Agglomeration, Offshoring and Heterogeneous Firms

Richard E. Baldwin and Toshihiro Okubo
Graduate Institute of International Studies, Geneva
April 2006; First draft: September 2005

ABSTRACT

Recent trade models determine the equilibrium distribution of firm-level efficiency endogenously and show that freer trade shifts the distribution towards higher average productivity due to entry and exit of firms. These models ignore the possibility that freer trade also alters the firm-size distribution via international firm migration (offshoring); firms must, by assumption, produce in their ‘birth nation.’ We show that when firms are allowed to switch locations, new productivity effects arise. Freer trade induces the most efficient small-nation firms to move to the large nation. The big country gets an ‘extra helping’ of the most efficient firms while the small nation’s firm-size distribution is truncated on both ends. This reinforces the big-nation productivity gain while reducing or even reversing the small-nation productivity gain. The small nation is nevertheless better off allowing firm migration.

JEL H32, P16.

Keywords: heterogeneous firms, economic geography, productivity effects, offshoring, delocation, hollowing-out.

1. INTRODUCTION

Recent years have seen a flourishing of heterogeneous-firms trade models including Eaton and Kortum (2002), 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), Falvey, Greenaway and Yu (2004), Yeaple (2005), and Demidova (2005). The key innovation in this literature is that it endogenizes the equilibrium distribution of firm-level efficiency and thereby the equilibrium firm-size distribution in an open economy. One result that has attracted much attention is the way in which freer trade has pro-productivity effects that stem from changes in the equilibrium distribution of firm-level efficiencies.

Helpman, Melitz and Yeaple (2004) allow for multinational production, i.e. production by a single firm in both home and foreign whose purpose is to avoid trade costs, but the literature to date ignores the issue of international firm migration – what is known as offshoring in North America, delocation in Europe and ‘hollowing out’ in Japan – i.e. where a firm ceases production in one nation and sets up production abroad. Since firm migration – the shifting of production location by

* We thank the Swiss National Science Foundation for support (Grant No. 100012-105675/1). An early draft was presented at the Vth Spring School in Economic Geography, Cagliari, Italy 24th May 2005 and at the annual conference of the European Trade Study Group, Dublin 10th September 2005 (see http://www.etsg.org); we thank the participants for useful and insightful comments especially Thierry Mayer, Pierre-Philippe Combes, Holger Breinlich and Peter Neary. We also thank Jota Ishikawa, Taiji Furasawa and other participants at the 2005 presentation of this paper at Hitotsubashi University.
individual firms – is an important empirical phenomenon, it is striking that heterogeneous-firms trade (HFT) models ignores this spatial relocation.

Our paper argues that the assumption of no spatial relocation in the HFT models is not innocent. We work in a framework where nations differ only in size.1 In this world, freer trade has two distinct effects. First, it has the well-known pro-productivity impact; freer trade truncates the firm-size distribution – driving out the smallest least efficient firms with the effect being symmetric in the two nations (see Helpman, Melitz and Yeaple 2003, for example). Second, it creates “home market effect” pressures that foster the shifting of production to the larger market. In the standard HFT model, all the home-market-effect pressure is alleviate via the entry and exit of firms. We argue that since entry/exit is not instantaneous, some of the home-market-effect pressure would be alleviated via delocation of the most efficient small-nation firms, if the possibility of spatial reallocation were not ruled out by assumption. Such firm-migration would have interesting implications for the firm-size distribution in both nations. The big country would get an ‘extra helping’ of the most efficient firms and the small nation’s firm-size distribution would be truncated on both ends. The most efficient firms would delocate and the least efficient firms would exit.

In this way, delocation adds a new dimension to the productivity impact of freer trade. The big nation gets an extra large productivity gain while the small nation’s productivity gain is mitigated or even reversed.

To demonstrate these interactions as cleanly as possible, our paper combines a simple New Economic Geography (NEG) model – namely, the Footloose Capital model of Martin and Rogers (1995) – with a simple heterogeneous-firms trade model. Since NEG and HFT models are both on the edge of analytic tractability, it is not surprising that the combined model is impossible to solve analytically when we allow entry/exit and delocation simultaneously. To explore the combined model analytically, we consider two polar cases: One where entry and exit of firms is instantaneous, but relocation is slow, and one where delocation is instantaneous, but entry/exit is slow.

1.1. Literature review

Our paper has antecedents from two strands of literature, the heterogeneous firms trade literature and the ‘new economic geography’ literature. We address these in turn.

One recent branch of trade theory has focused on differences among firms. The main theoretical papers in this rapidly expanding literature are Eaton and Kortum (2002), 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), Falvey, Greenaway and Yu (2004), and Yeaple (2005). For brevity’s sake, we refer to these as the heterogeneous-firms trade (HFT) models.

The HFT models were motivated by empirical evidence. 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).

The other strand of literature that is relevant to our work is the so-called New Economic Geography literature, of which Fujita, Krugman and Venables (1999) and Fujita and Thisse (2002) are the standard references. In this literature, increasing returns to scale and trade costs create forces that foster spatial agglomeration. Since well-known the core-periphery model was introduced by Krugman (1991), various models have been provided by the literature. The simplest of these models

---

1 Thus we are thinking of delocation from, e.g. Canada to the US, or Sweden to Germany rather than the US to China.
is the so-called Footloose Capital (FC) model proposed by Martin and Rogers (1995). The FC model features two, potentially asymmetric region and two-factor of production of which only capital is mobile between regions. Capital is owned by labourers who are inter-regionally immobile, so all capital earnings are repatriated. As a consequence, the FC model has no demand linkages.

The paper that is closest to ours in combining heterogeneous firms and NEG considerations is our earlier paper, Baldwin and Okubo (2006). That paper ignores two key elements that are crucial to the heterogeneous-firms literature, namely entry and exit of firms, and fixed market entry costs. Our main findings in Baldwin and Okubo (2006) are that the most efficient firms are likely to first relocate to the bigger market. This suggests that standard econometric tests of agglomeration economies are biased. Okubo (2005, 2006) and Okubo and Rebeyrol (2006) are other papers in this line.2

1.2. Plan of paper

The rest of the paper is organised into four sections. The next section introduces the model and solves it for the closed economy case. The following two sections works out the trading equilibrium and the welfare impact on the small nation. The final section presents our concluding remarks.

2. THE MODEL

The model works with two nations, two sectors and two factors of production. The nations – the North and the South – have identical tastes, technology, and openness to trade, but they differ in terms of size (North is larger by convention). The two sectors are M (manufactures) and a numeraire sector. The numeraire sector produces a homogenous good subject to constant returns, perfect competition and costless trade. The M-sector produces a continuum of varieties under conditions of increasing returns and Dixit-Stiglitz monopolistic competition. The factors are capital and labour.

Tastes of each consumer in either nation are quasi-linear:

\[ U = \mu \ln C_M + C_A, \quad C_M = \left( \int c_i^{1-1/\sigma} di \right)^{1/(1-1/\sigma)}, \quad 0 < \mu < 1 < \sigma \]

where \( C_M \) and \( C_A \) are, respectively, consumption of the composite of M-sector varieties and the numeraire, \( \sigma \) is the constant elasticity of substitution between any two M-sector varieties and \( \mu \) measures demand for manufactures. The integral is over the set of varieties available for consumption in a particular nation. All varieties are symmetric on the demand side.

Our M-sector firms face constant marginal production costs and a fixed start-up cost. We assume that the marginal cost involves only labour while the start-up cost involves only capital (the amount of start-up capital is normalised to one unit per variety). We also deviate from the standard Dixit-Stiglitz model by assuming that firms are heterogeneous in terms of their manufacturing efficiency (marginal cost). Following Melitz (2003), we work with a simplified version of the Hopenhayn (1992a, 1992b) mechanism of firm development whereby each M-sector firm is associated with a particular labour input coefficient – denoted as \( a_j \) for firm \( j \). Since each firm is associated with a unique unit of capital, it is natural to view the firms’ heterogeneity in marginal cost as a feature associated with its unit capital. Capital can be viewed either as physical capital or knowledge capital (i.e. some form of firm-specific knowledge).

