

Heterogeneous firms and trade: testable and untestable properties of the Melitz model

*Richard E. Baldwin**

May 2005

Abstract

This paper sets out a basic heterogeneous-firms trade model that is closely akin to Melitz (2003). The positive and normative properties of the model are studied in a manner intended to highlight the core economic logic of the model. The paper also studies the impact of greater openness at the firm-level and aggregate level, focusing on changes in the number and type of firms, trade volumes and prices, and productivity effects. The normative effects of liberalisation are also studied and here the paper focuses on aggregate gains from trade, and income redistribution effects, showing *inter alia* that the model is marked by a Stolper-Samuelson like effect.

A number of empirically testable hypotheses are also developed. These concern the impact of greater openness on the firm-level trade pattern, the variance of unit-prices, the stock market valuation of firms according to size, and the lobbying behaviour by size.

1. INTRODUCTION

Until the 1980s, ‘old’ trade theory adopted modelling approaches that assumed away intra-industry trade for simplicity’s sake, but empirical evidence revealed that the bulk of world trade was exactly of the assumed-away kind (Grubel and Lloyd 1975). In response, ‘new’ trade theory incorporated imperfect competition and increasing returns to account for intraindustry trade.¹ The modelling approaches adopted by the ‘new’ trade theory assumed away differences among firms for simplicity’s sake. Recent empirical evidence, however, shows that differences among firms are crucial to understanding world trade. For example, firm differences within sectors may be more pronounced than differences between sector averages, and most firms – even in traded-goods sectors – do not export at all (Bernard and Jensen 1995, 1999a,b, 2001; Clerides, Lach and Tybout 1998, Aw, Chung, and Roberts 2000, Eaton, Kortum, and Kramarz 2004; see Tybout 2003 for a survey). In response, what might be called the ‘new new’ trade theory incorporated firm-level heterogeneity to account for the many of the new firm-level facts. The main theoretical papers in this rapidly expanding literature are Bernard, Eaton, Jensen and Kortum (2003), Melitz (2003), Helpman, Melitz and Yeaple (2004), Bernard, Redding and Schott (2004), Bernard, Eaton, Jensen, and Schott (2003), Melitz and Ottaviano (2005), and Yeaple (2005).

These theoretical models add two critical elements to the new trade theory:

- 1) Heterogeneity with respect to firm’s marginal costs, and
- 2) Fixed market-entry costs that are added to the standard fixed cost of developing a new variety.

Thus one name for this new approach is HMCFMEC models, short for heterogeneous marginal costs and fixed market-entry costs models.

* Baldwin; Graduate Institute of International Studies, Avenue de la Paix 11A, 1202 Geneva, Switzerland, <http://heiwww.unige.ch/~baldwin>. I thank Toshi Okubo, Virginia di Nino and Theresa Carpenter for comments on earlier drafts.

¹ The pioneers in this field were (alphabetic order): Brander, Dixit, Eaton, Either, Grossman, Helpman, Krugman, Lancaster, Markusen, Norman, Spencer.

This paper presents the baseline model in a very uncluttered setting and demonstrates the main results in the literature. It also develops a sequence of hypotheses that should, in principle, be empirically testable.

2. THE BASIC HMCFMEC MODEL

The basic model is a slightly simplified version of Melitz (2003). Most of the model's elements are identical to the standard monopolistic competition trade model. Specifically, the model combines heterogeneity in firms' marginal costs with elements of the one of the simplest new trade theory models (Krugman 1980)) and its extension to include market-entry costs (Baldwin 1988, Baldwin and Krugman 1989, and Dixit 1988).

2.1. Assumptions

Consider a world with two identical nations, a single primary factor and a single consumption-good sector. Our analysis focuses exclusively on steady state equilibria and intertemporal discounting is ignored, but firms face a constant probability of 'death' according to a Poisson process with the hazard rate δ .²

The model assumes Dixit-Stiglitz monopolistic competition and iceberg trade costs (selling one unit in the export market requires shipment of $\tau \geq 1$ units). Marginal production costs are constant but increasing returns arise from three fixed costs paid just prior to production. First, all firms face the standard Dixit-Stiglitz fixed cost of developing a new variety. The second and third fixed costs are different. Selling a new variety also requires firms to pay fixed costs to enter each market. These fixed market-entry costs – or 'beachhead' costs³ – reflect the cost of introducing a new variety into a market, e.g. the cost of meeting market-specific standards and regulations, establishing a brand name, etc. Here the costs are denoted as F_D and F_X for the local and export markets, respectively (D and X are mnemonics for 'domestic' and 'export'). It is important that the innovation fixed cost be sunk, but the two beachhead costs may be overhead-type fixed costs (i.e. reoccurring each period), or one-time sunk costs given that HMCFMEC models ignore transitional dynamics. To be concrete, assume F_D and F_X are sunk just like the variety-development costs, which is denoted F_I (I is a mnemonic for innovation, we could use E for entry but that risks confusion with 'exporting').

Crucially, the model allows for heterogeneity with respect to firms' marginal production costs; in particular, Melitz (2003) works with a simplified version of the Hopenhayn (1992a, 1992b) mechanism of development of firms, each associated with a particular labour input coefficient – denoted as a_j for firm j . Where do the a_j 's come from? A potential entrant into the market has to pay the variety-development cost as in the standard Dixit-Stiglitz model, F_I . Just after paying this entry cost, the firm is randomly 'assigned' an ' a_j ' that is generated from a density function $G[a]$ whose support is $0 \leq a \leq a_0$. Intuition is served by thinking of the entry-cum-lottery as a single innovation process. That is to say, the innovation technology is stochastic in a very specific sense. At the cost of F_I units of labour, one gets the 'blueprint' for a new variety with certainty, but the associated marginal cost is random.

² Specifically, a firm works perfectly until it 'dies' after which it ceases to produce – an event describe by a Poisson process with hazard rate δ . In other words, a firm's ability to produce value is assumed to remain constant over its life and then fall to zero at the end. This so-called "one-hoss shay" (=one horse sleigh, in modern English) pattern of depreciation is not a bad approximation for many forms of capital (Griliches 1963). The name comes from the Oliver Wendell Holmes poem 'The wonderful one-hoss shay': "Have you heard of the wonderful one-horse shay/ That was built in such a logical way/ It ran a hundred years to a day/ And then, of a sudden, it--ah but stay/ I'll tell you what happened without delay,/ ... /How it went to pieces all at once/ All at once, and nothing first / Just as bubbles do when they burst. / End of the wonderful one-boss shay. / Logic is logic. That's all I say."

³ The idea is that firms have to pay a fixed cost to establish a 'beachhead' in a new market.

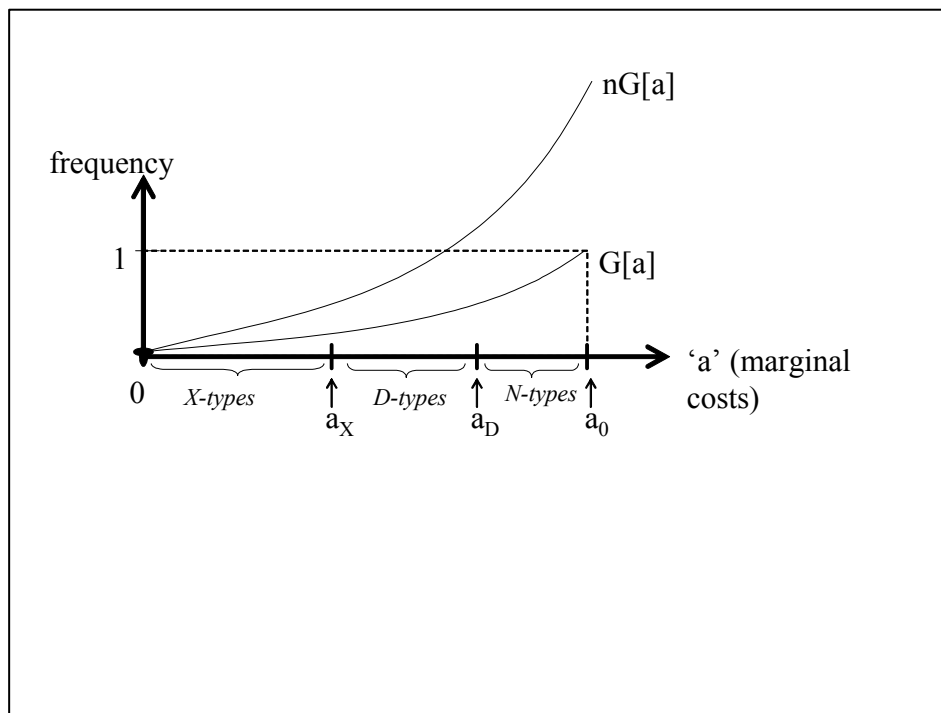
2.2. Equilibrium solutions

Although this is a general equilibrium trade model, it makes a number of assumptions that greatly simplify the usual complex interactions among markets (as is true of most trade models these days). Specifically, symmetry of nations equalises wages and eliminates concerns over trade balance. Since all wages are equal, we can take labour as numeraire, so wages are unity and all heterogeneity in firms' marginal costs stem from differences in firms' unit labour requirements, which are denoted as 'a'.⁴ Thus, we can refer without ambiguity to a firm's 'a' as its marginal costs.

2.2.1. D-types, X-types and N-types

Although each firm has its own marginal cost, it is useful to group firms into three types: firms that produce but sell only locally (D-types, short for domestic firms), firms that sell locally and also export (X-types, short for export firms), and firms that do not produce (N types, short for non-producers). As we shall see in the next section, this distinction is critical to the truly novel contributions of the HMCFMEC model. This section discusses these types and how they are determined, taking as given that a firm has already paid the F_1 needed to develop a new variety.

Figure 1: N, D and X types.



Standard Dixit-Stiglitz trade model results tell us that a firm's operating profit – call this π – is proportional to its revenue.⁵ The level of a firm's sales increases as its relative price falls and its relative price falls as its relative marginal cost falls (relative here means relative to that of its competitors). Given this, the standard logic of fixed costs tells us that only firms with sufficiently low marginal costs will enjoy operating profits that justify paying the beachhead costs. Since there are two beachhead costs, F_D and F_X , there will be two cut-off levels, i.e. thresholds, for marginal

⁴ Many other trade models, especially those in the so-called New Economic Geography literature achieve the same end by assuming a Walrasian sector with costless trade.

