q(c) &= \frac{L}{2\gamma}\left(c_D - c\right), & (9)\\ r(c) &= \frac{L}{4\gamma}\left[\left(c_D\right)^2 - c^2\right], & (10)\\ \pi(c) &= \frac{L}{4\gamma}\left(c_D - c\right)^2. & (11) \end{aligned}$$

As expected, lower cost firms set lower prices and earn higher revenues and profits than firms with higher costs. However, lower cost firms do not pass on all of the cost differential to consumers in the form of lower prices: they also set higher markups (in both absolute and relative terms) than firms with higher costs.

2.3 Free Entry Equilibrium

Prior to entry, the expected firm profit is $\int_0^{c_D} \pi(c)dG(c) - f_E$. If this profit were negative, no firms would enter the industry. As long as some firms produce, the expected profit is driven to zero by the unrestricted entry of new firms. Using (11), this yields the equilibrium free entry condition

$$\int_0^{c_D} \pi(c)dG(c) = \frac{L}{4\gamma}\int_0^{c_D}\left(c_D - c\right)^2 dG(c) = f_E, \tag{12}$$

which determines the cost cutoff c_D. This cutoff, in turn, determines the number of surviving firms, since $c_D = p(c_D)$ must also be equal to the zero demand price threshold in (3):

$$c_D = \frac{1}{\eta N + \gamma}\left(\gamma\alpha + \eta N\bar{p}\right),$$

which yields

$$N = \frac{2\gamma}{\eta}\frac{\alpha - c_D}{c_D - \bar{c}}, \tag{13}$$

where $\bar{c} = \left[\int_0^{c_D} c dG(c)\right]/G(c_D)$ is the average cost of surviving firms.[1] The number of entrants is then given by $N_E = N/G(c_D)$.

Given a production technology referenced by $G(c)$, average productivity will be higher (lower

[1] Given (7), it is readily verified that $\bar{p} = \left(c_D + \bar{c}\right)/2$.

$\bar{c}$) when sunk costs are lower, when varieties are closer substitutes (lower γ), and in bigger markets (more consumers L). In all these cases, firm exit rates are also higher (the pre-entry probability of survival $G(c_D)$ is lower). The demand parameters α and η that index the overall level of demand for the differentiated varieties (relative to the numeraire) do not affect the selection of firms and industry productivity – they only affect the number of firms. Competition is 'tougher' in larger markets as more firms compete and average prices $\bar{p} = (c_D + \bar{c})/2$ are lower. A firm with cost c responds to this tougher competition by setting a lower markup (relative to the markup it would set in a smaller market – see (8)).

2.4 Parametrization of Technology

All the results derived so far hold for any distribution of cost draws $G(c)$. However, in order to simplify some of the ensuing analysis, we use a specific parametrization for such distribution. In particular, we assume that productivity draws $1/c$ follow a Pareto distribution with lower productivity bound $1/c_M$ and shape parameter $k \geq 1$. This implies a distribution of cost draws c given by

$$G(c) = \left(\frac{c}{c_M}\right)^k, \ c \in [0, c_M]. \tag{14}$$

The shape parameter k indexes the dispersion of cost draws. When $k = 1$, the cost distribution is uniform on $[0, c_M]$. As k increases, the relative number of high cost firms increases, and the cost distribution is more concentrated at these higher cost levels. As k goes to infinity, the distribution becomes degenerate at c_M. Any truncation of the cost distribution from above will retain the same distribution function and shape parameter k. The productivity distribution of surviving firms will therefore also be Pareto with shape k, and the truncated cost distribution will be given by $G_D(c) = (c/c_D)^k, \ c \in [0, c_D]$.

