NBER WORKING PAPER SERIES

INTERNATIONAL TRADE, OFFSHORING AND HETEROGENEOUS FIRMS

Richard Baldwin
Toshihiro Okubo

Working Paper 16660
http://www.nber.org/papers/w16660

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2011

International Trade, Offshoring and Heterogeneous Firms
Richard Baldwin and Toshihiro Okubo
NBER Working Paper No. 16660
January 2011
JEL No. F1,F12,F2

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.

Richard Baldwin
Cigale 2
1010 Lausanne
SWITZERLAND
and CEPR
and also NBER
baldwin@graduateinstitute.ch

Toshihiro Okubo
RIEB Kobe University
2-1, Rokkodai cho, Nada-ku, Kobe,
657-8501 JAPAN
okubo@rieb.kobe-u.ac.jp

1. INTRODUCTION

Recent years have seen a flourishing of heterogeneous-firms trade models such as Eaton and Kortum (2002) and Melitz (2003). The key innovation in this literature is that it endogenises 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 boost productivity via changes in the equilibrium distribution of firm-level outputs as some the least efficient shut down and the most efficient expand.

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 individual firms – is an important empirical phenomenon, it is striking that heterogeneous-firms trade (HFT) models ignores this spatial relocation.[1]

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.[2] In this world, freer trade has two distinct effects. The first is the well-known pro-productivity effect discussed above. The second is that 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 (see, for example, the asymmetric case in Helpman, Melitz and Yeaple 2004). We argue that since entry/exit is not instantaneous, some of the home-market-effect pressure would be alleviated via relocation of the most efficient firms from the small market to the large market – as long as the possibility of spatial reallocation is 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 relocate and the least efficient firms would exit.

In this way, relocation 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 relocation 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 relocation 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.

---

[1] Many empirical studies found the offshoring of production using micro-data, such as Head and Ries (2003) and Tomiura (2007).
[2] Thus we are thinking of relocation from, e.g. Canada to the US, or Sweden to Germany rather than the US to China.

One recent branch of trade theory has focused on differences among firms. This literature is reviewed recently by Redding (2010) with the main theoretical papers being Eaton and Kortum (2002), Bernard, Eaton, Jensen and Kortum (2003), Melitz (2003), Helpman, Melitz and Yeaple (2004), Bernard, Redding and Schott (2007), Melitz and Ottaviano (2008), Yeaple (2005), Demidova (2005) and Chaney (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 the well-known core-periphery model was introduced by Krugman (1991), various models have been provided by the literature. The simplest of these models is the so-called Footloose Capital (FC) model proposed by Martin and Rogers (1995). The FC model features two, potentially asymmetric regions with only capital being inter-regionally mobile. Each firm is identified with one unit of capital; all 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, i.e. the shift of firms between regions does not change the size of market demands.

The paper that is closest to ours in combining heterogeneous firms and NEG considerations is our earlier paper, Baldwin and Okubo (2006b). 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. The main findings in that paper 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 (2009) and Okubo et al. (2010) are other papers in this line.[3]

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 and two sectors. 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.

Tastes of each consumer in either nation are given by a logarithmic quasi-linear utility function:

---

[3] Okubo (2009) studies the vertical linkage of input-output with heterogeneous firms in economic geography model. Okubo et al. (2010) employs the monopolistic competition model with linear demand and finds spatial sorting and co-agglomeration.

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

where C_M and C_A are, respectively, consumption of the composite of M-sector varieties and the numeraire, σ is the constant elasticity of substitution between any two M-sector varieties and μ 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). Capital is a "produced factor" rather than a primary factor. Creating a firm in the M-sector requires the creation of one unit of capital; this process involves only labour. 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).

Since there is one unit of capital per firm, international migration by firms (i.e. relocation) is synonymous with capital mobility in our model. As mentioned, all capital is owned by labourers who are immobile between nations and capital income is repatriated costlessly, so capital/firms move in search of the highest reward 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_I 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 the 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 δ.