⁵ A typical Dixit-Stiglitz first order condition is $p(1-1/\sigma)=wa$, where w is wage and p is price; rearranging, the operating profit, $(p-wa)c$, equals pc/σ with c defined as consumption. Thus, operating profit, $(p-wa)c$, is proportional to revenue, pc ; the factor of proportionality is $1/\sigma$.

cost. Following the notational protocol, the marginal cost cut-offs are labelled, a_D and a_X . Consistency with the stylized fact that not all firms in a sector export, we need $a_X < a_D$, i.e. the maximum marginal-cost threshold for exporting firms is below the maximum marginal-cost threshold for firms that sell domestically.

It is instructive to consider the innovation outcomes in more detail. Some unlucky innovators/firms will have paid F_I to get a new variety with a marginal cost higher than the maximum marginal cost that is consistent with breaking even in the local market, i.e. with marginal costs that exceed a_D . Such firms never produce, i.e. are N-types. That is to say, ex poste they find out that they wasted their F_I . Other firms are luckier. They receive a blueprint for a new variety that is associated with a very low marginal cost, i.e. an 'a' that is low enough to make it profitable to enter both the local and export markets. Finally, there will also be some firms that draw intermediate marginal costs, so that they find it worth paying F_D to sell in the local market but not F_X to sell in the export market. To summarise, firms that draw a's above a_D will be N-types, those with a's between a_D and a_X will be D-types and those with a's below a_X will be X-types. This is shown in Figure 1.

Some statistical results. Note that each new variety is an independent and identical draw from the underlying marginal cost distribution $G[a]$ and there will be, in general, more than one firm with each level of marginal cost. If innovators continuously draw from $G[a]$, and firms continuously die according to the Poisson process, what will be the steady-state distribution of a's? The answer is $G[a]$ times n , where n is the mass of firms with each level of 'a'. To think about this, suppose that the answer is right and consider whether it would be self-sustaining. The a's are continuously distributed, but intuition is served by making it discrete with, say, 'm' equal-length categories of a's between 0 and a_0 . Given the Poisson firm-death process, the chance of a firm 'leaving' (i.e. dying) is proportional to the mass of existing firms in each category, which, by hypothesis is equal to $nG[a]/m$. Moreover, given the innovation process, the chance of a new firm arriving in a given category is dictated by $G[a]/m$. Thus on average, we know that the firm 'deaths' and 'births' will match in every single category, at least in the long run. Finally, recall that we are working with a continuum of goods, so there is an uncountable infinity of firms. Consequently, there will be an uncountable infinity of firm deaths each period and an uncountable infinity of firm births each period. Thus, the death/birth matching by category will be perfect at every moment, as per the law of large numbers applied to each category. Thus, the shape of the distribution of a's will be exactly $G[a]$ in steady state. Of course, after a shock to the system, it may take a while for the actual distribution of existing firms to reach the new equilibrium distribution.

What all this means is that the distribution of the a's for N-, D- and X-types will be $nG[a]$. Obviously, it would be hard to observe the 'failures' in real world data (i.e. the N-types) since they are the varieties that never make it off the drawing board, so to speak, but the distribution of the active firm's a's would be a truncated version of $G[a]$.

2.2.2. The Cut-off Conditions

With these preliminaries out of the way, we turn now to specifying the exact conditions that define the two cut-off conditions.

Given standard Dixit-Stiglitz results, the level of a firm's sales in its local market is related to its marginal cost and the marginal costs of its competitors according to:⁶

$$(1) \quad \frac{a^{1-\sigma}}{\Delta} \left(\frac{E}{n} \right); \quad \Delta = \int_0^{a_D} a^{1-\sigma} dG[a|a_D] + \tau^{1-\sigma} \int_0^{a_X} a^{1-\sigma} dG[a|a_D]$$

Here Δ (a mnemonic of 'denominator' of the standard CES demand function) can be thought of as a weighted average of the marginal costs of all firms active in the market, $\sigma > 1$ is the constant elasticity of substitution among varieties, n is the mass of varieties produced in a typical nation, and τ is the iceberg trade cost.⁷ As usual with Dixit-Stiglitz monopolistic competition, operating profit earned in this market will be $1/\sigma$ times the value of sales.

Note that since the beachhead costs are sunk, firms must consider future operating profits when making their market-entry decisions. This sort of dynamic entry problem with uncertainty and sunk entry costs can get complex (see Dixit 1988, Baldwin and Krugman 1989), but Melitz (2003) and subsequent papers like Helpman, Melitz and Yeaple (2004) make a number of assumptions that eliminate such problems.⁸ Thus considering the costs and benefits of entering a market, the relevant comparison is between the present value of operating profit and the beachhead costs. Given the constant firm-death rate δ and the zero discount rate, the present value of a given firm is just π/δ , where π is the operating profit the firm would earn if it actually produces. The cut-off levels of marginal cost in the local and export markets are respectively defined by:⁹

$$(2) \quad \frac{a^{1-\sigma}}{\Delta} \left(\frac{E}{n\sigma\delta} \right) = F_D; \quad \frac{\phi a^{1-\sigma}}{\Delta} \left(\frac{E}{n\sigma\delta} \right) = F_X; \quad 0 \leq \phi \equiv \tau^{1-\sigma} \leq 1$$

where a_D and a_X are the cut-off marginal costs for entering the local market and the export market, respectively, E is expenditure (same in both markets by symmetry), and ϕ ranges from zero when trade is perfectly closed ($\tau = \infty$) to unity when trade is perfectly free ($\tau = 1$); we refer to ϕ as the 'freeness' of trade. Observe that the conditions in (2) are potentially very complex since Δ depends upon a_X and a_D but these in turn depend upon the Δ via (1).¹⁰

⁶ More formally, the standard CES demand function is $c_j = E(p_j)^{-\sigma} / (\int p_i^{1-\sigma} di)$ for variety j , where the integral is over all competing varieties and E is total expenditure on all varieties in the market. Given Dixit-Stiglitz 'mill price' (see appendix), the price-marginal cost mark-ups are all identical and thus cancel out, so $c_j = E(a_j)^{-\sigma} / (\int a_i^{1-\sigma} di)$. Now, given the distribution of a 's discussed in the text, this is also equal to $c_j = E(a_j)^{-\sigma} / n (\int_0^{a_D} a_i^{1-\sigma} dG[a] + \phi \int_0^{a_X} a_i^{1-\sigma} dG[a])$.

Note that we have here included all D-type varieties produced in the local market in the first integral and all varieties that are imported from the other market (i.e. varieties with a 's between zero and a_X) in the second integral. Here $\phi \equiv \tau^{1-\sigma}$, measures the iceberg trade costs that are passed on by foreign firms. Multiplying by p_j yields the expression in the text.

⁷ Note that the Dixit-Stiglitz markup cancels out from every price and we use $w=1$, so prices depend only on the a 's.

⁸ In particular, they assume all firm-level uncertainty is resolved at the same moment that the sunk cost is paid.

⁹ Defining $\pi[a]$ as the steady state operating profit earned by a firm with marginal cost ' a ', the present value with a

discount rate ρ and the Poisson firm-death process assumed is: $\pi[a] \int_0^{\infty} e^{-\rho t} e^{-\delta t} dt$ since the probability of the firm still being alive at t is $e^{-\delta t}$. Setting $\rho=0$ as Melitz does and solving the integral yields the expression in the text.

¹⁰ Note that Helpman, Melitz and Yeaple (2004) use a diagram that depicts these cut-off conditions as straight line functions of $a^{1-\sigma}$ and this appears to make the determination of the cut-offs simply the intersection of the lines with the x-axis. This is misleading however since the intersections depend upon the position of the lines and the position of the lines depends upon both intersections (via what we call Δ); note that their $B=E/n\Delta$ in our notation.

One firm-level fact about trade that has been firmly established is that only a fraction of all firms that produce in a nation actually export (see Tybout 2003 for a survey of such findings). In terms of our model, this means $a_X < a_D$ is a regularity condition. From (2), the necessary and sufficient condition for this is that:

$$(3) \quad \frac{a_X}{a_D} < 1 \quad \Leftrightarrow \quad \frac{F_X / \phi}{F_D} > 1$$

We assume this holds throughout the analysis.

2.2.3. The Free Entry Condition

Having determined the cut-offs facing potential entrants, we now turn to determining the mass of firms that will actually be active, namely n .

A potential entrant pays F_I to develop a new variety with a randomly assigned ‘ a ’. After developing the new variety and observing the associated ‘ a ’, the potential entrant decides whether to enter the local only, or both the local and export markets. As discussed above, entering both markets will be profit maximising if the new firm’s ‘ a ’ is such that $a \leq a_X$. Entering only the local market will be profitable if the $a_X \leq a \leq a_D$, and if $a_D \leq a$, the entrepreneur drops the project and lets bygones be bygones.

Plainly, sinking the innovation cost F_I is risky; there is some probability that it will be wasted entirely. Notice, however, from (2) that the reward to being a ‘winner’, i.e. a D- or X-type, increases as the mass of active firms ‘ n ’ falls. Following the standard logic of free entry in the Dixit-Stiglitz model, it is intuitively obvious that ‘ n ’ will adjust such that the expected reward to sinking F_I is zero. The calculation of the equilibrium ‘ n ’, however, is more complicated than in the homogenous firm model since we must deal simultaneously with the possibility that an entrant may end up as a D-, X- or N-type.

The easiest way to find n is to start with a simple fact. Because we focus on steady states, the ex ante likelihood of getting a ‘winner’ with any particular ‘ a ’ is exactly the same as the actual distribution of a ’s in the market (here ‘winner’ means D or X type). In other words, the ex ante expected operating profit of a winner must exactly match the average operating profit earned in the market. This average must be $E/\sigma n$ because total operating profit worldwide equals $2E/\sigma$ (due to the constancy of the Dixit-Stiglitz operating profit margin) and the worldwide mass of varieties is $2n$. Thus:

$$\text{Ex ante expected value of a ‘winner’} = \frac{E}{\sigma n \delta}$$

Of course, innovators do not come up with a winner every try; only new varieties with $a < a_D$ will be ‘winners’. It is straightforward to calculate the ex ante expected fixed cost of getting a winner (i.e. developing a D or X type patent). The answer is:

$$(4) \quad \bar{F} \equiv F_D + F_X \frac{G[a_X]}{G[a_D]} + F_I \frac{1}{G[a_D]}$$

The first right-hand term is the fixed cost for local sales – an expense that every winner will incur. The second term reflects the fact that some ‘winners’ will be X-types so their developer will find it profitable to incur F_X as well; $G[a_X]/G[a_D]$ is the probability of being an X-type conditional on being a winner. The third right-hand term reflects the ex ante expected variety development cost, i.e.