2 Okubo (2005) studies the impact of taxation and subsidy on spatial sorting. Okubo (2006) studies the impact of various sorts of anti-agglomeration subsidies on firm location and welfare, and Okubo and Rebeyrol (2006) examines the impact of market-specific sunk costs on the home market effect, when firms are heterogeneous.
Since there is one unit of capital per firm, international migration by firms (i.e. delocation) is synonymous with capital mobility in our model. All capital is owned by labourers who are immobile between nations. When capital is internationally mobile, we assume all capital income is repatriated costlessly to its immobile owner. Because of this, capital/firms move in search of the highest rate of return irrespective of national cost-of-living considerations.

The marginal costs, the ‘a’s, are determined during the capital creation process. A potential firm pays a start-up cost of $F_1$ units of labour to create a unit of capital. Just after sinking this cost, the capital is randomly assigned an ‘a’ from a distribution function, $G[a]$, which has positive probability for $a \in [0,a_0]$. It may be useful to think of this as a stochastic production function in the capital-creation process. Following Melitz (2003), we ignore intertemporal discounting but keep present values finite by assuming that capital depreciates in a particular way. Firms face a constant probability of ‘death’ according to a Poisson process with the hazard rate $\delta$.

M-sector firms face per-unit and per-market ‘selling’ costs on top of their production costs. The per-unit costs are standard – they are zero in the local market but involve ‘iceberg’ trade costs in the export market such that a firm must ship $1+t\geq 1$ units to sell one unit in the export market. The per-market costs – what we call ‘beachhead costs’ – reflect the fixed cost of establishing a ‘beachhead’ for each new variety in each market. We think of this as reflecting the cost of meeting market-specific standards and regulations, and establishing a brand name. The beachhead cost per market is denoted as $F$, which is sunk involving only labour inputs.

2.1. Intermediate results

Results for the numeraire sector in this sort of model are well known. Constant returns, perfect competition and zero trade costs equalise nominal wage rates across nations. We choose units of the numeraire good such that the equalised wage rate is unity. This means that all differences in M-firms’ marginal costs are due to differences in their a’s so we can refer to the a’s as marginal cost without ambiguity.

The M-sector is marked by all the usual Dixit-Stiglitz results. Firms’ prices are a constant mark-up of their marginal selling costs. In the local market, these marginal costs entail only production costs. The price in the export market will include the iceberg costs, $t\geq 1$, marked up by the constant Dixit-Stiglitz mark-up. Given the quasi-linear preferences, expenditure on manufactured goods is invariant to trade costs and firm location (assuming only that the region’s endowments are such that some of both goods are produced in both nations). These results allow us to normalise worldwide expenditure on manufactured goods to unity without loss of generality. Specifically, we choose units of labour such that $L_w=1/\mu$, so $E_w=1$.

The standard CES demand function for variety-$j$ produced and sold in the North can be written as

$$c_j = \frac{(p_j)^{\sigma}}{\bar{p}}E; \quad \bar{p} = \int_{i \in \Theta} p_j^{1-\sigma} di + \int_{h \in \Theta^*} \phi p_h^{1-\sigma} dh, \quad \sigma > 1 \geq \phi \equiv t^{1-\sigma} \geq 0$$

where $p_j$ is variety-$j$’s producer price (which equals its consumer price since it is produced locally), and $\bar{p}$ is the (weighted) average of consumer prices of varieties sold in the North. The first term in the definition of $\bar{p}$ reflects the prices of goods that are produced in the North (and so bear no iceberg trade costs). The second term reflects the imported varieties whose producer prices are $p_h$; $\Theta$ and $\Theta^*$ are sets of consumed goods that are produced in the North and the South, respectively.

---

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

4 Many HFT models allow separate beachhead costs for the local ($F_1$) and export market ($F_2$), but this plays no role in our model so we assume that are identical to reduce notational clutter in the expressions ($F_1=F_2=F$).
Note that the geometric weights are negative due to the regularity condition that \( \sigma > 1 \), and thus the weighted average price, \( \bar{p} \), falls as individual prices rise. A parameter that plays a critical role in our paper is \( \phi \); we refer to it as the ‘free-ness’ (phi-ness) of trade, and note that \( \phi \) ranges from zero when trade is perfectly un-free (\( t = \infty \)) to unity when trade is perfectly free (\( t = 0 \)). Southern demand functions are isomorphic.

2.1.1 Solution difficulties

The model is marked by a complex matrix of interactions involving firm heterogeneity, export behaviour, free entry-exit and international relocation. The spatial allocation of firms poses the greatest difficulties. The location choice depends upon the degree of competition in each market but this depends in turn upon the mass and efficiency of the firms that choose to locate and enter in each market. The crux of the difficulty lies in defining \( \bar{p} \) and \( \bar{p}^* \) when the production location of varieties is not determined a priori. To determine the equilibrium location of firms, we need to know the degree of competition in each market (\( \bar{p} \) and \( \bar{p}^* \)), but this in turn depends upon the distribution of the marginal costs in each market. In general, firms with any levels of marginal costs could locate in either nation, so it is not obvious how one would apply the density function for the \( a \)’s in forming the integrals behind the \( \bar{p} \)’s.

The literature has leaned on two extreme assumptions to get past this difficulty. The traditional NEG literature ignores firm heterogeneity so the spatial distribution of firms boils down to the number of identical firms producing in each market (i.e. \( G \) is degenerate so the integrals in \( \bar{p} \) and \( \bar{p}^* \) are trivial). The standard HFT model rules out relocation by fiat so the distribution of active firms in each nation is simple – firms must, by assumption, produce in the nation they are ‘born’ so the distribution of active firms in a nation is a simple transformation of the underlying \( G[a] \).

Determining the equilibrium when firms are heterogeneous and internationally mobile is the central technical difficulty of our model. To tackle this problem, we start with the closed-economy case.

2.2. Equilibrium under autarky

We start with autarky, defined here as zero mobility of goods (due to zero free-ness of goods i.e. \( \phi = 0 \)) and firms. As in Melitz (2003), firms are assumed, by fiat, to produce where they are born. The first task is to work out the steady state distribution of active firms.

Firms are active if being so is profitable. Given (2), the Poisson death process and well-known Dixit-Stiglitz mark-up, present value of being an active firm with a marginal cost of \( a_j \) is:

\[
\frac{p_j}{\sigma \delta \bar{p}} E \quad \text{where} \quad p_j = \frac{a_j}{1 - 1/\sigma}, \quad \bar{p} \equiv \int_{\Theta} p_i^{1-\sigma} \, di
\]

Here \( \bar{p} \) is the weighted average of all competing varieties (the weights are negative so higher prices imply a lower \( \bar{p} \)), \( \Theta \) is the sets of varieties sold. Since this present value increases as a firm’s marginal cost falls, it is easy to see that only varieties with sufficiently low \( a \)’s will find it worth their while to sink the ‘beachhead cost’ \( F \). Defining this threshold marginal cost as \( a_D \), the cut-off condition is:

\[
(3) \quad a_D^{1-\sigma} B = f \quad \text{where} \quad B = \frac{E}{\bar{m}}, \quad \bar{m} = \int_0^{a_D} a^{1-\sigma} dG[a|a_D], \quad f = \sigma \delta F
\]

where \( \bar{m} \) is the weighted average of marginal costs of active firms (the geometric weights are negative), ‘n’ is the mass of firms, and \( G[a|a_D] \) is the conditional distribution function of ‘a’
(conditional on \(a \leq a_0\)); we cancel the mark-ups and define ‘f’ and ‘B’ for notational convenience; B is the per firm demand.

Note that we have switched from integrating over varieties to integrating over a’s. As Melitz (2003) shows, the distribution of a’s comes from the features of the firm birth and death processes. In steady state, a continuous flow of new varieties is generated to replace the continual flow of ‘dying’ varieties. Given variety-generation process assumed, the a’s of new varieties are distributed according to \(\delta nG[a]\), \(0 \leq a \leq 1\), where ‘n’ is the mass of firms. In other words the distribution of new varieties is a simple transformation of the underlying G. In particular, the beachhead cost implies that not all new varieties are produced, so the distribution of varieties that are actually produced is a truncated version of \(\delta nG[a]\), namely \(\delta nG[a|\alpha_D]\), where \(G[a|\alpha_D]\) is the distribution function conditional on ‘a’ being less than \(\alpha_D\).