Given this parametrization, the cutoff cost level c_D determined by (12) is then

$$c_D = \left[\frac{2(k+1)(k+2)\gamma\, (c_M)^k f_E}{L}\right]^{\frac{1}{k+2}}, \tag{15}$$

where we assume that $c_M > \sqrt{[2(k+1)(k+2)\gamma f_E]/L}$ in order to ensure that $c_D < c_M$ as was previously anticipated. From (13), the corresponding number of firms is

$$N = \frac{2(k+1)\gamma}{\eta} \frac{\alpha - c_D}{c_D}. \tag{16}$$

As expected, these results show that the properties derived in the case of a generic distribution $G(c)$ still hold: average productivity is higher when sunk costs are lower, when varieties are closer substitutes, and in bigger markets. These are explained by the tougher competition in bigger markets, which leads to lower average prices and markups. In addition, under the Pareto assumption, higher c_M leads to higher c_D, whereas higher k generates higher c_D only when k is small and c_M is large.

The average performance measures from (7)-(11) can then be written:

$$\begin{aligned} \bar{c} &= \frac{k}{k+1} c_D, & \bar{q} &= \frac{L}{2\gamma} \frac{1}{k+1} c_D = \frac{(k+2)(c_M)^k}{(c_D)^{k+1}} f_E, \\ \bar{p} &= \frac{2k+1}{2k+2} c_D, & \bar{r} &= \frac{L}{2\gamma} \frac{1}{k+2} (c_D)^2 = \frac{(k+1)(c_M)^k}{(c_D)^k} f_E, \\ \bar{\mu} &= \frac{1}{k+1} c_D, & \bar{\pi} &= f_E \frac{(c_M)^k}{(c_D)^k}. \end{aligned}$$

As with the cost average $\bar{c}$, the average for a performance measure $z(c)$ is given by $\bar{z} = \left[\int_0^{c_D} z(c) dG(c)\right] / G(c_D)$. These values can be substituted into (5) to yield:

$$U = 1 + \frac{1}{2\eta} (\alpha - c_D) \left(\alpha - \frac{k+1}{k+2} c_D \right). \tag{17}$$

Welfare naturally increases with decreases in the cutoff c_D, as the latter induces both increases in product variety N and decreases in the average price $\bar{p}$.

In this model, market size induces some important changes to the distribution of firms and their performance measures. In addition to being more productive and setting lower prices in bigger markets, firms are also bigger (in terms of both output and sales) and earn higher profits (although average markups are lower). Although profits are higher, the average profitability of the industry, measured as the profit to sales ratio $\bar{\pi}/\bar{r}$, does not vary with market size. Finally, it can also be easily verified that the variances of costs, prices, and markups are lower in bigger markets, while the variance of output and sales are higher.[2]

[2] All derivations are based on the assumption that consumers have positive demands for the numeraire good. Consumers derive all of their income from their labor: there are no redistributed firm profits as ex-ante industry profits (net of the entry costs) are zero. We therefore need to ensure that each consumer spends less than this unit income on the differentiated varieties. Spending per consumer on the varieties is $N\bar{r}/L = (\alpha - c_D) c_D (k+1) / [\eta (k+2)]$. A sufficient condition for this to be less than 1 is $\alpha < 2\sqrt{\eta (k+2) / (k+1)}$.

3 Open Economy

In the previous section we used a closed economy model to assess the effects of market size on various performance measures at the industry level. This closed economy model could be immediately applied to a set of open economies that are perfectly integrated through trade. This scenario, however, can not be extended to the case of goods that are not freely traded. Furthermore, although trade is costly, it nevertheless connects markets in ways that preclude the analysis of each market in isolation. To understand these inter-market linkages, we now extend our model to a two-country setting.

Consider two countries, H and F, with L^H and L^F consumers in each country. Consumers in both countries share the same preferences, leading to the inverse demand function (2). The two markets are segmented, although firms can produce in one market and sell in the other, incurring a per-unit trade cost. Specifically, the delivered cost of a unit with cost c to country l ($l = H, F$) is $\tau^l c$ where $\tau^l > 1$. Thus, we allow countries to differ along two dimensions: market size L^l and barriers to imports τ^l.