F_1 times the inverse probability of getting a winner on a random draw since, roughly speaking, $1/G[a_D]$ is the number of ‘tries’ needed to get a winner.¹¹

Given the expected cost and benefit of a winner, the expected pure profit of devoting resources to the development of new varieties will be $(E/\sigma n\delta - \bar{F})$. The free entry condition is thus that the expected cost and benefit of sinking entry must match, thus:¹²

$$(5) \quad \frac{E/\sigma}{n\delta} - \bar{F} = 0$$

Pure profits in equilibrium? Luck rentiers

It is important to note that all active firms earn pure profits throughout their entire life in the sense that their revenue exceeds their variable costs by more than would be needed to amortize their sunk costs. The reason, of course, is that ‘n’ must be such that the ex ante expected value of pure losses on N-types is balanced by the ex ante expected value of pure profits on D types and X types. Note that since zero discounting is assumed, these profits are not a payment to reward the foregone consumption wrapped up in the sunk costs. They are rents – rents earned for being lucky. This is one of the unique elements of the HCMFMEC model – one that has not really yet been fully exploited in the literature.

2.2.4. Analytic intractability and the Pareto distribution

All of the analysis up to this point has been conducted without resort to a functional form for G . Indeed much of the subsequent analysis can also be conducted in this manner, but the reasoning is clearer when we have explicit solutions, and this requires an explicit $G[a]$. The point is that given the dependence of Δ and \bar{F} on the conditional and unconditional distributions of the a 's, we cannot solve (2) or (5) explicitly without assuming an explicit functional form for G .

Fortunately, the empirical literature on firm size distribution suggests that a Pareto distribution is a reasonable approximation (e.g., Cabral and Mata 2001). In the HCMFMEC model, size and marginal cost are inversely related, so it is understandable that the Pareto distribution has been adopted as fairly standard practice in the theoretical literature (also because the Pareto is so easy to work with). Thus, henceforth we assume a Pareto distribution whose cumulative density function is:

$$(6) \quad G[a] = \left(\frac{a}{a_0}\right)^k \quad 0 \leq a \leq a_0$$

¹¹ Since we work with a continuum of goods, there are an uncountable number of varieties and a fraction δ of these must be replaced at all moments. Thus innovators must develop an uncountable infinity of new varieties to replace depreciation. By the Law of Large numbers, this means that the average fixed cost per new varieties is exactly \bar{F} at all instants. Or, to put it differently, there is no aggregate uncertainty in this model.

¹² The direct approach to formulating the condition for zero-expected-profit-from-innovation is to calculate the ex ante expected benefit net of market entry costs, i.e. $\int_0^{a_D} \{a^{1-\sigma}(E/\sigma n\Delta\gamma) - F_D\} dG[a] + \phi \int_0^{a_X} \{a^{1-\sigma}(E/\sigma n\Delta\gamma) - F_X\} dG[a] + 0$, which simplifies to $(E/n\sigma\gamma\Delta)(A[a_D] + \phi A[a_X])$ minus $(G[a_D]F_D + G[a_X]F_X)$, where $A[x] \equiv \int_0^x a^{1-\sigma} dG[a]$. Notice, however, that $G[a_D]\Delta = (A[a_D] + \phi A[a_X])$ from (2) and the fact that $dG[a] = g[a]/G[a_D] da$, where $g[a]$ is the pdf of $G[a]$. Thus, the benefit minus beachhead costs is $G[a_D](E/n\sigma\gamma) - G[a_D]F_D - G[a_X]F_X$. This is set equal to F_1 in the direct approach. Dividing this through by $G[a_D]$ validates our indirect approach in (5). We adopt the indirect approach since it allows a direct comparison with standard monopolistic trade models (which have only one fixed costs) and it facilitates analysis by concentrating the impact of openness and parameter changes in the ex ante expected cost of getting a winner, namely \bar{F} .

where k and a_0 are the ‘shape’ and ‘scale’ parameters, respectively. Notice that $k=1$ implies the uniform distribution. One extremely useful feature of the Pareto distribution is its ‘fractals’ nature. That is, any continuous slice of a Pareto distribution is itself a Pareto distribution with the same shape parameter, but with different upper and lower bounds (to think about this, think of the uniform distribution which is a special case of the Pareto). This fractals nature is not common. For example, a truncated normal distribution is not itself a normal distribution. The usefulness of the fractals property will become abundantly clear in the analysis below.

Given the Pareto distribution and assuming the regularity condition $\beta \equiv k/(\sigma-1) > 1$ (so the integrals in Δ converge), we can explicitly solve for Δ and \bar{F} to get¹³:

$$(7) \quad \Delta = \frac{a_D^{1-\sigma}(1+\Omega)}{1-1/\beta}, \quad \bar{F} = \frac{F_D(1+\Omega)}{1-1/\beta}; \quad 0 \leq \Omega \equiv \phi^\beta T^{1-\beta} \leq 1, \quad T \equiv \frac{F_X}{F_D}, \quad \beta \equiv \frac{k}{\sigma-1} > 1$$

Where Ω (a mnemonic for ‘openness’) summarises the impact of beachhead trade barriers and iceberg trade barriers.

Aside on the openness parameter. The variable Ω summarising the two types of trade barriers in the model, so it is worth pointing out four features of Ω that facilitate intuition and subsequent analysis: (1) Ω measures the combined protective effects of higher fixed and variable trade costs; (2) $\Omega=0$ with infinite τ and/or infinite F_X/F_D , (3) $\Omega=1$ with zero iceberg costs and $F_X=F_D$; (4) we can also express Ω as $\phi(F_X/F_D\phi)^{1-\beta}$ and this tells us that as long as the inequality in (3) holds, Ω is bound between zero and unity.

Expenditure. Finally, we note that the assumption of a zero discount rate means that the foregone consumption necessary to create new varieties requires no compensation. Thus, the only source of current income is labour, so:¹⁴

$$E = L$$

If the discount rate ρ is not assumed to be zero, the ‘capital’ in this model – namely the foregone consumption tied up in the sunk costs – must be paid a reward. Here E equals permanent income, namely, the income from labour, L , plus the rental rate on the steady-state value of the nation’s capital stock, namely $E=L+\rho n\bar{F}$. Taking $\rho=0$ as in most applications of the HCMFMEC model, $E=L$ and using this, (7), the two cut-off conditions (2), and the free entry condition (5), we get explicit, closed form solutions for n , a_D and a_X :¹⁵

$$(8) \quad n = \frac{L(\beta-1)/\delta}{\beta F_D(1+\Omega)\sigma}; \quad a_D = a_0 \left(\frac{F_I(\beta-1)}{F_D(1+\Omega)} \right)^{\frac{1}{k}}; \quad a_X = a_0 \left(\frac{\Omega(\beta-1)F_I}{(1+\Omega)F_X} \right)^{\frac{1}{k}}; \quad n_C = \frac{L(\beta-1)}{\sigma F_X \beta} \left(\frac{T+\Omega}{1+\Omega} \right)$$

¹³ The expression for Δ follows directly from (2) and (6); specifically $\Delta = \{k/(1-\sigma+k)\}(a_D^{1-\sigma})(1+\phi(a_X/a_D)^{1-\sigma+k})$. Using the ratio of the cut-offs, $\beta \equiv k/(\sigma-1)$ and our definition of Ω , $\Delta = \{1/(1-1/\beta)\}(a_D^{1-\sigma})(1+\Omega)$. As for the expected fixed cost \bar{F} we plug our new expression for Δ into the D-type cut-off condition to get $a_D^{1-\sigma}(E/n\sigma)\{(1-1/\beta)/(1+\Omega)\}=F_D$. But from (5) we know, $\bar{F} = E/n\sigma$, so must equal $\bar{F} = F_D\{(1+\Omega)/(1-1/\beta)\}$.

¹⁴ To see this, note that expenditure is total income less spending on the replacement of ‘dead’ varieties, i.e. $E=L+E/\sigma-L_I$, where L is labour income, E/σ is total payments to ‘capital’, i.e. operating profit, and L_I is the labour employed in developing new varieties and market entry costs. The free entry condition with discounting is $E/\sigma = n(\delta)\bar{F}$, and the replacement cost is $\delta n\bar{F}$, so $E=L$.

¹⁵ To find n , plug Δ from (7) into the D-type cut-off condition. To find a_D , plug the closed form solution for n into the free entry condition (4), using the ratio of cut-off conditions to evaluate $G[a_X]/G[a_D]$. a_X then follows from this the expression for a_D and the ratio of the cut-off conditions.

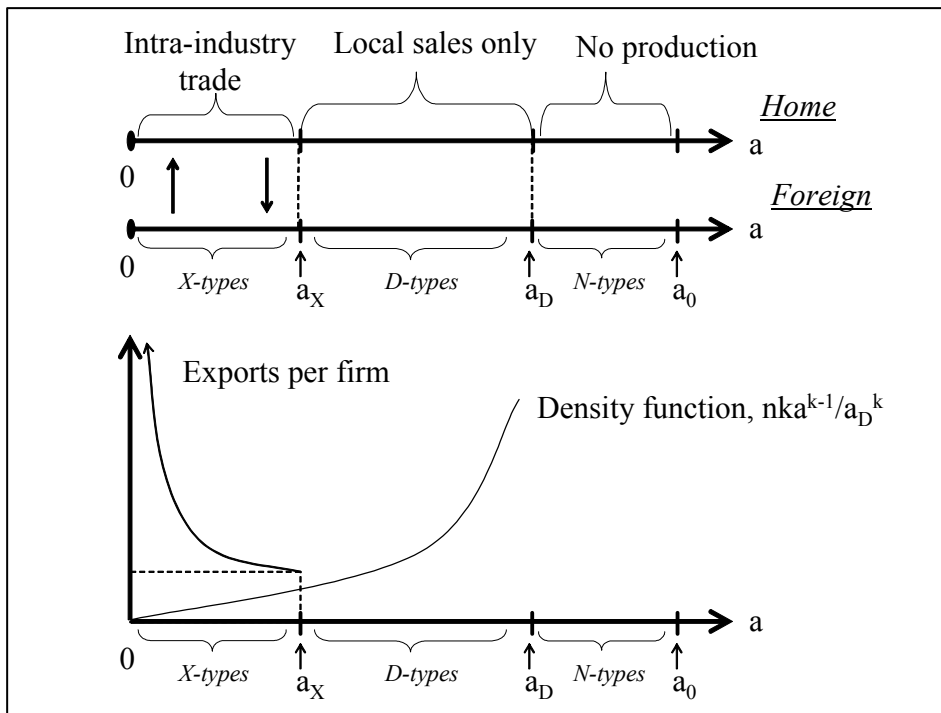
Notice that unlike the standard Dixit-Stiglitz trade model, not all varieties are consumed by all agents since some are sold only locally. Defining n_C as the number of varieties consumed in a typical nation, note that Baldwin and Forslid (2004) derive the expression for n_C in (8), where $T \equiv F_X/F_D$ is a measure of the fixed trade barriers.¹⁶

A number of additional intermediate results that prove useful in subsequent analysis are:

$$(9) \frac{a_X}{a_D} = \left(\frac{\phi}{T}\right)^{\frac{1}{\sigma-1}}, \quad \Delta = a_0^{1-\sigma} \left(\frac{F_D}{\beta F_I}\right)^{\frac{1}{\beta}} \left(\frac{1+\Omega}{1-1/\beta}\right)^{1+\frac{1}{\beta}}, \quad P = a_0 \frac{(\sigma\delta/L)^{\frac{1}{\sigma-1}}}{1-1/\sigma} \left(F_I \frac{\beta-1}{1+\Omega}\right)^{\frac{1}{k}} (F_D)^{(\beta-1)/k}$$

where P is the standard CES price index.