The cut-off condition defines the optimal sales strategy for a firm with a given ‘a’ (they sell only if the marginal cost is below \(\alpha_D\)), thus the equilibrium present value of operating profits for a variety with \(a_j\) is:

\[
a_j^{1-\sigma} \frac{B}{\sigma} \text{ for } 0 \leq a \leq \alpha_D; \quad \text{zero for } \alpha_D \leq a;
\]

In the first stage – the innovation stage where \(F_I\) is sunk – potential firms form expectations as to their likely future profits by considering these two present values and the likelihood that they will draw an ‘a’ below \(\alpha_D\). Specifically, the integral of these present values over all possible a’s, weighted by the probability of each ‘a’ (i.e. the unconditional density function) provides the expected benefit of sinking \(F_I\). Free entry ensures that the mass of firms ‘n’ rises to the point where the expected benefit of innovation is driven to zero and this gives us the free-entry condition:

\[
\int_0^{\alpha_D} (a^{1-\sigma} B - f) dG[a] = f_I \quad \text{where } f_I = \sigma \delta F_I
\]

These two conditions, namely the cut-off condition (3) and the free entry condition (4), characterise the two equilibrating variables \(\alpha_D\) and \(n\).

To get an analytic solution, we assume \(G\) is a Pareto distribution:

\[
G[a] = \left(\frac{a^\rho}{a_0^\rho}\right), \quad 0 \leq a \leq a_0 \equiv 1, \quad \rho \geq 1
\]

where the parameters \(\rho\) and \(a_0\) are Pareto’s shape and scale parameters; without loss of generality, we choose units such that the maximum marginal cost \(a_0\) equals 1. Using this, we solve (4) and (3) to get:

\[
a_D = \left(\frac{f_I (\beta - 1)}{f} \right)^{1/\rho}, \quad n = \frac{E(\beta - 1)}{f \beta} \quad \text{where } \beta = \frac{\rho}{\sigma - 1}
\]

Here we introduce the collection of parameters, \(\beta\), for convenience and note that \(\beta > 1\) is a regularity condition, assumed throughout the paper (it ensures the integrals converge). The South’s equilibrium conditions are isomorphic, namely:

\[\text{5} \quad \text{Given the quasi-linear preferences, } E \text{ and } E^* \text{ can be taken as parameters since they equal } \mu L \text{ and } \mu L^*, \text{ where } L \text{ and } L^* \text{ are the number of Northern and Southern residents respectively (one unit of labour per resident).}\]
\[(7) \quad a_D = \left( \frac{f_i(\beta - 1)}{f} \right)^{1/\rho}, \quad n^* = \frac{E^*(\beta - 1)}{f \beta} \]

where \(n^*\) is the mass of South firms with any given ‘a’s.

3. EQUILIBRIUM WITH TRADE

Starting from the closed-economy steady state, consider the impact of lowering the iceberg trade cost. As trade opens up, pressures for the well-known Home Market Effect (HME) appear. That is, there is a tendency for deeper integration to make the large market more attractive to firms in the increasing-returns sector while simultaneously making the small market less attractive. There are two ways for the equilibrium to adjust to the HME pressure:

- Firms can physically migrate, moving their units of capital from the small South to the big North;
- Firm births can exceed firm deaths in the North, while the opposite occurs in the South.

The fact that delocation and/or entry/exit can occur makes characterisation of the equilibrium difficult. We know that the distribution of firm-births and firm-deaths in equilibrium must be \(\delta n G[a]\) and \(\delta n^* G[a]\), but we do not know, a priori, where these varieties will be produced. The crux of the difficulty is that determining production location requires knowledge of the degree of local competition, namely the \(\mathcal{m}\)'s, but defining these requires knowledge of where each variety is produced.

Figure 1: NEG and HFT solution to calculation difficulties.

To tackle this problem, it is useful to work out the equilibrium assuming that entry/exit and delocation occur in sequence, as illustrated by Figure 1.

- First we assume entry/exit is instantaneous, and then we allow delocation (we move from the origin in the figure to the second circle by going over and then up); this is the approach implicit in the standard HFT literature, e.g. Melitz (2003).
- Second, we assume delocation is instantaneous and then allow free entry/exit (we move from the origin in the figure to the second circle by going up and then over); roughly speaking, this is what is assumed in NEG models.
We consider these alternative adjustment paths in isolation before considering the general case.

### 3.1. Instantaneous entry/exit followed by delocation

As we shall see, when entry/exit is instantaneous but delocation is slow, all adjustment to the HME pressure occurs via entry/exit. The solution therefore is very similar to that of the well-known Melitz (2003) equilibrium. The cut-off conditions for local and export sales are:

\[ a_D^{1-\sigma} B = f, \quad \phi a_X^{1-\sigma} B = f \]

where the two beachhead costs are assumed to be identical for simplicity’s sake, and \( \bar{m} \) and \( \bar{m}^* \) (weighted average marginal selling costs in the Northern and Southern markets) are:

\[
\bar{m} = n \int_0^{a_d} a^{1-\sigma} dG[a|a_D] + n^* \int_0^{a_d} a^{1-\sigma} dG[a|a_D], \quad \bar{m}^* = \phi n \int_0^{a_d} a^{1-\sigma} dG[a|a_D] + n^* \int_0^{a_d} a^{1-\sigma} dG[a|a_D]
\]

Here \( n \) and \( n^* \) are the masses of firms producing in the North and the South.

The North and South free entry conditions are:

\[
\int_0^{a_d} \left( a^{1-\sigma} B - f \right) dG[a] + \phi \int_0^{a_d} \left( a_X^{1-\sigma} B^* - f \right) dG[a] = f_I
\]

\[
\int_0^{a_d} \left( a_X^{1-\sigma} B - f \right) dG[a] + \phi \int_0^{a_d} \left( a_X^{1-\sigma} B^* - f \right) dG[a] = f_I
\]

Since we have normalised the size of the world economy such that world expenditure on M-varieties is unity (i.e. \( \mu (L + L^*) \) equals 1), the four equilibrium conditions in (8) and (9) can be solved analytically to yield:

\[
a_D = \left( \frac{f_I (\beta - 1)}{f (1 + \phi^\beta)} \right)^{1/\rho}, \quad a_X = \left( \frac{f_I \phi^\beta (\beta - 1)}{f (1 + \phi^\beta)} \right)^{1/\rho}, \quad n^w = \frac{1 - 1/\beta}{f (1 + \phi^\beta)}, \quad s_n = \frac{1}{2} + \frac{1}{1 - \phi^\beta} (s_E - \frac{1}{2})
\]

where

\[
s_E \equiv \frac{E}{E + E^*}, \quad s_n \equiv \frac{n}{n + n^*}, \quad n^w \equiv n + n^*
\]

Here we introduce the standard NEG ‘share notation’ where \( s_n \) and \( s_E \) are the North’s share of the worldwide mass of firms \( n + n^* \) and worldwide expenditure on all M-sector varieties \( E + E^* \) (\( s_E \) is exogenous in our model and \( s_n \), which characterises the spatial allocation of industry, is one of the key endogenous variables).

Note that an implication of the equilibrium conditions is that \( B = B^* \), i.e. the per-firm demand is equalised despite the market size differences. Intuitively, this result is quite obvious. Free entry means that the number of competitors in each market rises to the point where further entry is

---

6 In principle, there are four cut-off conditions defining the threshold marginal costs for D-type and X-type firms in the North and the South, but as is well-known from the HFT literature, free entry implies that the thresholds are identical despite size difference between the two nations; see Helpman, Melitz and Yeaple (2004) and Falvey, Greenaway and Yu (2004).

7 The first integral in the expression for \( \bar{m} \) captures the marginal costs of locally produced goods, i.e. goods that have marginal costs between zero and \( a_D \). The term \( dG[a|a_D] \) reflects the mass of firms with marginal cost ‘\( a \)’ conditional on the fact that the firm is producing at all, i.e. that it’s ‘\( a \)’ is less than \( a_D \). The limits of the second integral are from zero to \( a_X \) which is the cut-off marginal cost for Southern firms exporting to the North. The expression for \( \bar{m}^* \) is isomorphic, but \( \phi \) applies to sales in the Northern market. Recall that the geometric weights are negative, so \( \bar{m} \) and \( \bar{m}^* \) fall as prices rise.
uneconomic. In other words, the number of firms per market must adjust to the point where each market is equally attractive to firms, in particular to the point where potential firms are just indifferent to sinking $F_i$.