Let p^l denote the price threshold for positive demand in market l. Then (3) implies

$$p^l = \frac{1}{\eta N^l + \gamma}\left(\gamma\alpha + \eta N^l \bar{p}^l\right), \quad l = H, F, \tag{18}$$

where N^l is the total number of firms selling in country l (the total number of domestic firms and foreign exporters) and $\bar{p}^l$ is the average price (across both local and exporting firms) in country l. Let $p^l_D(c)$ and $q^l_D(c)$ represent the domestic levels of the profit maximizing price and quantity sold for a firm producing in country l with cost c. Such a firm may also decide to produce some output $q^l_X(c)$ that it exports at a delivered price $p^l_X(c)$.

Since the markets are segmented and firms produce under constant returns to scale, they independently maximize the profits earned from domestic and exports sales. Let $\pi^l_D(c) = \left[p^l_D(c) - c\right] q^l_D(c)$ and $\pi^l_X(c) = \left[p^l_X(c) - \tau^h c\right] q^l_X(c)$ denote the maximized value of these profits as a function of the firm's marginal cost c (where $h \neq l$). Analogously to (6), the profit maximizing prices and output levels must satisfy: $q^l_D(c) = \left(L^l/\gamma\right)\left[p^l_D(c) - c\right]$ and $q^l_X(c) = \left(L^h/\gamma\right)\left[p^l_X(c) - \tau^h c\right]$. Let c^l_D denote the cutoff cost level of a firm selling in its domestic market. This firm will earn zero profits from domestic sales: $c^l_D = \inf\{c : \pi^l_D(c) = 0\}$. Similarly, let c^l_X denote the cutoff cost level of an exporting firms. This firm earns zero profit from exporting: $c^l_X = \inf\{c : \pi^l_X(c) = 0\}$. These cutoffs must then satisfy:

$$\begin{aligned} c_D^l &= p^l, \\ c_X^l &= \frac{p^h}{\tau^h}, \end{aligned} \tag{19}$$

which implies that, in each market, the cutoffs are lower for foreign exporters than for domestic firms: $c_X^l = c_D^h/\tau^h$. Therefore, due to trade barriers, survival is tougher for exporters.

As was the case in the closed economy, the cutoffs summarize all the effects of market conditions relevant for firm performance. In particular, the optimal prices and output levels can be written as functions of the cutoffs:

$$\begin{aligned} p_D^l(c) &= \frac{1}{2}\left(c_D^l + c\right), \qquad & q_D^l(c) &= \frac{L^l}{2\gamma}\left(c_D^l - c\right), \\ p_X^l(c) &= \frac{\tau^h}{2}(c_X^l + c), \qquad & q_X^l(c) &= \frac{L^h}{2\gamma}\tau^h\left(c_X^l - c\right), \end{aligned} \tag{20}$$

which yield the following maximized profit levels:

$$\begin{aligned} \pi_D^l(c) &= \frac{L^l}{4\gamma}\left(c_D^l - c\right)^2, \\ \pi_X^l(c) &= \frac{L^h}{4\gamma}\left(\tau^h\right)^2\left(c_X^l - c\right)^2. \end{aligned} \tag{21}$$

3.1 Free Entry Condition

Entry is unrestricted in both countries. Firms choose a production location prior to entry and paying the same sunk cost f_E. Then, free entry of firms in country l implies zero expected profits in equilibrium:

$$\int_0^{c_D^l} \pi_D^l(c)dG(c) + \int_0^{c_X^l} \pi_X^l(c)dG(c) = f_E.$$

We also assume the same Pareto parametrization of firm cost draws (14) in both countries. Given (21) this allows us to re-write the free entry condition as

$$L^l\left(c_D^l\right)^{k+2} + L^h\left(\tau^h\right)^2\left(c_X^l\right)^{k+2} = \gamma\phi, \tag{22}$$