Figure 2: The trade pattern.



2.3. Trade pattern

One of the stark differences between the HCMFMED model and the standard homogenous-firms trade model concerns the export pattern. In particular, only a fraction of firms export their goods. This is a big step forward in reality and, we shall see below, is crucial to several of the models unique results on productivity and welfare. The model displays standard intra-industry trade in differentiated varieties produced by X-types, but the varieties of D-types are non-traded even though they would be classified as being in a ‘traded goods’ sector. This is shown in the top panel of Figure 2.

¹⁶ In a pdf file posted on various internet sites and marked “Preliminary and Incomplete”, Melitz and Ottaviano (2003) study a heterogeneous-firms trade model assuming a linear demand system. Crucially, however, they do not allow for fixed market entry costs, so their model lacks one of the two critical and novel elements of the HCMFMED model. In this alternative model, they derive n_C . The linear demands imply that price-cost mark-ups vary with the degree of competition so their model displays a number of features that are ruled out by Dixit-Stiglitz competition. For example their model predicts that firms in large nations charge a lower markup.

2.3.1. Trade volume and pattern

The value of exports of a typical X-type firm tends to infinity as ‘a’ approaches zero and equals $\sigma\delta F_X$ for $a=a_X$. For X-firms with intermediate a’s, the value of exports is: $a^{1-\sigma}\phi L/\Delta\sigma$, but the export cut-off condition tells us $\phi L/\Delta n=\sigma\delta F_X/a_X^{1-\sigma}$, thus:

$$(10) \quad v[a] = \left(\frac{a}{a_X}\right)^{1-\sigma} \delta \sigma F_X, \quad V = L \frac{\Omega}{1+\Omega}, \quad a \in [0, a_X]$$

where $v[a]$ is the per firm export. Integrating over all X-types (weighting by frequency), we have V , which is defined as the total value of exports in terms of the numeraire, i.e. $\int_0^{a_X} v[a] dG[a]$ integrated from 0 to a_X . Notice that a standard measure of openness, i.e. imports to GDP is proportional to, $\Omega/(1+\Omega)$, since GDP, which equals factor income $L+E/\sigma$, equals $L(1+1/\sigma)$. The trade volume as a function of ‘a’ is depicted in Figure 2.

2.3.2. Unit values and quality

It is interesting to note that the underlying assumptions behind the Greenaway, Hine and Milner, (1994 and 1995) approach to ‘horizontal’ intra-industry trade (IIT) and vertical IIT fail in this model. Greenaway and Milner, and more recently, Schott (2003), use unit value indices to deduce product quality – with higher prices indicating higher quality; the underlying assumption is that the trade classifications are too broad and thus group together goods that are fundamentally different. In the HCMFMEC model, the goods are absolutely symmetric in terms of product characteristics but nevertheless have very different prices. In particular, the Greenaway-Milner approach would view the varieties produced by relatively inefficient X-types as being higher quality.

2.4. The ‘knowledge capital stocks’

There is only one primary factor in the HCMFMEC model, namely labour. However, the sunk cost nature of the variety-development cost, F_1 , makes it useful to introduce the notion of a second factor of production, which we can call capital. That is, investment of F_1 units of labour produces a tangible body of knowledge – the knowledge of how to produce a unique variety – that may earn a positive reward throughout its life. In this terminology, the factor reward paid to a unit of knowledge capital – or unit of capital for short – depends upon its associated marginal cost. Thus, as shown by Baldwin and Forsild (2003), there are three types of capital in this model, D-type, X-type and N-type. The factor reward for a D-type and X-type are, respectively:

$$r_D[a] = a^{1-\sigma} \frac{E}{n\Delta\sigma}; \quad a_X < a < a_D, \quad r_X[a] = a^{1-\sigma} \frac{E(1+\phi)}{n\Delta\sigma}; \quad 0 < a < a_X$$

Where r_D is the reward to D-type capital and r_X is the reward to X-type capital. The factor reward to N-type knowledge capital is zero. Given the firm-death process assumed, a unit of capital earns its reward right up to the moment it dies.

2.5. Social optimality of the thresholds and mass of firms

There are two standard welfare issues to consider: 1) does the market provide the right mass of varieties? And, 2) does the market provide the right cut-offs? These questions were answered by Baldwin and Robert-Nicoud (2005) following the Grossman and Helpman (1991 Appendix A.3.3) procedure of focusing on two externalities. The entering firm does not consider the extra consumer surplus it generates since it cares only about profit, but it also does not consider the impact it has on the profits of existing varieties. As it turns out, these two exactly offset each other in this model.

2.5.1. Optimal mass of firms

Denoting the sum of all operating profit as Π , the indirect utility function is:

$$U = \ln E - \ln P; \quad P \equiv (n\Delta)^{1/(1-\sigma)} / (1 - \frac{1}{\sigma}); \quad E \equiv L - L_1 + \Pi$$

where L_1 is the labour employed in meeting the fixed costs of new varieties. We consider a small perturbation of the laissez-faire n and evaluate whether this raises utility. Differentiating with respect to n at time T :

$$(11) \quad \frac{dU}{dn} = \frac{1}{E_T} \frac{dE_T}{dn} - \frac{1}{P} \frac{d(P)}{dn}$$

Noting that Δ is independent of n , differentiation of P yields the consumer surplus effect directly:

$$(12) \quad \frac{d \ln P}{dn} = \frac{-1}{n(\sigma - 1)}$$

The profit destruction effect is more subtle. We are considering a slight perturbation of the steady-state n . Having an extra variety at time t_0 entails a slight increase in investment at t_0 , but this creates a flow of operating profit as well. Since we are working from the steady-state, where the cost and benefit of an extra variety are equal, the direct impact of the extra variety on E is nil. However, this extra variety depresses the profit of all existing varieties, so $dE/dn = d\Pi/dn$. Specifically, the average domestic operating profit falls from $E/\sigma n$ to $E/\sigma(n+\Delta)$. Multiplying this by the unperturbed n :

$$\frac{d\Pi}{dn} = -\frac{E}{\sigma n} + \frac{1}{\sigma} \frac{dE}{dn}$$

Using this and $dE/dn = d\Pi/dn$, we have $(1/E)dE/dn = -1/n(\sigma-1)$. Plugging this and (12) into (11) we see that $dU/dn = 0$. In short, the consumer surplus effect and the profit destruction effect exactly cancel out, so the laissez-faire mass of varieties is also socially optimal (as usual in the static Dixit-Stiglitz model and its endogenous growth applications; this is true in a constrained optimum sense only, see Dixit and Stiglitz, 1977).

2.5.2. Optimal cut-offs

Evaluation of the socially optimal cut-offs follows a similar procedure. We consider the steady state with the market-determined cut-offs and consider a small perturbation in these cut-offs. Consider first a small increase in a_D , i.e. each nation would see a small increase in the number of locally produced varieties. As before, introduction of these new varieties would have a direct effect – the cost of their development minus the present value of their operating profit – but this would be zero given the cut-off condition that defines a_D ; the same holds for a_X . Additionally, there would be an indirect effect on consumer surplus and profit destruction as before, however once again these will cancel out. The basic reason is that $1/P = (n\Delta)^{1/(\sigma-1)} / (1-1/\sigma)$ and the average profit of home firms falls from $\Delta(E/n\Delta\sigma + \phi E/n\Delta\sigma)$ to $\Delta(E/n(\Delta+d\Delta)\sigma + \phi E/n(\Delta+d\Delta)\sigma)$. Using the same procedure as above, it is easy to show that $dU/d\Delta = 0$ when considering a slight perturbation of Δ around its steady-state. Plainly this is true whether the perturbation of Δ comes from a perturbation of a_D or a_X . To summarise:¹⁷

¹⁷ This was first shown by Baldwin and Robert-Nicoud (2005).

Result 1: The laissez-faire mass of varieties and cut-off points along the steady-state growth path are socially optimal.