The HME is clear from the fourth expression in (10), the expression of the equilibrium $s_n$. By inspection, we see that as trade gets freer ($d\phi>0$), the share of firms in the North rises above the North’s relative size (i.e. $s_n>s_E$). Compared with the homogeneous firms model, where the HME derivative is $d s_n/ds_E = (1+\phi)/(1-\phi)$, the HME derivation in our model is smaller, namely $(1+\phi\beta)/(1-\phi\beta)$, because of $\beta>1$. This implies that firm heterogeneity dampens agglomeration in terms of firm shares.

3.1.1 Allowing delocation

When firms are allowed to change locations (i.e. delocate), each existing firm compares the operating profit it would earn in the two markets. For X-type firms, which are already selling in both markets, the question boils down to the minimization of trade costs. Its operating profit when located in the North is $a^1-\sigma(B+\phi B^*)/\sigma$; when located in the South it is $a^1-\sigma(\phi B+B^*)/\sigma$. Since $B=B^*$ in equilibrium (as explained above), these are identical so all X-types are indifferent to relocation. Likewise, no D-type could gain by relocating. Indeed, since relocation would involve re-sinking $F$, no D-type would move. Likewise, every newly born firm is completely indifferent to its location. To summarise:

**Result 1:** Instantaneous entry/exit relieves all pressure to delocate; instantaneous entry/exit is a perfect substitute for delocation, so the instant-entry-then-delocation case is identical to the standard HFT model.

In the next section, Section 0, we consider the diametrically opposed case, where we move from autarky to trade assuming that delocation is instantaneous and then free entry occurs. However first we characterise the trade and productivity effects of freer trade. These results are familiar from the standard HFT literature, but we quickly review them to fix ideas and introduce notation and analytical techniques that are useful for the instant-delocation-then-entry case in Section 0.

3.1.2 Trade, production and productivity effects

The effects of greater openness are well known. The rise in $aX$ means more firms in both nations export, so exports rise and this, in turn, heightens competition in both markets, so the sales of D-type firms fall in both markets and the marginal D-types (those with $a$’s near $aD$) are driven out. Both the selection and the share-shifting raise average productivity in both nations. Due to the HME, $n$ rises and $n^*$ falls; this increases the degree of inter-industry trade as resources are shifted out of the M-sector in the small nation and into the numeraire sector.

Formally, productivity can be measured by the firm-level productivity weighted by firm’s production share. For the North, this productivity measure – i.e. the production-share-weighted average of the a’s – is:

\[
\bar{a} = \int_0^a \left( a^1-\sigma B(1+\phi) \right) ndG[a|a_D] + \int_0^a \left( a^1-\sigma B \right) ndG[a|a_D];
\]

where $\bar{a}$ is referred to as the weighted average unit input coefficient; this is a measure of average productivity since the geometric weights are negative, but a drop in the input-coefficient enhances productivity. North’s total production, $TP$ is given by:

---

\(^8\) See Martin and Rogers (1995) for instance.
\[ TP = \int_{a_X}^{a_0} a^{1-\sigma} B(1+\phi) n \, dG[a_D] + \int_{a_X}^{a_0} a^{1-\sigma} B \, n \, dG[a_D]. \]

Note that the ratios in large parentheses are production share of firms. The first integral shows the a’s for X-type firms, while the second integration shows the a’s for D-type firms. A drop in the average ‘a’ is an improvement in productivity.

As we shall see below, the productivity effects are substantially more complex when delocation occurs. To set the stage for the more complex analysis, we replicate Melitz’s well-know productivity results using a flexible analytic technique. The basic idea is to view \( \bar{a} \) as the weighted average of the invariant underlying distribution \( G[a] \) and see how freer trade shifts the weights.

Using (5) to solve the TS integrals and the equilibrium expression for \( a_D \) in (10), we can write:

\[ \bar{a} = \int_{a_X}^{a_0} W_X(a^{1-\sigma+\rho}) da + \int_{a_X}^{a_0} W_D(a^{1-\sigma+\rho}) da \]

where the weights on X-type firms, \( W_X \), and the weights on D-type firms \( W_D \) are:

\[ W_X = \frac{(1+\phi)}{f_j(\beta-1)/\overline{f}^{1-1/\beta}} \left( \frac{\rho(1-1/\beta)}{(f_j(\beta-1)/\overline{f})^{1-1/\beta}} \right) \]

\[ W_D = \frac{1}{(1+\phi)^{1/\beta}} \left( \frac{\rho(1-1/\beta)}{(f_j(\beta-1)/\overline{f})^{1-1/\beta}} \right) \]

Plainly, the weight on the a’s of X-type firm’s, \( W_X \), rises as trade gets freer (d\( \phi > 0 \)), and the weight on the a’s of D-types, \( W_D \), declines in \( \phi \).

Using the standard formula for the derivative of a definite integral:

\[ \frac{d\bar{a}}{d\phi} = \int_{a_X}^{a_0} \frac{dW_X}{d\phi} a^{1-\sigma+\rho} da + \int_{a_X}^{a_0} \frac{dW_D}{d\phi} a^{1-\sigma+\rho} da + \frac{d a_X}{d\phi} \left( W_X a_X^{1-\sigma+\rho} - W_D a_D^{1-\sigma+\rho} \right) + \frac{d a_D}{d\phi} W_D a_D^{1-\sigma+\rho} \]

When \( d\bar{a}/d\phi \) is written in this way, we see that the productivity impact of freer trade depends upon the impact of d\( \phi \) on the ‘weighting’ terms – the dW/d\( \phi \)’s – under the integrals – and the impact on the limits of integration – i.e. the last two terms. We know that the derivative (12) is negative (i.e. there is a pro-productivity effect) since freer trade uniformly shifts weight from high values of \( a^{1-\sigma+\rho} \) to low values. Figure 2 shows this graphically.

The term that is being weighted, namely \( G[a] \equiv a^{1-\sigma+\rho} \), is shown with the rising solid line and we note that it is invariant to the freeness of trade. The weighting function before the increase in trade freeness is shown with the solid horizontal line, “Weight (pre)”. After the opening, the weighting function shifts to “Weight (post)”, so the weight on a’s below \( a_X \) rises and the weight on a’s above \( a_X \) falls. The impact of the change in the limits of integration is fourfold and corresponds to the fourth and third terms in (12). The first and second terms in (12) represent the switch of the weighting lines holding \( a_D \) and \( a_X \) at the initial level. The third term shows the impact of shift in weight for the range of a’s between the pre and post export cut-off. The final terms shows the shift in weight (from positive to zero) for the range of a’s between the pre and post D-type cut-off.

Plainly the post-integration weighting curve unambiguously shifts mass to lower values of a, so we know that \( \bar{a} \) falls, i.e. manufacturing productivity rises since the average unit labour input coefficient falls. Freer trade is pro-productivity in the South for the same reasons.

---

9 Please see guide to calculations.
10 Note \( d\ln\{W_X\}/d\phi \) is \( (1-\phi^{\beta-1})/\{(1+\phi)(1+\phi^{\beta})\} \); this is positive since \( \beta > 1 \) (regularity condition).
In a sense, freer trade “pivots” the density clockwise around $a_X'$, lowering the weight on a’s above $a_X'$ and raising it on a’s below $a_X'$. This unambiguously lowers the average marginal cost since $a^{1-\sigma+\rho}$ is upward sloped.

3.2. **Instantaneous delocation followed by entry/exit**

The assumption of instantaneous birth-and-death of firms is clearly unrealistic. But if entry/exit takes time, then some firms may adjust to the HME pressure by moving to the big market. In other words, some of the HME adjustment may occur via delocation of firms from the small market to the large market (as in the New Economic Geography literature) rather than via entry/exit as in the HFT literature.

To explore this possibility, this section considers the other polar assumption that delocation is instantaneous and then considers possibility of entry/exit.