where $\phi = 2(k+1)(k+2)(c_M)^k f_E$ is a technology index that combines the effects of better distribution of cost draws (lower c_M) and lower entry costs f_E.[3] Given $N_E^l > 0$, there will be $G^l(c_D^l)N_E^l$ firms selling in the domestic market and $G^l(c_X^l)N_E^l$ exporters. It can further be shown that $c_X^l < c_D^l$, which implies that only a subset of more productive firms choose to export. The remaining higher cost firms (with costs between c_X^l and c_D^l) only serve their domestic market. $G^l(c_D^l)N_E^l$ is then also the number of surviving firms in country l. The number of firms *selling* in country l, N^l, must satisfy:

$$G^l\left(c_D^l\right)N_E^l + G^h\left(c_X^h\right)N_E^h = N^l. \tag{23}$$

3.2 Prices, Product Variety, and Welfare

The prices in country l reflect both the domestic prices of country-l firms, $p_D^l(c)$, and the prices of exporters from h, $p_X^h(c)$.Using (19) and (20), these prices can be written:

$$\begin{aligned} p_D^l(c) &= \frac{1}{2}\left(p^l + c\right), \ c \in [0, c_D^l], \\ p_X^h(c) &= \frac{1}{2}\left(p^l + \tau^l c\right), \ c \in [0, c_D^l/\tau^l]. \end{aligned}$$

In addition, the cost of domestic firms $c \in [0, c_D^l]$ and the delivered cost of exporters $\tau^l c \in [0, c_D^l]$ have identical distributions over this support, given by $G_D^l(c) = c^k / \left(c_D^l\right)^k$. The price distribution in country l of domestic firms producing in l, $p_D^l(c)$, and exporters producing in h, $p_X^h(c)$, are therefore also identical. The average price in country l is thus given by

$$\bar{p}^l = \frac{2k+1}{2k+2} c_D^l.$$

Combining this with the threshold price in (18) determines the number of firms selling in country l:

$$N^l = \frac{(2k+2)\gamma}{\eta} \frac{\alpha - c_D^l}{c_D^l}. \tag{24}$$

These results for product variety and average prices are identical to the closed economy case. This is due to the matching of price distributions across domestic firms and exporters. Similarly, welfare

[3] The condition (22) will hold so long as there is a positive mass of entrant $N_E^l > 0$. Otherwise, $\int_0^{c_D^l} \pi_D^l(c)dG(c) + \int_0^{c_X^l} \pi_X^l(c)dG(c) < f_E^l$ and $N_E^l = 0$. For the sake of parsimony, we rule out this case by assuming that α is large enough.

in country l can be written in an identical way to (17) as:

$$U^l = 1 + \frac{1}{2\eta}\left(\alpha - c_D^l\right)\left(\alpha - \frac{k+1}{k+2}c_D^l\right). \tag{25}$$

Once again, welfare changes monotonically with the domestic cost cutoff, which captures both the effects of product variety and average prices.[4]

3.3 Different Market Sizes

In order to emphasize the role of asymmetric market sizes, we assume that trade costs are symmetric ($\tau^H = \tau^F = \tau$). Using (19), the free entry conditions for both countries can be written as a system of equations in the two domestic cutoffs levels c_D^H and c_D^F:

$$L^l \left(c_D^l\right)^{k+2} + \tau^{-k} L^h \left(c_D^h\right)^{k+2} = \gamma\phi, \quad l,h = H,F; l \neq h, \tag{26}$$

which can be solved for the cutoffs:

$$c_D^l = \left[\frac{\gamma\phi}{L^l\left(1+\rho\right)}\right]^{\frac{1}{k+2}}, \quad l = H,F, \tag{27}$$

where $\rho = \tau^{-k} \in (0,1)$ is an inverse measure of trade costs. The cutoffs, in turn, determine the number of firms selling in both countries (N^H and N^F) using (24). These can then be used to solve for the number of entrants in both countries using (23):

$$N_E^l = \frac{\left(c_M\right)^k}{1-\rho^2}\left[\frac{N^l}{\left(c_D^l\right)^k} - \rho\frac{N^h}{\left(c_D^h\right)^k}\right], \quad l,h = H,F; l \neq h. \tag{28}$$