3. POSITIVE EFFECTS OF LIBERALISATION

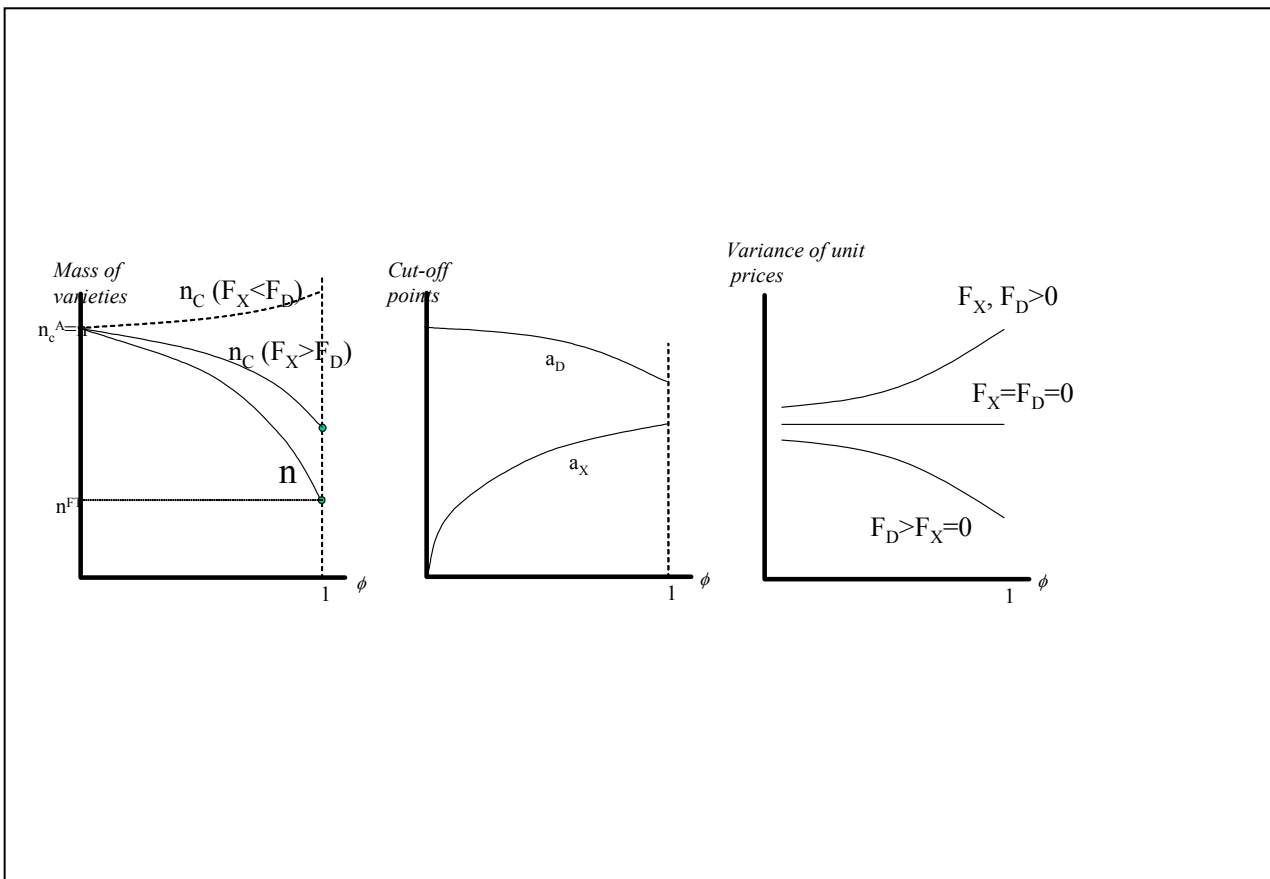
The HCMFMEC model has been used for a number of purposes and will undoubtedly be used for many more. However, the early papers Melitz (2003) and Bernard, Eaton, Jensen and Kortum (2003) focused on productivity effects of trade liberalisation. Before turning to the demonstration of these results, however, we study the standard questions that arise when liberalisation occurs, namely how does freer trade affect the trade pattern, the trade volume and welfare.

Importantly, there are two natural definitions of trade liberalisation in this model, one concerns the variable cost of trade ϕ and the other concerns the differential beachhead cost for local and imported varieties F_X/F_D . As we shall see, the two types of trade barriers usually affect variables in an isomorphic manner since they are combined into the aggregate measure of openness Ω .

3.1. Lower marginal cost of trade

We begin by considering lowering marginal trade costs, i.e. freer trade in the sense of $d\phi > 0$.

Figure 3: Varieties produced and consumed and cut-off points



3.1.1. Impact on cut-offs and mass of firms

As inspection of (8) reveals, greater openness (i.e. $d\Omega > 0$, which corresponds to $dT < 0$ and/or $d\phi > 0$) lowers a_D and raises a_X . When trade is at zero freeness (i.e. infinite trade costs) $a_X = 0$, i.e. even a firm with zero marginal cost cannot export. Greater openness lifts a_X while lowering a_D , but the two do not meet even when trade is costless (i.e. $\phi = 1$), as long as $T = F_X/F_D > 1$. In words, these results

mean that freer trade lowers the maximum marginal cost of active firms and raises the maximum marginal cost of exporting firms. The two cut-off marginal costs meet when trade is costless, only if there is no regulatory protection, i.e. $F_X=F_D$. In the focal case where $F_X>F_D$, not all varieties are exported (i.e. $a_D>a_X$) even when trade is costless. This is shown in the middle panel of Figure 3.

Turning to the mass of varieties, inspection of (8) shows that freer trade lowers the number of varieties produced in each nation. The proportional change in n with respect to a proportional change in trade freeness is:

$$\hat{n} = -\beta \frac{\Omega}{1+\Omega} \hat{\phi}$$

where we have used the standard ‘hat’ notation for proportional changes (e.g. \hat{x} equals dx/x), and $\Omega/(1+\Omega)$ is the import share (i.e. expenditure share on all imported varieties in a typical market).¹⁸ Since the import share rises with openness, this expression tells us that the proportional decline in ‘ n ’ is magnified, as trade gets progressively freer, as shown in the left panel of Figure 3.

The impact on the range of varieties available to a typical consumer could be ambiguous, in principle (Melitz 2003). Greater openness raises a_X and thus raises the fraction of Foreign-made varieties that are imported to Home so the rise in imported varieties could more than offset the drop in locally produced varieties. In the model at hand, however, inspection of (8) the number of varieties bought by a typical consumer falls monotonically as the freeness of trade rises – as long as $T \equiv F_X/F_D > 1$. Thus, lower variable cost of trade will produce an ‘anti-variety’ effect, i.e. the range of consumed varieties falls as trade gets freer. The two cases are shown in the left panel of Figure 3.

The basic intuition for this result flows most easily by first examining the knife-edge case of no fixed-cost protection, i.e. $F_X/F_D=1$. As inspection of (8) shows, the number of varieties consumed is constant with respect to ϕ . If there is no intrinsic difference between local and imported varieties, changes such as trade liberalisation that introduce more imported varieties will produce a one-for-one reduction in local varieties.

More generally, when the ratio of beachhead costs F_X/F_D exceeds unity, n_C falls as ϕ rises because imported varieties have systematically lower prices than the domestic varieties they displace. Restoring zero profits thus requires more than one D-type variety to be displaced by each extra X-type variety is imported. Conversely, if the beachhead costs are lower for imported varieties, the relationship is reversed and freer trade means a wider range of varieties available to consumers. To summarise these results, we write:

Result 2: Raising the freeness of trade unambiguously lowers a_D and raises a_X . The two coincide under perfectly free trade ($\phi=1$) only if the two market entry costs are identical, i.e. $F_D=F_X$.

Result 3: Raising the freeness of trade unambiguously lowers the range of varieties produced in each nation (Melitz 2003). The range of varieties consumed in each nation falls with openness as long as the beachhead costs for imported varieties exceeds the beachhead costs for local varieties; if the two costs are equal, there is no change in n_C and if $F_D>F_X$, the range of varieties consumed increases (Baldwin and Forslid 2004).

¹⁸ The market share of imported variety j is $\phi a_j^{1-\sigma}/\Delta n$, so integrating over all foreign varieties $a_j \in [0..a_X]$, the import share equals $\Omega/(1+\Omega)$.

3.1.2. Trade Volume and Pattern

As inspection of (10) shows, total trade flows increase in proportion with the level of trade freeness, ϕ . Moreover, the effect is dampened as trade gets freer:

$$\hat{V} = \frac{\hat{\Omega}}{1 + \Omega}; \quad \hat{\Omega} = \beta \hat{\phi} + (1 - \beta) \hat{T}$$

The change in the trade pattern can easily be seen from (8), (10) and Figure 2. Freer trade raises a_X and lowers a_D , so a higher fraction of the nation's GDP is exported, but all of the new trade (which comes from the rise in a_X) is two-way trade in similar products.

There are four more subtle points that could suggest empirically testable hypotheses. First, the exports of each existing X-type firm expand in proportion. From (10), the change in firm-level export as a function of 'a' (and thus the size of firms) is:

$$\frac{dv[a]}{d\Omega} = v[a] \left(\frac{\sigma - 1}{a_X} \right) \frac{da_X}{d\Omega}$$

Second, every new exporter should be smaller (in the sense of value of domestic sales) than every existing export since the drop in a_X affects firms with a's that were just below the cut-off before the liberalisation. To summarise:

Result 4: Freer trade should raise the exports of all firms in proportion (regardless of firm size) and all of the new exporters should be smaller than all of the existing exporters.

Plainly, this hypothesis would be clouded if one considers transitional effects and/or a firm-entry process that is more realistic than the Hopenhayn-Melitz approach assumed here.

The third type of prediction concerns the range of unit prices. To illustrate this, we step slightly outside the simple model that we are working with. Suppose that the economy has three manufacturing sectors instead of just one, with constant expenditure shares on each. Moreover, suppose trade-statisticians lump together all the trade in each of the three sectors, ignoring variety level differences. Furthermore, suppose that the first manufacturing sector is exactly as described above, marked by firm heterogeneity and beachhead costs for local and export markets. The second manufacturing sector is marked by firm heterogeneity but no export beachhead costs (although we allow domestic market entry costs), so $F_D > F_X = 0$. The third manufacturing sector is marked by firm heterogeneity but no fixed beachhead costs of any type, i.e. $F_D = F_X = 0$.

The impact of freer trade on the variance of unit price indices for the first sector, i.e. the sector with heterogeneous marginal costs and fixed market entry costs (HMCFMEC) is easily calculated using the usual Dixit-Stiglitz markup, $p = a/(1 - 1/\sigma)$, and (6), namely:

$$\text{var}^{het1} = k \left(\frac{a_X}{1 - 1/\sigma} \right)^2 \left(\frac{\phi}{T} \right)^\beta \left(\frac{1}{2 + k} - \frac{k(\phi/T)^\beta}{(1 + k)^2} \right)$$

The link between unit price dispersion and trade freeness is reversed in sector two where $F_D > F_X = 0$. Namely:

$$\text{var}^{het2} = k \left(\frac{a_D}{1 - 1/\sigma} \right)^2 \left(\frac{1}{2 + k} - \frac{k}{(1 + k)^2} \right)$$

For the third sector that lacks all fixed beachhead costs,

$$\text{var}^{\text{het}3} = k \left(\frac{a_0}{1-1/\sigma} \right)^2 \left(\frac{1}{2+k} - \frac{k}{(1+k)^2} \right)$$

Given result 2, it is plain that the change in the variance of unit price indices provides of discriminating among sectors as to the importance of firm heterogeneity and fixed market entry costs. To summary:

Result 5: Freer trade should raise the variance of unit prices in sectors marked by firm-level heterogeneity and beachhead costs for domestic and export sales (i.e. $d(\text{var}^{\text{het}1}/d\phi > 0$); but for sectors marked by heterogeneity and beachhead costs only for domestic sales ($F_D > F_X = 0$), the sign of the impact is reversed $d(\text{var}^{\text{het}2})/d\phi < 0$. For sectors without firm-level heterogeneity, and/or without any fixed market entry costs, $d(\text{var}^{\text{het}2})/d\phi = 0$.

This is depicted schematically in the right panel of Figure 3. As we shall see below, free trade has a pro-productivity effect in either of the first two cases.