**The initial conditions.** As before we start from the autarky equilibrium. With entry/exit ruled out by fiat initially, the two free entry conditions and the domestic cut-off condition are suspended, so all active firms that were active in autarky remain active initially when trade and firm-migration become possible.

Using our share notation, the initial $n$ and $n^*$ and the initial distribution of active firms $a \in [0, a_D]$ from the autarkic equilibrium (i.e. $\phi=0$) characterised in Section 2.2. Combining (6) and (7):

\[
\begin{align*}
n^w &= \frac{\beta - 1}{f \beta}, \\
a_D &= \left( \frac{f_i (\beta - 1)}{f} \right)^{1/\rho}, \\
s_n &= s_E
\end{align*}
\]

When we allow delocation, the first question to ask is: “Which firms move first?” To answer this, we calculate the incipient gains that atomistic firms would anticipate when no delocation has yet occurred. The present-value gain from delocation for X-type firms would be:

\[
\nu_X[a] = \frac{a^{1-\sigma}}{\delta \sigma} (1 - \phi)(B - B^*) > 0; \quad 0 \leq a \leq a_X
\]
Figure 3: Delocation gains in initial, no-delocation situation.

Note that the magnitude of this change is highest for firms with the lowest marginal cost and it is always positive since B>B* in the initial situation. For Southern D-type firms migrating to the North, the gain would be:

\[
\nu_D[a] = \frac{d^{1-\sigma}}{d\sigma} (B - B^*) > 0; \quad a_X \leq a \leq a_D
\]

Figure 3 graphs (14) and (15) and shows that the firms that have the most to gain are the most efficient firms, thus it is the most efficient firms that will tend to move first. This result is intuitively obvious, since the HME is driven by firm’s desire to minimise trade costs by locating near the big market and large/efficient firms sell more and thus have more to gain from moving to the big market. Note that by the definition of \(a_X\), firms with \(a=a_X\) are just indifferent to exporting so \(v_X[a]\) and \(v_D[a]\) touch at \(a_X\) as shown in Figure 3.

**Result 2:** The most efficient export firms in the small nation move to the large nation first. The local small firms have less incentive to move to the large nation.

**The order of delocation**

Since the most efficient firms have the most to gain from moving, the most efficient Southern firms move first. The delocation itself, however, raises the degree of competition in big North market and lowers it in the South, so the value to any given firms of moving declines as the range of firms that have moved expands. The process continues until the gain from moving is zero, i.e. the location condition, B=B*, holds (i.e. firms move until all firms are indifferent to location which happens when B=B*).

Formally, taking the range of southern firms that have moved to the North as \(a \in [0,a_R]\), where ‘R’ is a mnemonic for relocation, the \(\bar{m}\) ’s are:

\[
\bar{m} = s_E \int_{0}^{a_R} a^{1-\sigma} dG[a|a_D] + (1-s_E) \int_{0}^{a_X} a^{1-\sigma} dG[a|a_D] + (1-s_E) \int_{a_X}^{a_D} a^{1-\sigma} dG[a|a_D]
\]

\[
\bar{m}^* = \phi s_E \int_{0}^{a_X} a^{1-\sigma} dG[a|a_D] + (1-\phi)(1-s_E) \int_{0}^{a_X} a^{1-\sigma} dG[a|a_D] + (1-s_E) \int_{a_X}^{a_D} a^{1-\sigma} dG[a|a_D]
\]

The key change here is that now Southern varieties with a’s between zero and \(a_R\) are produced in the North and so are not subject to trade costs in the northern market. Solving these and using the solutions to define the B’s, we have:

---

11 See guide to calculations for a demonstration.
12 See Appendix for a more formal proof of these involving quadratic migration costs.
Solving the location condition, $B = B^*$, yields the ‘relocation threshold’ $a_R$ as a function of the export threshold $a_X$:

$$a_R = \frac{s_n + s_E - 1}{(1 - s_n)(1 - \phi)} x a_X^\alpha$$

Employing the initial condition $s_n = s_E$, the definition of $B$ in (176) and the export cut-off condition, we have:

$$a_R = \frac{\phi(2s_E - 1)}{(1 - \phi)(1 - s_E)} x a_X^\alpha, \quad \phi a_X^{1-\sigma} \frac{(\beta - 1) f_j}{(f_j (\beta - 1) / f)^{\beta - 1} + \phi x a_X^\alpha} = f$$

where $a_D$ and $n^w$ are pinned down by the initial conditions (no entry/exit yet). The second expression cannot be solved analytically for $a_X$ since $\alpha$ and $1-\sigma$ are different non-integer powers, but it has a unique and positive solution since the left-hand side is monotonically falling as $a_X$ rises and it starts at infinity when $a_X = 0$.

**Two phases of migration, XX and DD**

Inspection of (18) reveals that $a_R$ starts out at zero but gets progressively closer to $a_X$ as the level of trade freeness (i.e. $\phi$) rises. This implies that the migrating firms are initially X-types. When trade is sufficient, $a_R$ equals $a_X$. At this level of trade freeness, all southern X-types will have moved to the North. Beyond this level of trade freeness, a second phase of delocation begins where all the migrating firms are D-types. We refer to these two phases of migration as XX and DD migration.

The exact level of $\phi$ where DD migration begins is simple to calculate. Setting $a_R = a_X$ and solving for $\phi$ yields:

$$\phi^p = \frac{1 - s_E}{s_E}$$

where $\phi^p$ is the ‘partitioning’ level of trade freeness.

**Phase 2 delocation: DD migration.** When $\phi$ falls below $\phi^p$, migration affects the B’s differently because the south no longer exports manufactures to the North and the migrating firms no longer sell back into the southern market. The $m$’s are now:

$$m = \int_0^{a_E} a^{1-\sigma} dG[a | a_D] + s_E \int_{a_D}^{a_E} a^{1-\sigma} dG[a | a_D], \quad m^* = \phi \int_0^{a_E} a^{1-\sigma} dG[a | a_D] + (1 - s_E) \int_{a_D}^{a_E} a^{1-\sigma} dG[a | a_D]$$

which solve to:

$$m = \frac{a_R^{\alpha} + s_E (a_D^{\alpha} - a_R^{\alpha})}{(1 - 1/\beta) a_D^{\rho}}, \quad m^* = \frac{\phi a_X^{\alpha} + (1 - s_E)(a_D^{\alpha} - a_R^{\alpha})}{(1 - 1/\beta) a_D^{\rho}}$$

Using (20) to define the B’s and solving the delocation condition for $a_R$ yields:
\[ a_R = a_X \left( \frac{\phi_E}{1 - s_E} \right)^{1/\alpha} \]

The export cut-off condition for sales from the North to the South is the same as in the XX migration case given. From (10), we see that as before, \( a_R \) continues to rise gradually as trade gets freer.

Using the ratio of the cut-off conditions (21) can be written in terms of \( a_D \). Solving for the \( \phi \) where \( a_R = a_D \) tells us the level of openness where all firms have left the small nation – what is known as the ‘sustain point’ in economic geography, \( \phi^S \). It is:

\[ \phi^S = \left( \frac{1 - s_E}{s_E} \right)^{1/\beta} \]

Note that since \( \beta > 1 \), and \( se > \frac{1}{2} \) the \( \phi^P < \phi^S < 1 \). This means that as trade freeness, \( \phi \), limits to 1, it will eventually pass the sustain point \( \phi^S \) and so full agglomeration will occur. However, as \( \beta \) increases, full agglomeration is less likely. At extreme, when \( \beta = \infty \), full agglomeration never happens until trade costs are zero; in this case the location of production becomes completely indeterminate. For \( \beta \) to approach infinity, \( \rho \) must approach infinity (firm distribution is very skewed to the left towards the highest marginal cost, \( a_0 \)), or \( \sigma \) must approach unity.

**Figure 4: Geographic distribution of firms with free delocation.**

To summarise:

**Result 3:** (Two phases of firm delocation). The gradual reduction of iceberg trade costs produces a gradual delocation of firms from the small nation to the big nation as in NEG models, but here firms leaving in the order of rising marginal cost. The delocation process is marked by two phases. Migration of small-nation X-types (who remain X-types after their relocation) takes place first. In the second phase, which starts once all X-type firms have left the small nation, migration consists of small-nation D-type firms (who remain D-types in their new location). The ‘partitioning’ threshold between XX and DD migration is defined by the level of trade freeness, \( \phi^P \), and this equals \( (1-s_E)/s_E \); the sustain point is \( ((1-s_E)/s_E)^{1/\beta} \).