Together (24), (27), and (28) determine the open economy equilibrium and highlight the role of size asymmetries – as the results of the closed economy are reproduced: $L^l > L^h$ implies $c_D^l < c_D^h$, $N^l > N^h$, and $N_E^l > N_E^h$. Thus, firms selling in the larger country are more productive and larger. The larger market also exhibits higher product variety and lower markups. Again, welfare in the large market is higher due to the combined effect of lower prices and more product variety.[5] Finally,

[4] The previously derived condition for the demand parameters α and η, and k again ensure that $q_0^l > 0$ as has been implicitly assumed.

[5] It interesting to point out that (27) implies that the size of a country's trading partner does not affect the selection of firms at home and thus its welfare. Trading with a bigger country has both benefits and drawbacks. First, from the export market perspective, the benefits of a bigger export market are cancelled out by the effect of tougher competition in the export market. Second, from the domestic market perspective, the effect of increased competition from imports is cancelled out by the effect of a lower number of entrants in the domestic market. In this

it can also be easily verified that the variance of costs, prices, and markups are lower in the bigger country, while the variances of output and sales are higher.

4 Trade liberalization

We have just shown that, in an open economy – given symmetric trade costs – the effects of country size on the firm performance measures are similar to those derived for the closed economy. There are, nonetheless, other crucial differences. To highlight these, we consider three types of trade liberalization: bilateral and unilateral liberalizations in a two-country world, and preferential trade liberalization in a three-country world.

4.1 Bilateral liberalization

Bilateral trade liberalization (higher ρ) emphasizes the crucial difference between the closed and the open economy cases: the gaps in the performance measures between different-sized countries depend on the level of trade barriers. When trade barriers are prohibitive ($\rho = 0$), (27) indicates that we recover the closed economy results. However, as trade barriers fall (ρ rises), the difference between c_D^l and c_D^h rises: cutoffs fall in both countries, but fall faster in the larger country. This implies that the numbers of sellers and entrants rise (hence product variety increases) in both countries, but rises faster in the larger country. All other performance variables behave accordingly: average costs, prices, and markups fall in both countries but fall faster in the bigger country. Average outputs and sales rise in both countries but rise faster in the bigger one. (The variances also exhibit similar properties.) Thus, welfare rises in both countries, but rises disproportionately more in the larger country.

4.2 Unilateral liberalization

To assess the impact of unilateral trade liberalization, we focus on the case of symmetric country size ($L^H = L^F = L$) where one country unilaterally reduces its trade barriers from an initial symmetric equilibrium. Thus, trade barriers are no longer symmetric: $\tau^l \neq \tau^h$.

Solving the model with asymmetric trade barriers delivers the following free entry conditions :

$$L\left(c_D^l\right)^{k+2} + \rho^h L\left(c_D^h\right)^{k+2} = \gamma\phi, \quad l, h = H, F; l \neq h,$$

respect, the model is special as these effects exactly cancel out. However, it highlights an important point about the ambiguous effects of trading partner's size.

where $\rho^l = \left(\tau^l\right)^{-k}$ is an inverse measure of trade costs for exports to country l. These yield the new cutoffs

$$c_D^l = \left[\frac{\gamma\phi}{L}\frac{1-\rho^h}{1-\rho^l\rho^h}\right]^{\frac{1}{k+2}}, \quad l = H, F; l \neq h, \tag{29}$$