The final point concerns ‘zeros’ in the trade matrix. Empirically oriented trade economists have long know that there are many zero bilateral trade flows in the world.¹⁹ The facts have been more recently documented in a systematic fashion by Feenstra and Rose (1997) and Helpman, Melitz and Rubinstein (2004). Stepping slightly outside of our two nation model, one simple empirically testable implication of the HMCFMEC model concerns the pattern of zeros.

First, as has already been indirectly pointed out by Helpman, Melitz and Rubinstein (2004), the pattern of zeros bilateral aggregate trade flows should follow a geographical pattern, assuming that trade costs increase with distance. This is easily testable and indeed the first stage of the Helpman-Melitz-Rubinstein results seem to confirm this. Notice, however, that one could get much more power by testing for zeros at the very finely defined commodity level and using the time dimension of the data. For example, taking the US’s very finely disaggregated and easily available export data, one should find the likelihood of a zero (controlling for the usual gravity equation issues like economic size of the importing nation) increasing with distance. Likewise, as a result of any well defined liberalisation exercise, such as the phase in of Uruguay Round tariff cuts, one should find that the impact of distance on the zeros diminishes as tariffs are cut.

Result 6: Due to the existence of market-specific beachhead costs, the likelihood of observing a zero bilateral trade flow should increase with variable trade costs, e.g. as proxied for by distance, when one controls for other factors such as the size of the import market. It is possible that the size of the beachhead cost and the market size are correlated, so one might observe an interaction between distance and market size that is either positive or negative. The estimated coefficient on distance should diminish during a liberalisation of other variable trade costs.

Second, some form of trade liberalisation are more likely to reduced beachhead costs than variable trade costs. One common source of beachhead costs are known as technical barriers to trade (TBTs) many of which involve health, safety and environment certification of new products (see Baldwin 2000 for details). One way in which such measures are reduced is via international agreements – e.g. Mutual Recognition Agreements either on testing (US-EU MRA) or product norms (New Approach Directives in the EU). These agreements should diminish the probability of observing a zero in any given bilateral trade flow that is affected. A simple difference-in-difference approach should pick

¹⁹ For example the old gravity equation literature pondered on the best solution to this with some authors dropping these observations, others performing Tobit regressions and other plugging in small positive values. See, e.g. Wang and Winters 1992.

this up on aggregate or disaggregate data. Notice that a reciprocal MRA predicts that the effects should be two-way in the affected sectors.

Result 7: Trade liberalisation agreements that focus on fixed market entry costs (e.g. MRAs) should reduce the positive impact of distance on the likelihood of a zero. If the MRA is reciprocal (most are), the impact should occur in both directions in the affected sectors.

Third, and more to the heart of the HMCFMCEC model logic, one should find a pattern in firm-level zeros in the data. Specifically, there should be a positive correlation between a firm's domestic market share (which varies with its marginal cost) and the number of markets to which it exports, or the likelihood that it has a zero in any given market, controlling for standard market-specific factors. If one expands the model to allow for the standard proximity-versus-scale FDI a la Helpman-Melitz-Yeaple, then the prediction is still true but it is for the number of markets in which the firm sells (via location production is it is going for proximity or exports if it is going for scale economies).

Result 8: In firm-level trade data, the HMCFMCEC model has clear empirically testable predictions. One should observe a positive correlation between a firm's domestic market share and the likelihood that it has positive sales (via exports or local production) to any given market. The firm-level pattern of zeros should indicate that bigger firms have fewer zeros and those zeros are for markets that are further away.

Result 9: As with the national-level bilateral trade data, one should observe an impact on the zeros pattern of both variable and fixed cost trade liberalisation schemes. If the liberalisation is well-defined in time, such effects should appear clearly in difference-in-difference regressions.

3.1.3. Productivity effects of liberalisation

We turn next to the productivity effects of liberalisation. Melitz (2003) shows that liberalisation has a strong impact on the average productivity of an industry via selection effects (lowering the maximum marginal cost of producing firms by lowering a_D) and production reallocation effects (reallocating production shares from the least efficient to the most efficient firms). Or, to use a sports analogy, the batting average rises since freer trade eliminates the worst batters and gives the best batters more swings.

Defining 'measured productivity' as the ratio of real output to labour input, it is easy to show that openness boosts productivity. More specifically and considering only the manufacturing sector, total sales and labour input per nation are E and L respectively, so the average value of sales per production worker is $E/L=1$ (recall $E=L$ when zero discounting). The ideal price index for national firms is $(n\Delta)^{1/(1-\sigma)}/(1-1/\sigma)$. Using (7) and (8), labour productivity measured as real output per worker is²⁰:

$$\frac{1-1/\sigma}{a_D} \left(\frac{L}{\sigma \delta F_D} \right)^{1/(\sigma-1)}$$

Since greater openness lowers a_D , we see it also raises industry's average productivity.

²⁰ The expression is $\{E/L\}/P^P$, where P^P is the producer price index. Substituting $P^P = \{a_D/(1-1/\sigma)\} \{n/(1-1/\beta)\}^{1/(1-\sigma)}$, which is the solution with the Pareto distribution, yields the expression in the text.

Selection and share shifting

The liberalisation-induced productivity rise is decomposed to a selection and share-shifting effect by Melitz (2003). The traditional focus is on lower marginal trade costs, but liberalisation such as the EU’s Single Market Programme probably had a big impact on the fixed cost of trade as well, so we consider the two separately.

A typical X-firm’s sales are $a^{1-\sigma}(1+\phi)E/n\Delta$, while that of a D-type is $a^{1-\sigma}E/n\Delta$. Employing (7) and (8), these are:

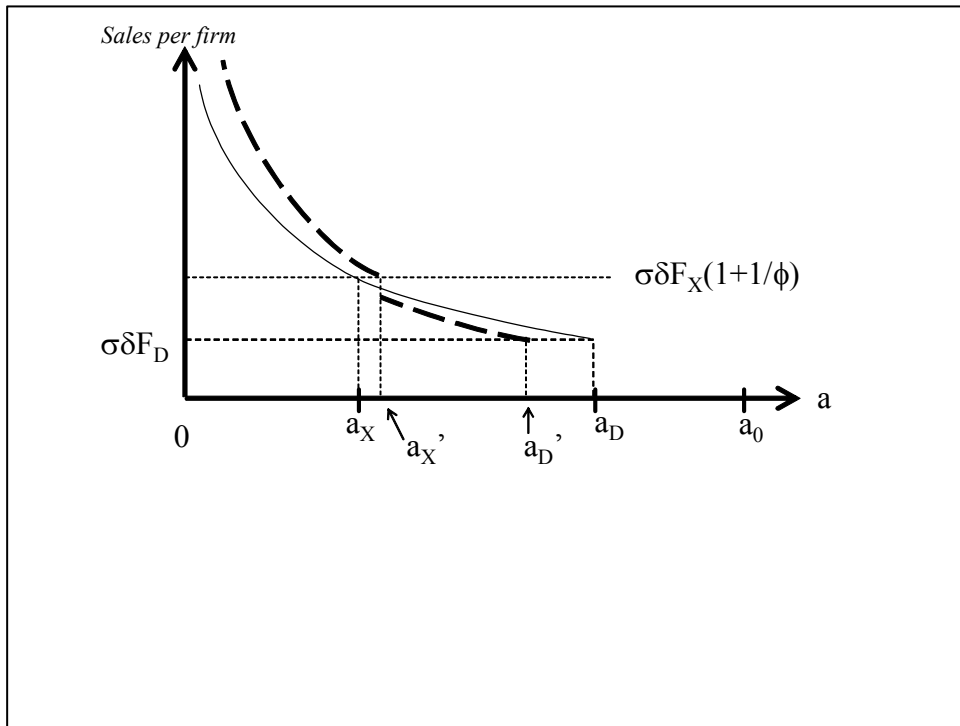
$$S_X[a] = \left(\frac{a}{a_D}\right)^{1-\sigma} (1+\phi)\sigma\delta F_D; \quad S_D[a] = \left(\frac{a}{a_D}\right)^{1-\sigma} \sigma\delta F_D$$

where S’s represent firm sales and the first expression is for X-types and the second for D-types.

(13)
$$\hat{S}_X[a] = \frac{\phi - \Omega}{(1 + \phi)(1 + \Omega)} \hat{\phi}; \quad \hat{S}_D[a] = \frac{-\Omega}{1 + \Omega} \hat{\phi}$$

The second expression is obviously negative and the first is positive as long as $T > \phi$, which is guaranteed by the regularity condition (3).

Figure 4: Freer trade’s impact on firm-level sales.



Demonstrating the selection is even easier. The selection effect refers to the fact that freer trade eliminates the firms with the highest marginal cost. This is perfectly captured by the change in a_D . As we know from the results above a_D drops as ϕ rises. The two effects are illustrated in Figure 4.

Turning to the productivity impact of lower fixed trade costs, F_X , we note that all the effects are qualitatively identical due to Result 11.

An alternative productivity measure

The results for measured productivity discussed above are based on ideal price indices and were first demonstrated by Baldwin and Robert-Nicoud (2004). This, however, was not the measure used in the early HMCFMEC papers. For example, Melitz (2003) focuses on the weighted average of the a 's using relative output shares as weights. His measure is basically the CES price index ignoring the mass of firms and the constant mark-up, i.e. $\Delta^{1/(\sigma-1)}$ in our notation, or $\tilde{\varphi}$ in his.²¹ It is easy to see, using (9) that the same selection and share-shift effect appear with this measure of productivity. To summarise:

Result 10: Greater openness raises industrial productivity via a selection and share-shifting effect.

3.2. Lower fixed-cost protection and domestic de-regulation

The classic notion of liberalisation is a lowering of marginal trading costs, however many of the trade barriers remaining among industrialised nations are related to standards and regulation that make it difficult to introduce foreign-produced varieties into a market. These barriers, called technical barriers to trade (TBTs in WTO circles) are some of the few remaining barriers to trade in manufactures among the US, Canada, the EU and Japan. Moreover, since classic trade barriers were eliminated in Western Europe by the mid-1970s, the last four decades of trade liberalisation in Western European nations has been mainly concerned with TBTs. This suggests that it is important to analyse the positive and normative implications of lowering the gap between the beachhead cost facing local and imported varieties. In particular, in this section we assume that these regulatory barriers are reflected in the beachhead costs and that this explains why $F_X > F_D$, so liberalisation involves a moving $T = F_X/F_D$ towards unity. Of course, it is not possible to change T without change F_D and/or F_X , so we must be explicit about how T falls. To be concrete, consider a ceteris paribus reduction in F_X as a fixed-cost trade liberalisation, while a lowering of F_D and F_X in tandem (such that T is unaffected) is a domestic de-regulation.