**Result 4:** As the distribution of firms is skewed towards low productivity, the sustain point
rises implying less agglomeration for any given level of trade freeness. At the extreme of an infinity skewed distribution (\(\rho = \infty\)), full agglomeration cannot occur.

The spatial distribution of firms according to their level of efficiency is depicted schematically in Figure 4. Note that the big northern market has an extra allocation of the most efficient firms while the southern market has none. The diagram shows the situation for a level of trade freeness, \(\phi\), where the delocation is still in the XX phase.

### 3.2.1 Allowing entry/exit

In parallel with the previous section, we turn now to consider entry/exit. That is, we now add back the two free entry conditions and the domestic cut-off condition. To be concrete, we assume that \(\phi\) is such that marginal migration is of the XX type, i.e. \(0 < \phi < \phi^P\), although the analysis would be qualitatively identical for \(\phi > \phi^P\). As always, the B’s are critical. Using (17) to eliminate \(a_R\) and the ratio of cut-offs to eliminate \(a_X\), the domestic cut-off condition is:

\[
a_D^{1-\sigma} B = f, \quad B = \frac{s_E (1 - 1/\beta) a_D^{1-\sigma} / n^w}{s_a a_D^\sigma + s_E \phi^B a_D^\sigma}
\]

Solving this for \(n^w\):

\[
(22) \quad n^w = \frac{(\beta - 1) / f \beta}{s_a / s_E + \phi^B}
\]

Since \(s_n = s_E\) in the case at hand, the comparison of (22) and (13) makes it clear that allowing entry/exit after instantaneous delocation results in a reduction in the number of active firms. In other words, delocation does not fully relieve the HME pressure; global mass of firms must fall to restore expected pure profits to zero. This of course, requires the firm-death rate to exceed the firm birth rate. Since the location condition assures the equality of profitability in the two nations, we assume that this affects each potential innovator in the world in the same way. This is natural to assume that the overall reduction in \(n^w\) will fall pro rata on Northern and Southern varieties, so the share of firms owned by Northern citizens, \(s_n\), will not change. In short, we assume that \(s_n = s_E\) even after the global mass of firms adjusts.

Using the ratio of cut-offs and the free entry conditions we have:

\[
\frac{1}{n^w} = f + f \left( \frac{a_X}{a_D} \right)^\alpha + f_I \left( \frac{1}{a_D} \right)^\alpha
\]

Using (22) to solve this for \(a_D\) and using the ratio of cut-off conditions to get \(a_X\), we have:

\[
\frac{E^w}{n^w} - \bar{j} = 0
\]

---

13 While (9) is the standard way of writing the free-entry condition, it is conceptually much simpler to work with a global free entry condition and \(B = B^*\). That is, instead of using the pair of free-entry conditions, we can use a single, global entry condition and \(B = B^*\) to determine \(n^*\) and \(s_n\). The global free entry condition starts from the fact that 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 is the average operating profit earned in the market. This average operating profit is \(E^w/\sigma n^w\). Each unit of capital developed, however, does not lead to a ‘winner.’ The ex ante expected fixed cost of getting a winner (i.e. developing a D- or X-type variety) is: 

\[
F = F + F \frac{G[a_X]}{G[a_D]} + F_I \frac{1}{G[a_D]}
\]

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 \(F\) 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. \(F_I\) times \(1/G[a_D]\) is the number of ‘tries’ needed to get a winner. The global free entry condition is thus:

\[
\frac{F^w}{n^*} - \bar{j} = 0
\]
(23) \[
a_D = \left(\frac{f_i(\beta - 1)/f}{\beta(s_n/s_E - 1) + (1 + \phi \beta)}\right)^{1/\rho}, \quad a_X = \left(\frac{\phi \beta f_i(\beta - 1)/f}{\beta(s_n/s_E - 1) + (1 + \phi \beta)}\right)^{1/\rho}
\]

(22), (23) and (17) characterise the equilibrium. Comparing (23) to (13), we see that the impact of trade and delocation involves the standard HFT model effects of lowering the threshold for D-type firms where \(s_n = s_E\).

### 3.2.2 Trade, production and productivity effects

The effects of integration are much richer in the instantaneous-delocation-then-entry case. As in the instantaneous entry case in section 3.1, we get share-shifting and selection effects via the change in \(a_D\). However, we get a new effect from delocation, namely the change in \(a_R\). This means that freer trade has an extra large productivity effect on the big nation since the firms that migrate to the big nation in response to freer trade are systematically more efficient than the least efficient of the existing firms. Correspondingly, the small nation experiences less of a pro-productivity effect. Indeed as we shall see, the small nation’s productivity may actually fall as trade gets freer.

Figure 5, which is akin to Figure 2, shows the impact on the North’s and South’s average productivity when the economy is in the XX-phase of firm migration (delocating firms are X-types). For the North, freer trade pivots the density around \(a_X\) and so unambiguously shifts weight to lower values of \(a^{1-\sigma+\rho}\). Thus the North’s weighted-average ‘a’ falls, i.e. average productivity rises. The right panel shows the impact on the South. Here, we get the usual share-shifting and selection effects, but since the South loses its most efficient firms (the weight on \(a^{1-\sigma+\rho}\) between \(a_R\) and \(a_X\) falls to zero), the overall productivity impact on the sales-weighted ‘a’ is dampened. Indeed if the \(a^{1-\sigma+\rho}\) curve is rising steeply enough, the average can actually rise, as we shall see below.

![Figure 5: Productivity effects, free delocation case.](image)

In the XX-case, our productivity measure is:

\[
\bar{a}^* = \int_{a_X}^{a_D} a^{1-\sigma} B(1 + \phi) \frac{T S^*}{n^*} dG[a|a_D; a_R] + \int_{a_X}^{a_R} a^{1-\sigma} B(1 + \phi) \frac{T S^*}{n^*} dG[a|a; a_R]
\]

where

\[
T S^* = \int_{a_X}^{a_D} a^{1-\sigma} B n^* dG[a|a_D; a_R] + \int_{a_X}^{a_R} a^{1-\sigma} B(1 + \phi) n^* dG[a|a_D; a_R], \quad dG[a|a_R; a_D] = \frac{\rho a^{\sigma-1}}{a_D^\rho - a_R^\rho} da
\]
Solving the integral, we get:

\[
\overline{a}^* = \left( \frac{\alpha}{1 + \alpha} \right) \left( 1 + \phi \right) a_X^{1+\alpha} - (1 + \phi) a_R^{1+\alpha} + a_D^{1+\alpha} - a_R^{1+\alpha}
\]

In the case of DD-phase delocation (firms that move are D-types), we have explicit solutions and the formal expression for the weight average ‘a’ can be written as:

\[
\overline{a}^* = \int_a^{a_D} \left( a^{1-\sigma} B \right) n^* dG[a_D; a_R]; \quad TS^* = \int_{a_R}^{a_D} a^{1-\sigma} Bn^* dG[a_D; a_R]
\]

Solving the integral, we get:

\[
\overline{a}^* = \left( \frac{\alpha}{1 + \alpha} \right) a_D^{1+\alpha} - a_R^{1+\alpha}
\]

Since \(a_D\) falls with freer trade and \(a_R\) rises, the productivity effect is ambiguous. In particular, using (21), the ratio of cut-offs and (23), we have

\[
(24) \quad \overline{a}^* = \left( 1 + \phi \right)^{\frac{1}{\rho}} \left( 1 - \left( \frac{\phi \beta s}{(1-s)} \right)^{1-\sigma} \right) \left( \frac{1 - \sigma + \rho f(1-\sigma + \rho)}{2 - \sigma + \rho f(\sigma - 1)} \right)
\]

where \(s=s_n=s_E\) is given by the initial conditions. Note that the first right-hand term is decreasing in \(\phi\) and so tends to imply that freer trade is pro-productivity; the second term however is increasing in \(\phi\) and thus anti-productivity. 14 We can vary the relative importance to the two conflicting terms by varying \(\rho\). When \(\rho\) is small (the regularity condition only requires \(\rho\) to be bigger than 1-\(\sigma\)), the pro-productivity term gets stronger relative to the anti-productivity term, so the overall derivative is negative (freer trade lowers the average a). Conversely, when \(\rho\) is large, freer trade tends to be anti-productivity. Heuristically speaking, when \(\rho\) is high, then the line \(a^{1+\rho-\sigma}\) is very steep and this give great import to the loss of weight on low values of a’s due to the rise in \(a_R\).