Once more, together with (24), the cutoffs yield the number of firms selling in both countries (N^H and N^F). These, together with (23), can then be used to solve for the number of entrants in both countries:

$$N_E^l = \frac{(c_M)^k}{1-\rho^2}\left[\frac{N^l}{\left(c_D^l\right)^k} - \rho\frac{N^h}{\left(c_D^h\right)^k}\right], \quad l, h = H, F; l \neq h.$$

From (29), we see that a liberalizing country, say l, experiences a deterioration in firm productivity (higher cutoff c_D^l when ρ^l rises) – but that its trading partner concurrently experiences productivity improvement (lower c_D^h). This implies that the numbers of sellers and entrants drop in the liberalizing country while they rise in the other. Average markups and prices thus rise in the liberalizing country, and fall in the other. Average outputs and sales fall in the former country and rise in the latter.

These responses highlight the crucial role of 'market potential' – the firms' access to all national markets – for the open economy equilibrium. In this scenario, countries share the same size but firms face higher relative export costs when producing in the liberalized country. Therefore, while both countries provide domestic firms with the same access to local consumers, the relatively more protected country provides domestic firms with better access to foreign consumers. By transforming this country into an attractive 'export base', unilateral trade liberalization harms the liberalizing country.[6] This result is similar to that in Melitz (2003), where the source of productivity gains from liberalization are driven by improved access to foreign markets rather than fiercer competition by foreign exporters.

4.3 Preferential liberalization

The role of market potential is even more evident in the multi-country case. We now introduce a third country, T. (Our model can easily be extended to an arbitrary number of trading partners.) For simplicity, we return to the assumption of symmetric trade costs between any country pair. Moreover, since we focus on the role of market potential, we also assume symmetric country sizes.

To understand the effects of preferential trade liberalization, we consider the case where two countries reduce their trade barriers bilaterally from an initial symmetric situation. As a result,

[6] See, e.g., Helpman and Krugman (1989) for a discussion of the parallel effect in the case of representative firms.

trade barriers between these two countries are then lower than those between them and the third country.

With three countries, the free entry conditions become:

$$L^l\left(c_D^l\right)^{k+2}+\rho^{lh}L^h\left(c_D^h\right)^{k+2}+\rho^{lt}L^t\left(c_D^t\right)^{k+2}=\gamma\phi,\ l,h,t=H,F,T;\ l\neq h\neq t, \tag{30}$$

where $\rho^{lh}=\left(\tau^{lh}\right)^{-k}\in(0,1)$ is an inverse measure of trade costs for exports from country l to country h. After imposing $\rho^{lh}=\rho^{hl}$ and $L^l=L$, (30) can be used to solve for the three cutoffs:

$$c_D^l=\left[\frac{\gamma\phi}{L}\frac{\left(1-\rho^{ht}\right)\left(1+\rho^{ht}-\rho^{lt}-\rho^{hl}\right)}{1+2\rho^{lh}\rho^{ht}\rho^{lt}-\left(\rho^{ht}\right)^2-\left(\rho^{lt}\right)^2-\left(\rho^{hl}\right)^2}\right]^{\frac{1}{k+2}},\quad l=H,F,T;l\neq h\neq t, \tag{31}$$

In this expression, the differences between national cutoffs come entirely from the term $\left(1-\rho^{ht}\right)$ $\left(1+\rho^{ht}-\rho^{lt}-\rho^{hl}\right)$, which shows that the country featuring the lowest sum of bilateral trade barriers has the lowest cutoff. This occurs because this country is the best export base (or 'hub'), which delivers the largest numbers of sellers and entrants (hence more product variety). This is also associated with the lowest average costs, markups, and prices; and the highest average outputs and sales.