As an aside, we note that the formulation of the model makes F_D and F_X entirely distinct. That is, we cannot think of F_D being part of the cost of, say, establishing a product's safety in the home market, and F_X as the extra (lower) cost of using the basic domestic results to establish the product's safety in the foreign market. If this were the case, then the export and local market entry conditions would be linked in the sense that the cost of entering the foreign market would be higher for firms that had not entered the domestic market.²²

Fixed cost trade liberalisation has qualitatively identical effects as $d\phi > 1$ when it comes n , a_D , a_X and the volume of trade. That is to say, lowering F_X lowers n , a_D and raises a_X and the trade volume.

Result 11: Lower fixed trade costs (i.e. $dF_X < 0$) has the same impact as lower marginal trade costs (i.e. $d\phi < 0$) for Result 2, Result 4 and Result 5.

When consider the impact on the number of varieties consumed, however, the impact is ambiguous. The direct effect of lowering F_X is to raise n_C since more varieties are imported. However, the

²¹ For the closed economy case, this definition of weighted productivity is justified as follows in the Melitz article (see footnote 9). Writing $\tilde{\varphi} = \int_0^{a_D} a \left(\frac{a^{-\sigma}}{\tilde{\varphi}^{-\sigma}} \right) \frac{ka^{k-1}}{a_D^k} da$ and we see that the idea is that the weight on firm j 's marginal cost, a_j ,

is the ratio of j 's output (which is proportional to $a_j^{-\sigma}$) over the output of a firm with the weighted average marginal cost equal to $\tilde{\varphi}$ (which is proportional to $\tilde{\varphi}^{-\sigma}$).

²² To see this, consider $\phi=1$; if $F_X < F_D$, then in the standard HMCFMEC model, $a_X > a_D$, in other words, some domestic producers would only export and not sell in the local market. If F_X is viewed as an extra cost on top of F_D , then $a_X \leq a_D$ for all ϕ .

indirect effect, via the impact on overall openness is negative, as long as we start with $T > 1$. Of course, if entering the foreign market is cheaper than entering the domestic market, then both effects work in the same direction.

4. WELFARE EFFECTS OF LIBERALISATION

We turn next to the welfare effects of liberalisation in this model, focusing on the aggregate and redistributive impact of greater openness.

4.1.1. Aggregate Gains from Trade

As noted above, utility of a typical agent in this model can be described by the indirect utility function E/P , where P is the standard CES price index. Ignoring any possible differences between consumers (due, for instance to their ownership of various forms of capital) and long differentiating P from (9):

$$(14) \quad \hat{U} = -\hat{P}; \quad \hat{P} = \hat{a}_D, \quad \hat{a}_D = \frac{\Omega}{k(1+\Omega)}(\beta\hat{\phi} + (1-\beta)\hat{T})$$

Thus we see that the so-called selection effect of greater openness is a necessary and sufficient condition for overall gains from trade. To summarise:

Result 12: The aggregate gain from greater openness is always positive; the gains from marginal liberalisation are increasing as the world gets more open (i.e. Ω approaches unity).

4.1.2. Stolper-Samuelson result

Income distribution effects can also be easily worked out. Indeed, this model displays classic Stolper-Samuelson-like behaviour (Baldwin and Forslid 2004).

There is only one primary factor in this model, however we can think of firm owners as owning ‘knowledge capital’. In particular, we can think of there as three types of capital in this model: D-type capital, X-type capital and N-type capital, where the reward to D-type and X-type capital are the operating profit on D-type and X-type firms, respectively. Recall that although the average reward to the three types of capital must be zero (zero profit condition), this average consists of pure losses for some balanced by pure profits for others; D-type and X-type firm owners earn pure profits while drawers of ‘losing’ varieties earn the flow equivalent of minus F_L .

The reward to capital is a firm’s Ricardian surplus, namely S_X/σ or S_D/σ , for X and D types respectively, where the S ’s are defined in (13). Thus:

$$(15) \quad r_X[a] = \left(\frac{a}{a_D}\right)^{1-\sigma} (1+\phi)\delta F_D; \quad r_D[a] = \left(\frac{a}{a_D}\right)^{1-\sigma} \delta F_D$$

where we define r_D and r_X as the Ricardian surplus of typical D and X type firms.

Of the three ‘factors’, labour is the simplest to deal with. Labour is numeraire so freer trade has no impact on the wage in terms of the numeraire good. The impact on the rental rates on D-type and X-type capital are also as simple to derive. As noted above, a firm’s total operating profit is proportional to its sales, so (13) implies:

$$\frac{\hat{r}_D}{\hat{\phi}} = \frac{-\Omega}{1+\Omega} \leq 0, \quad \frac{\hat{r}_X}{\hat{\phi}} = \frac{\Omega}{(1+\phi)(1+\Omega)} \left(\frac{\phi}{\Omega} - 1\right) \geq 0, \quad \hat{w} = 0$$

where we use the usual hat notation.

Turning to real factor rewards, it is easy to show that if the elasticity of substitution among varieties is sufficiently high (specifically, $\sigma > 2$), then we get a Stolper-Samuelson chain:²³

$$(16) \quad \hat{r}_D - \hat{P} < 0 < \hat{w} - \hat{P} < \hat{r}_X - \hat{P}$$

Note that even if σ violates the condition $\sigma > 2$, we still have that the real gain to X-types exceeds that of D-types.

An interesting implication of (16) when combined with the fact that rental rates are inversely proportional to α 's is that the income distribution among active-firm owners follows a fractals-like pattern. That is, capital rental rates will follow a Pareto distribution with the shape parameter $\rho + 1 - \sigma$. Thus, if, for example, $y\%$ the gains from liberalisation accrue to the top $x\%$ of the income distribution, then the same is true of the top $x\%$ of the top $x\%$. This fractal-like income distribution has received some empirical support from income distribution studies. To summarise:

Result 13: Freer trade widens the factor-income distribution in the sense that it widens the gap between wage and the average pure profit earned by active firms. It also widens the income gap between X-type firm-owners and D-type firm-owners since it raises r_X and lowers r_D . Finally, the distribution of profits among the winners, namely the X-type firm-owners, follows a fractal-like pattern, with most of the gain accruing to the firm-owners that earn the highest profit initially.

Result 14: As long as varieties are sufficiently substitutable, the model is marked by a standard Stolper-Samuelson inequality chain with some reinterpretation. We interpret the claims to D-type and X-type varieties as D-type and X-type capital respectively. Noting that D-type varieties are import competitors while X-type varieties are net exporters, we find that the factor used intensively in exporting (X-type capital) gains absolutely while the import-competing factor loses absolutely. Labour is used intensively in a sector without net trade and its reward is unaffected by liberalisation.

Result 15: This Stolper-Samuelson-like results should be testable via stock market data for large and small firms. The impact of a clearly defined liberalisation ‘treatment’ should be asymmetric for large and small firms.

It is important to recall that each D-type and each X-type earns pure profits throughout its life time, so these Stolper-Samuelson results are not transitory. They are permanent, firm by firm. Of course, new firms that become active will not experience an increase in their Ricardian surplus, but they will earn a reward that is higher than it would have been without the liberalisation.

²³ Noting that a firm's Ricardian surplus is $1/\sigma$ times its sales, the proportional change in the r 's are identical to the proportional changes in sales, which are listed in (13). Thus, we know $\hat{r}_D < 0 = \hat{w} < \hat{r}_X$. To establish real factor reward changes, we must compare these to the proportional change in P shown in (14). The real changes are:

$$\hat{r}_D - \hat{P} = (\sigma - 2)\hat{a}_D, \quad \hat{r}_X - \hat{P} = (\sigma - 2)\hat{a}_D + \frac{\phi}{1 + \phi}\hat{\phi} \quad \text{and} \quad \hat{w} - \hat{P} = 0 - \hat{a}_D.$$

Since \hat{a}_D is negative when trade freeness rises ($d\phi > 0$), we know that workers always gain, and $\sigma > 2$ is a necessary and sufficient condition for D-type firm owners to lose. For X-type owners note that using (14), $\hat{r}_X - \hat{P} = \frac{\sigma - 2}{\sigma - 2} \left(\frac{\Omega}{1 + \Omega} \right) \hat{\phi} + \frac{\phi}{1 + \phi} \hat{\phi}$. Since $\phi/(1 + \phi)$ exceeds $\Omega/(1 + \Omega)$ – see (13) – a sufficient condition for X-types to gain in real terms is that $\sigma > 3/2$.

An interesting point that might be interesting to pursue in future work stems from the observation that trade liberalisation raises the reward to developing really low cost varieties. In the Hopenhayn-Melitz approach to innovation, developers have absolutely no way of influencing the chances of getting a low marginal cost variety. But if they could, the Stolper-Samuelson result suggests that they would.

Given the correlation between type and firm size in this model, Result 14 and Result 15 should be testable empirically. Simply put, the rise in a firm's stock market price in reaction to a reciprocal trade liberalisation should be positive for firms that are sufficiently large and negative for firms that are sufficiently small.

4.1.3. Fixed cost liberalisation

Inspection of (15) and Result 2, show that the income distribution impact of fixed cost trade barrier liberalisation will be quite different compared to variable trade cost liberalisation, at least for X-type firms. In particular, lowering F_X lowers a_D and this, as per (15), lowers the reward to both X-type and D-type capital. In other words, greater regulatory liberalisation reduces both operating profits by the same proportion. Thus:

Result 16: Regulatory liberalisation, defined as a lowering of F_X/F_D , lowers the reward to both D-type and X-type capital without altering the wage.

The intuition for this result is clear. The beachhead costs create barriers to enter that must, in equilibrium, be compensated for by higher operating profits. Lowering the beachhead costs thus lowers flow reward to active firms.

5. POLITICAL ECONOMY EFFECTS: BIG VS LITTLE FIRMS

One often observes large, exporting firms supporting trade liberalisation, while smaller firms oppose it. Or more precisely, big firms are active supporters of reciprocal trade liberalisations – such as WTO rounds and free trade agreements – but small firms are much less so. Moreover, most firms, large and small, oppose unilateral liberalisation of their final goods market. As it turns out, these are a very natural implication of the HMCFMEC model.

The basic point is simple. All firms in HMCFMEC model would oppose a unilateral reduction in their nation's level of τ , since this would raise the degree of competition in their local market, thereby costing them sales and operating profits. Indeed, according to the usual home market effect, we know that this sort of liberalisation would lead to a fall in the number of home firms and a rise in the number of foreign firms (see Helpman, Melitz and Yeaple 2004). Thus all firms would oppose unilateral liberalisation.