For example, in the special case of \(\sigma=\rho=2\), the productivity impact of freer trade, \(d\overline{a}/d\phi\), equals \(\phi (s(2+\phi)/(1-s) - 1)/2(1+\phi^2)^{3/2}\), which is positive since \(s>1/2\). In other words, freer trade lowers productivity in the small nation in this case.

To summarise, we write

**Result 5: (new productivity effects).** Allowing delocation introduces a new productivity effect; since the most efficient firms migrate first and freer trade encourages such delocation, freer trade leads to an extra large productivity gain in the large nation which receives the migrating firms, but dampens the productivity gains in the small nation. We showed that during the DD-phase of migration, freer trade always boosts productivity in the large nation, but may raise or lower productivity in the small nation. If \(\rho\) is large enough, freer trade lowers the small nation’s productivity.

### 3.3. General model

Modelling the entry/exit and delocation process more generally than we have with our two polar cases runs into the fundamental indeterminacy of the system. We have five equilibrium conditions – the two cut-off conditions, the two free-entry conditions and the location condition – and five

14 Expressing the second term as \((1-x^3)/(1-x)\) which is increasing in x if and only if \(a>1\).
equilibrating variables, \( a_D, a_X, n, n^* \) and \( a_R \). However, the two free-entry conditions and the location condition are not independent; any two imply the third. For example, if the two free-entry conditions hold, the location condition \( B=B^* \) holds automatically. If the location condition and one free entry condition hold, the other free entry condition is automatically satisfied.

This means that we cannot determine \( s_n, n^w \) and \( a_R \) in general; the most we can say is that:

\[
(25) \quad a_R^\alpha = \frac{s_n + s_E - 1}{(1 - s_n)(1 - \phi)} a_X^\alpha, \quad n^w = \frac{(1 - 1/\beta)/f}{s_n / s_E + \phi\beta}
\]

where the cut-offs are defined as in (23). The HFT standard model assumes \( a_R=0 \) and is thus able to find \( s_n \). The delocation-then-entry/exit case fixes \( s_n \) with initial conditions and is thus able to find \( a_R \).

To pin down \( s_n, n^w \) and \( a_R \) in general would require us to introduce quadratic adjustment costs for both delocation and entry/exit. The result would be a system of six differential equations for the six state variables (the three state variables \( n^w, s_n \) and \( a_R \), and their corresponding co-state variables). Working with more than two differential equations is difficult and rarely rewarding. In this case, the final result would be that some of the Home Market Effect adjustment would occur via delocation and some would occur via entry/exit. Thus, \( s_n \) would be higher than \( s_E \) in the general model as some of the HME pressure would have been relieved via delocation and some via entry/exit.

4. SMALL NATION WELFARE

Section 3.2.2 showed that delocation may, under some circumstances, lower the productivity of the small nation’s M-sector, while Section 3.1.2 showed that the small-nation productivity always rose when delocation was forbidden from freer trade. This raises the question of whether the small nation might, under certain parameter conditions, benefit from forbidding delocation while trade costs fall. The large nation, as we saw, always experiences a pro-productivity effect of lower trade costs, so the welfare impact on the large nation is unambiguously positive (the pro-productivity effect reinforces the lower cost of imports).

Given (1), per-capita welfare in the small nation can be written as:

\[
U^* = E - \mu + \mu \ln(\mu m^* \sigma^{-1})
\]

Where \( m^* \) is the weighted average of marginal selling costs in the small market, see (16), which covers both imports and locally produced varieties.

Plainly, all effects will come through \( m^* \), so the axis of investigation will compare the impact of freer trade on \( m^* \) in 1) the instantaneous-entry-then-delocation equilibrium (recall that there is no delocation in this case), and 2) the instantaneous-delocation-then-entry case. For convenience, we call case 1) the no-delocation case and case 2) the latter the delocation case.

Since the small-nation domestic cut-off condition, namely \( a_D^{1-\sigma}B^*=f \), holds in both cases and \( B^*=(1-s_E)/m^* \), we have that \( m^* = a_D^{1-\sigma}(1-s_E)/f \). The comparison of \( m^* \)'s in the two cases boils down to a comparison of \( a_D \)'s in the two cases. In particular, \( m^* \) in the no-delocation case exceeds \( m^* \) in the delocation case if and only if, \( a_D^{1-\sigma} \) is higher in the no-delocation case. Given our welfare criteria, the no-delocation case is preferred if and only if \( a_D^{1-\sigma} \) is higher in the no-delocation case. From (10) and (23), the \( a_D \)'s in the two cases are:

\[
(26) \quad (a_D^{\text{deloc}})^{1-\sigma} = \left( \frac{\beta(s_n / s_E - 1) + 1 + \phi\beta}{f^*(\beta - 1)/f} \right)^{\sigma^{-1}/\rho}, \quad (a_D^{\text{no deloc}})^{1-\sigma} = \left( \frac{1 + \phi\beta}{f^*(\beta - 1)/f} \right)^{\sigma^{-1}/\rho}
\]
By inspection, we see that welfare comparison turns entirely on the issue of whether \( s_n \) exceeds \( s_E \) when delocation is allowed. Formally, our metric for the welfare effects of delocation can be written as:

\[
\frac{m^{\Delta_{deloc}}}{m^{\Delta_{no
deloc}}} = (1 + \beta \frac{s_n/s_E - 1}{1 + \phi''})^{1/\beta}
\]

where delocation is welfare improving if this ratio exceeds unity, i.e., if \( s_n > s_E \) when delocation is allowed.

What determines the relative size of \( s_n \) and \( s_E \) when delocation is allowed? Section 3 considered two extreme cases: instant entry/exit then delocation, and instantaneous delocation then free entry/exit. In the instantaneous delocation case, delocation equalised the attractiveness of the two markets at all moments, so we assumed that any entry/exit would affect all potential firms equally in the two nations. As a consequence, the fall in the world number of varieties was spread on a pro rata basis across the two nations — so \( s_n \) remained equal to \( s_E \). Combining this result with (27), we see that if delocation is instantaneous, it has no welfare effects. In other words, the small nation’s welfare is the same under instantaneous delocation and under forbidden delocation.

In the general case discussed in Section 3.3, however, neither free entry/exit nor delocation is instantaneous. Freer trade will make the large nation more attractive in transition to the long-run equilibrium (this is what creates the Home Market Effect pressure). When delocation is allowed but occurs slowly, some of the HME pressure will be relived via delocation to the North of firms that were create in the South, and some of it will be relived by faster firm creation in the North so that \( B = B^* \) in the long run. We cannot determine what \( s_n/s_E \) will be in the general case, but we can be sure that it will exceed its autarky value of unity, so entry will have been faster in the big market. When delocation is forbidden, by contrast, we know that \( a_D \) will be characterised by the second expression in (26). Given (27), this tells us that delocation is always welfare improving for the small nation. To summarise:

**Result 6:** Allowing delocation in the face of trade liberalisation improves in the small nation’s welfare.

### 5. CONCLUDING REMARKS

Eaton and Kortum (2002) and Melitz (2003) opened a new line of research in international trade by providing tractable models where the entry and exit of firms determines the distribution of firm-level efficiency endogenously. One of the most notable results in this heterogeneous-firms trade literature concerns the impact of trade integration on productivity; freer trade raises average productivity by forcing out the weakest firms and shifting production shares to the most productivity firms. These models, however, ignore the possibility that freer trade may alter the firm-size distribution via international firm migration, what is called offshoring in North America and delocation in Europe. Firms in the standard HFT models are assumed to produce in the nation in which they are ‘born.’