We now turn to the effects of preferential liberalization. Initially, when trade barriers are symmetric ($\rho^{ht}=\rho^{lt}=\rho^{hl}=\rho$), the cutoffs must be identical, and are given by (31):

$$c_D^l=\left[\frac{\gamma\phi}{L\left(1+2\rho\right)}\right]^{\frac{1}{k+2}},\ l=H,F,T.$$

However, as preferential trade liberalization induces divergence in the trade costs, the cutoffs respond. For concreteness, consider a preferential trade agreement between H and F such that $\rho^{HF}=\rho^{FH}=\rho'>\rho=\rho^{FT}=\rho^{HT}$. The cutoffs are then:

$$\begin{aligned}c_D^H &= c_D^F=\left[\frac{\gamma\phi}{L}\frac{\left(1-\rho'\right)\left(1-\rho\right)}{1+2\rho'\rho^2-2\rho^2-\left(\rho'\right)^2}\right]^{\frac{1}{k+2}} \\ c_D^T &= \left[\frac{\gamma\phi}{L}\frac{\left(1-\rho'\right)\left[\left(1-\rho\right)+\left(\rho'-\rho\right)\right]}{1+2\rho'\rho^2-2\rho^2-\left(\rho'\right)^2}\right]^{\frac{1}{k+2}}\end{aligned} \tag{32}$$

It is then readily verified that preferential liberalization causes the cutoffs to decrease in the liberalizing countries but concurrently causes the cutoff in the excluded country to rise. The reason, again, is that the liberalizing countries become better export bases: they gain better access to each other's market while maintaining the same ease of access to the third country's market.

Therefore, the effects of preferential liberalization are opposite to the case of unilateral liberalization: welfare improves in the liberalizing countries and deteriorates in the protectionist country. The numbers of sellers and entrants (hence product variety) rise in the liberalizing countries but fall in the other. Average costs, prices, and markups decrease in the former and rise in the latter. The average outputs and sales respond in the opposite direction.

5 Conclusion

We have presented a rich though tractable model that predicts how a wide set of industry performance measures respond to changes in the world trading environment. First, we show how a location with higher 'market potential' exhibits a larger numbers of sellers and entrants (hence, more product variety), more productive, bigger firms, and lower markups and prices. Second, reciprocal trade liberalization increases the numbers of sellers and entrants (hence product variety) in all markets, but with a larger effect in countries with higher market potential. All other performance variables respond accordingly. In particular, average costs, markups, and prices fall in all countries but proportionately more in the country with higher market potential. Third, following unilateral trade liberalization, the number of sellers and entrants fall in the liberalizing country but rise in the other countries. The other performance measures respond accordingly, and welfare is reduced in the liberalizing country. Fourth, the effects of preferential liberalization are opposite to the case of unilateral liberalization: the numbers of sellers and entrants (hence product variety) rise in the liberalizing countries but fall in the other country. Again, all other performance measures respond accordingly, and welfare rises in the liberalizing countries. We hope that this model provides a useful foundation for future empirical investigations.

References

[1] 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.

[2] Bernard, A. and B. Jensen (1999) Exceptional Exporter Performance: Cause, Effect, or Both?, Journal of International Economics 47, 1-25.

[3] Campbell, J. and H. Hopenhayn (2002) Market size matters, NBER Working paper n.9113.

[4] Clerides, S.K., S. Lach and J. R. Tybout (1998) Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco. The Quarterly Journal of Economics 113, 903-47.

[5] Helpman, E. and P. Krugman (1989), Trade Policy and Market Structure, MIT Press.

[6] Melitz, M. (2003) The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, forthcoming, Econometrica.

[7] Ottaviano, G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and Trade Revisited, International Economic Review 43, 409-436.

[8] Pavcnik, N. (2002) Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants, Review of Economic Studies 69, 245-276.

[9] Syverson, C. (2002) Price Dispersion: The Role of Product Substitutability and Productivity, mimeo, University of Chicago.

[10] Tybout, J. (2002) Plant and Firm-Level Evidence on New Trade Theories. In Handbook of International Economics, J. Harrigan (ed.), Vol. 38, Basil-Blackwell.