The story, however, is starkly different when it comes to a reciprocal trade liberalisation. Modelling this as simply as possible, we could say that the Home and Foreign τ 's would be lowered in tandem. As we saw in the previous section, such a liberalisation would provide real gains to large, export oriented firms, but real loses to small domestic-oriented firms. See Baldwin and Robert-Nicoud (2005) for a detailed development of this line of thinking.²⁴

²⁴ The idea was discussed informally in Baldwin (2000 p.275); the Brookings Panel draft of the paper contained a formal model that was removed at the request of the editor, Dani Rodrik.

5.1. Contrast with models without HMCFMEC

To illustrate the importance of the two main elements of the HMCFMEC model in driving this small-versus-big firm result, it is useful to consider the implications of the reciprocal liberalisation without heterogeneity in firm-level marginal costs and without fixed beachhead costs.

If all firms – both home and foreign – had the same marginal cost, but faced fixed beachhead cost, no firms would export. It would never be possible for an exporter to earn operating profits that were sufficient to cover the beachhead costs. The reason, of course, is that local entry would proceed until operating profits were just sufficient to cover local entry costs, thus leaving no possibility of earning the higher operating profits necessary to cover F_X . If we drop both heterogeneity and set $F_X=0$, we are in the standard Helpman-Krugman model world. Here, we note that all firms have exactly the same attitude towards liberalisations, both unilateral (which they oppose) and reciprocal. In particular, since total operating profit is invariant to ϕ , all firms would be perfectly indifferent to reciprocal trade liberalisation.²⁵

If we changed the model to allow for variable operating profit margins – as in Ottaviano, Tabuchi and Thisse (2002), or Ottaviano and Melitz (2005) – then reciprocal trade liberalisation will, typically, have a pro-competitive effect. This means all the identical firms lose and so all would oppose reciprocal liberalisation.

Finally, consider a model with firms that have heterogeneous marginal costs but no fixed market entry costs (e.g. Baldwin and Okubo 2004). In such a model, there would be no distinction between D-types and X-types, so again there would be no possibility of a divergence of views between large and small firms. Of course, if one eliminates the fixed beachhead costs, but switches to a linear demand system, then D vs X type distinction re-emerges since some firm's marginal costs will exceed the y-axis intercept of the demand curve in their export market, but not in their domestic market.

Result 17: If a sector is marked by HMCFMEC, then the largest firms in the sector should support reciprocal trade liberalisation while small firms in the same sectors should oppose it. By contrast, all firms in a sector should oppose unilateral liberalisation of their final good market.

6. CONCLUDING REMARKS

Trade models with heterogeneous firms and beachhead costs constitute an important new instrument in the toolbox of international trade theorists. Of course, nothing under the sun is entirely new – many, many trade theorists have published model in which firms have different marginal costs, and the late 1980s saw a flourishing of papers and books on models with beachhead costs. Nevertheless, the assumption of continuous marginal-cost heterogeneity in a monopolistic competition setting teamed with beachhead costs and a Melitz-Hopenhayn selection mechanism constitutes more than just an incremental improvement on existing models. In particular, the two main elements – heterogeneity and beachhead costs – allow trade theorists to address many real-world aspect of trade that have been henceforth largely ignored. While other uses of the model may emerge, the uses to date have focused on two aspects of the model: there are firm-level differences with respect trade behaviour, and all active firms earn pure profits.

This paper sets out a basic heterogeneous-firms trade model that is closely akin to Melitz (2003). The positive and normative properties of the model are studied in a manner intended to highlight

²⁵ In some models, liberalisation would expand total expenditure on the sector and thus raise all profits. However, free entry would again force all existing firms to be indifferent liberalisation.

the core economic logic of the model. The paper also studies the impact of greater openness on both the firm-level and aggregate level, focusing on changes in the number and type of firms, trade volumes and prices, and productivity effects. The normative effects of liberalisation are also studied and here the paper focuses on aggregate gains from trade, and income redistribution effects, showing inter alia that the model is marked by a Stolper-Samuelson like effect.

A number of empirically testable hypotheses are also developed. These concern the impact of greater openness on the bilateral trade pattern, the firm-level trade pattern, the variance of unit-prices, the stock market valuation of firms according to size, and the lobbying behaviour by size.

References

- Aw, B., Chung, S. and Roberts, M., 2000, 'Productivity and Turnover in the Export Market: Micro Evidence from Taiwan and South Korea', *World Bank Economic Review* 14, 65-90.
- Baldwin, R. and F-R Nicoud (2005). "Juggernaut".
- Baldwin, Richard E. (1988). "Hysteresis in Import Prices: The Beachhead Effect," *American Economic Review*, 78, 4, pp 773-785.
- Baldwin, Richard E. (2000). "Regulatory Protectionism, Developing Nations and a Two-Tier World Trading System," *Brookings Trade Forum*, S. Collins and D. Rodrik (eds), The Brookings Institution, Washington DC, pp 237-293, 2000. Re-published in *Quantifying the impact of technical barriers to trade: Can it be done?* K. Maskus and J.S. Wilson (eds), Michigan University Press, 2001.
- Baldwin, Richard E. and Paul Krugman (1989). "Persistent Trade Effects of Large Exchange Rate Shocks," *Quarterly Journal of Economics*, pp 635-654, No. 419.
- Baldwin, Richard E. and Rikard Forslid (2004). "Trade liberalisation with heterogenous firms," CEPR Discussion Paper 4635.
- Bernard A. and Jensen B, 1995, 'Exporters, Jobs and Wages in U.S. Manufacturing, 1976-1987' *Brookings Papers on Economic Activity, Microeconomics*, 67-119
- Bernard A. and Jensen B., 2001, 'Why Some Firms Export', NBER working paper 8349
- Bernard A. and Jensen, B., 1999b, 'Exporting and Productivity: Importance of Reallocation', NBER working paper 7135.
- Bernard, A. and Jensen, B., 1999a, 'Exceptional Exporter Performance: Cause, Effect, or Both?' *Journal of International Economics* 47, 1-26
- Bernard, A.B., J. Eaton, J.B. Jensen and S. Kortum (2003) "Plants and Productivity in International Trade", *American Economic Review*, Vol. 93, No. 4, September, 1268-1290.
- Bernard, Andrew B., Stephen Redding, Peter K. Schott (2004), "Comparative Advantage and Heterogeneous Firms," NBER Working Paper No. 10668.
- Bernard, Andrew, J. Bradford Jensen, and Samuel Kortum (2003). "Plants and productivity in international trade", *American Economic Review*, 93, 4, 1268-1290.
- Bernard, Andrew, Jonathan Eaton, J. Bradford Jensen, and Peter Schott (2003). "Falling Trade Costs, Heterogeneous Firms and Industry Dynamics", mimeo, Tuck Business School.
- Cabral, L. M. B. and J. Mata (2001) "On the Evolution of the Firm Size Distribution: Facts and Theory", CEPR Discussion Paper 3045. Forthcoming in *American Economic Review*.

- Clerides S., Lach, S. and Tybout, J., 1998, 'Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco', *Quarterly Journal of Economics*, 133, 903- 904
- Dinopoulos, E. and Segerstrom P, (1999b), "A Schumpeterian Model of Protection and Relative Wages", *American Economic Review*, 89, 450-473.
- Dixit, Avinash (1988). "Entry and exit decisions under uncertainty," *Journal of Political Economy*, 97, 4, pp 620-638.
- Eaton, Jonathan and Samuel Kortum (2002). "Technology, Geography and Trade," *Econometrica*, 70, 5, pp 1741-1780.
- Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2004), "Dissecting Trade: Firms, Industries, and Export Destinations", *American Economic Review Papers and Proceedings*, 94, 2004, 150-154.
- Feenstra, R. and A. Rose, 1997, "Putting Things in Order: Patterns of Trade Dynamics and Macroeconomics", NBER working paper 5975.
- Griliches, Z. (1963) "Capital Stock in Investment Functions: Some Problems of Concept and Measurement." In Christ, C., et al. (eds.) *Measurement in Econometrics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, 115-37. Stanford: Stanford University Press.
- Grossman, G.M. and Helpman, E., 1989. "Trade; Innovation; and Growth," *Papers 154*, Princeton, Woodrow Wilson School - Public and International Affairs.
- Grossman, Gene and Elhanan Helpman. (1991). *Innovation and Growth in the World Economy*. Cambridge, MA: MIT Press.
- Grubel, Herbert G., P.J. Lloyd (1975) *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products* (London, Macmillan).
- Helpman, Elhanan and Paul Krugman (1985), *Market structure and Trade*, MIT Press.
- Helpman, Elhanan, Marc Melitz and Stephen Yeaple (2004). "Export versus FDI with heterogeneous firms," *American Economic Review*, 94, 1, pp. 300-317.
- Helpman, Elhanan, Marc Melitz and Yona Rubinstein, 2004, *Trading Partners and Trading Volumes*, preliminary and incomplete pdf file, August 24, 2004.
- Hopenhayn, Hugo (1992a). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica* 60:1127-1150.
- Hopenhayn, Hugo (1992b). Exit, Selection, and the Value of Firms, *Journal of Economic Dynamics and Control* 16:621-653.
- Keller, W. (2004). "International technology diffusion," forthcoming *Journal of Economic Literature*.
- Krugman, P. (1980). "'". *American Economic Review*.
- Melitz, M. (2003). "The impact of trade on intraindustry reallocations and aggregate industry productivity," *Econometrica*, 71, pp 1695-1725.
- Melitz, Marc and Gianmarco Ottaviano (2003), "Market Size, Trade and Productivity", mimeo.
- Ottaviano G., Tabuchi T., and Thisse JF, "Agglomeration and Trade Revisited" *International Economic Review*, May 2002, vol. 43, no. 2, pp. 409-436(28).
- Rivera-Batiz, L. and P. Romer, 1991a, Economic integration and endogenous growth, *Quarterly Journal of Economics* 106, 531-555.

- Rivera-Batiz, L. and P. Romer, 1991b, "International trade with endogenous technological change", *European Economic Review* 35, 715-721.
- Romer, P., 1986, "Increasing returns and long-run growth", *Journal of Political Economy* 94, 1002-1037.
- Romer, Paul (1990). "Endogenous Technological Change," *Journal of Political Economy*, 98, S71-S102.
- Tybout, J., 2003, "Plant- and Firm-level Evidence on 'New' Trade Theories", in E. K. Choi and J. Harrigan (eds) *Handbook of International Trade* Blackwell, Oxford.
- Wang, Z. K. and Alan Winters, 1992, "The Trading Potential of Eastern Europe," *Journal of Economic Integration*, Vol. 7, 1992, pp. 113-136.
- Yeaple, S. (2005), "Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, 65(1), pp 1-20.