Our paper relaxes this assuming in presenting a model that allows the equilibrium firm-size distribution to be influenced by delocation as well as entry and exit. We show that this additional channel of adjustment implies that freer trade affects average productivity and the firm-size distribution in new ways. In particular, the Home-Market-Effect pressure created by freer trade results in the most efficient firms moving from the small nation to the large nation. This in turn implies that freer trade has an extra large pro-productivity impact in the large nation. In the small
nation, the delocation always mitigates the positive productivity effect, and may, in some circumstances, result in an anti-productivity effect.

The paper also shows that delocation is a complex phenomenon in the presence of heterogeneous firms and fixed market-entry costs. When trade costs are high, the firm-migration (delocation) involves small-nations exporting firms (i.e. firms that sell locally and export both before and after moving). When trade costs get low enough, all exporting firms will have switched from the small nation to the large and subsequent delocation involves firms that sell only locally. In this phase, delocation reduces the number of varieties available in the small market.

Regardless of the anti-productivity anti-variety effects of delocation, we show that the small nation always benefits by allowing delocation.

REFERENCES


Okubo, T (2006) “Anti-agglomeration Subsidies with Heterogeneous Firms”, mimeo


**APPENDIX**

To determine the order of delocation, however, we must consider the delocation process in more detail. As we shall see, quadratic adjustment costs smooth out the firm-migration process and ensure that the firms with the most to gain are the first to move. The specific formulation of the quadratic adjustment costs is:

\[
\text{cost} = \gamma \dot{x}
\]

where \(\dot{x}\) is the flow of delocating firms and \(\gamma\) is the adjustment cost parameter.

### 5.1.1 Quadratic adjustment costs and the order of delocation

Atomistic firms observe the actual level of adjustment costs and move if the gain from doing so exceeds the adjustment cost. We observe from (14) and (15) that if the flow of delocation were zero, then all Southern firms would wish to move instantaneously (since \(B>B^*\)), but at first, the flow and thus the cost would be infinite. Since the gain to the most efficient Southern firms (those with \(a=0\)) is infinity, only the most efficient Southern firms will delocate at the first instant. Once these firms have moved, the flow of delocation would drop from infinite to a finite level, bringing down the delocation cost to a high, but finite level. Firms for whom the value of delocating is less
than or equal this finite level will move next. Given (14) and (15), this implies that the delocation process is marked firms moving in order of efficiency, with the most efficient firms moving first.

The process continues with both the cost and value of moving approaching zero. Importantly, the delocation process makes the big North market more competitive and the Southern market less competitive, so the value to any given firm of moving declines as the range of firms that have moved expands. This slows the rate of delocation and thereby the level of adjustment costs, so the process occurs smoothly. Ultimately, the process stops when the range of delocated firms rises to the point where the two markets are equally attractive, i.e. $B = B^*$. The range of firms that delocate in response to any given level of trade freeness is $[0, aR]$, where $aR$ (the ‘R’ is a mnemonic for relocate) is the threshold level of marginal cost defined by what we call the ‘delocation condition’:

$$B = B^*$$

While delocation is occurring, it is characterised by the equivalence of the benefit and cost of changing locations, namely:

$$v[a'] = \gamma \left[ n^* (a')^{\rho} a_D^{-\rho} \right]$$

where $a'$ is the marginal cost of the current migrants and the right-hand-side terms in brackets represent the rate of firm migration (recall that the $n^*$ and $aD$ are fixed by the initial conditions described above).

As mentioned, delocation alters average competitiveness of each market, so the definition of the B’s in the delocation condition, (16), is more complicated than the definition the Section 3.1. Specifically for near-zero levels of trade freeness, $B = sE/n^w \bar{m}$ where $\bar{m}$ is now defined to take account of the delocation of firms with a’s in the $[0, aR]$ range:

$$\bar{m} = sE \int_0^{aD} a^{-\sigma} dG[aD] + (1 - sE) \int_0^{aE} a^{-\sigma} dG[aD] + (1 - sE) \int_{aD}^{aE} a^{-\sigma} dG[aD]$$

where we have set $sH = sE$ by the initial conditions. The key change here is that now southern varieties with a’s between zero and $aR$ are produced in the north and so are not subject to trade costs in the northern market. $B^*$ equals $(1-sE)/n^w \bar{m}^*$, where the expression for $\bar{m}^*$ is similar to the above expression for $\bar{m}$.

The delocation condition $B = B^*$ can be solved analytically to get $aR$ as a function of $aX$:

$$aR = \left( \frac{\phi(2sE - 1)}{(1 - \phi)(1 - sE)} \right) \frac{1}{1 - \sigma + \rho} aX$$

Using the definition of B with $\bar{m}$ in (16) and $aR$ as defined in (18), the export cut-off condition implicitly defines $aX$ as:

$$\frac{\phi aX^{1-\sigma}}{\phi aX^{1-\sigma + \rho} + ((\beta - 1)f_1 f)^{\beta - 1}} = \frac{f/f_1}{\beta - 1}$$

---

15 This follows from the fact that when Southern firms with marginal costs below $a'$ have migrated to the North, the mass of firms that has moved is $\int_{a'}^n n^* dG[aD]$, which solves to $(n^* (a')^{\rho})/aD^{-\rho}$; taking the time derivative yields the result in the text.
This condition cannot be solved analytically since $\sigma$ and $\rho$ are not necessarily integers, but it is clear that there is a unique solution since the left-hand side is monotonically falling as $a_X$ rises and it starts at infinity when $a_X=0$. 
GUIDE TO CALCULATIONS FOR REFEREES (NOT FOR
PUBLICATION)

Footnote 9 calculations.

Solving the integrals, \( TP = \frac{nB}{(1-1/\beta)a_D^{-\sigma}}(\phi a_X^{-1-\sigma}\rho + a_D^{-1-\sigma}\rho) \). Using the ratio of cut-offs it becomes \( TP = \frac{nB(1+\phi^\beta)a_D^{-1-\sigma}}{(1-1/\beta)} \). Using this and the definition of the conditional density function in (11):

\[
\bar{a} = \int_0^a \left( a^{1-\sigma} B(1+\phi)(1-1/\beta) \right) \left( n\rho a^{-1-\sigma} a_D^{-\rho} \right) da + \int_{a}^\infty \left( a^{1-\sigma} B(1-1/\beta) \right) \left( n\rho a^{-1-\sigma} a_D^{-\rho} \right) da
\]

which simplifies to:

\[
\bar{a} = \int_0^a \left( (1+\phi)\rho(1-1/\beta) \right) a^{1-\sigma+\rho} da + \int_{a}^\infty \left( \rho(1-1/\beta) \right) a^{\rho} da
\]

Gathering terms and then using the expression for the equilibrium \( a_D \) from (10), namely

\[
a_D = \left( \frac{f_I(\beta - 1)}{(1 + \phi^\beta)} \right) \frac{1}{\rho},
\]

this simplifies to:

\[
\bar{a} = \int_0^a \left( (1+\phi)\rho(1-1/\beta) \right) a^{1-\sigma+\rho} da + \int_{a}^\infty \left( \rho(1-1/\beta) \right) a^{\rho} da
\]

Straightforward rearrangement yields the expression in the text.

Footnote 11 Calculations:

\( B>B^* \) in the absence of adjustment is what drives the HME. Formally, it is proved by contradiction. Suppose \( B^* \geq B \), then

\[
B = \{s_E(1-1/\beta)a_D^{-\rho}/n^\alpha\}/\{s_E a_D^{-\alpha} + (1-s_E)\phi a_X^{-\alpha}\} \quad \text{where } \alpha=1-\sigma+\rho
\]

and

\[
B^* = \{(1-s_E)(1-1/\beta)a_D^{-\rho}/n^\alpha\}/\{(1-s_E)a_D^{-\alpha} + s_E\phi a_X^{-\alpha}\} \quad \text{where the export cut-off conditions imply } a_X \geq a_X^* \text{ since the marginal firm export into the market with the bigger per-firm demand can have higher marginal costs. But substituting } a_X \geq a_X^* \text{ into } B^* \geq B \text{ implies a contradiction with } s_E>\frac{1}{2} \text{ and } B^* \geq B \text{, so it must be that } B>B^*.}