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 foothold for each new variety in each market. We think of this as reflecting the cost of meeting market-specific standards and regulations, and locally establishing a brand name. The beachhead cost per market is denoted as *F;* it involves only labour inputs.[4]

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 in the numeraire sector 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.

---

[4] Many HFT models allow separate beachhead costs for the local (F_D) and export market (F_X), but this plays no role in our model so we assume that are identical to reduce notational clutter in the expressions (F_D=Fx=F).

The M-sector is marked by all the well known 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, by contrast, will include the iceberg costs, $t \geq 1$, marked up by the constant Dixit-Stiglitz mark-up. Given the logarithmic 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 world expenditure on M goods, namely E^w, is unity.

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

$$(2) \qquad c_j = \frac{(p_j)^{-\sigma}}{\overline{p}^{1-\sigma}} E; \qquad \overline{p} \equiv \left(\int_{i \in \Theta} p_i^{1-\sigma} di + \int_{h \in \Theta^*} \phi p_h^{1-\sigma} dh \right)^{1/(1-\sigma)}, \qquad \sigma > 1 \geq \phi \equiv t^{1-\sigma} \geq 0$$

where p_j is variety-j's producer price (which equals its consumer price since it is produced locally), and $\overline{p}$ is the usual CES price index in the North. The first term in the definition of $\overline{p}$ 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; Θ and Θ^* are sets of consumed goods that are produced in the North and the South, respectively. A parameter that plays a critical role in our paper is ϕ; we refer to it as the 'free-ness' (phi-ness) of trade, and note that ϕ 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 $\overline{p}$ and $\overline{p}^*$ 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, but this in turn depends upon the distribution of the marginal costs in each market. In general, a firm with any marginal cost 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 $\overline{p}$'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 degenerated so the integrals in $\overline{p}$ and $\overline{p}^*$ 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 <u>and</u> internationally mobile is the central technical difficulty of our model. We start on this problem by working through the closed-economy case.

2.2.   *Equilibrium under autarky*

We start with autarky, defined here as zero mobility of goods (i.e. $\phi = 0$) and firms since, 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, the steady-state present value of being an active firm with a marginal cost of a_j is:

$$\frac{p_j^{1-\sigma}}{\sigma \delta \overline{p}^{1-\sigma}} E \qquad where \quad p_j = \frac{a_j}{1-1/\sigma}, \quad \overline{p} \equiv \left(\int_{i \in \Theta} p_i^{1-\sigma} di \right)^{1/(1-\sigma)}$$

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) $\qquad a_D^{1-\sigma} B = f \quad where \qquad B \equiv \frac{E}{\overline{m}}, \quad \overline{m} \equiv n \int_0^{a_D} a^{1-\sigma} dG[a|a_D], \quad f \equiv \sigma \delta F$

where $\overline{m}$ reflects the marginal costs of active firms, 'n' is the mass of firms, and $G[a|a_D]$ is the conditional distribution function of 'a' (conditional on $a \le a_D$); here we have cancelled out the mark-ups and defined '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 the variety-generation process assumed, the a's of new varieties are distributed according to $\delta n G[a]$, $0 \le a \le 1$, 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 $\delta n G[a]$, namely $\delta n G[a|a_D]$, where $G[a|a_D]$ is the distribution function conditional on 'a' being less than a_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 a_D), thus the equilibrium present value of operating profits for a variety with a_j is:

$$a_j^{1-\sigma} \frac{B}{\delta\sigma} \ for \ 0 \le a < a_D; \qquad zero \ for \ a_D \le a;$$

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 a_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:

(4) $\qquad \int_0^{a_D} (a^{1-\sigma} B - f) dG[a] = f_I \qquad where \quad f_I \equiv \sigma \delta F_I$

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

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

(5) $\qquad G[a] = (\frac{a^\rho}{a_0^\rho}), \qquad 0 \le a \le a_0 \equiv 1, \quad \rho \ge 1$

where the parameters ρ 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:

$$\textbf{(6)} \qquad a_D = \left(\frac{f_I(\beta-1)}{f} \right)^{1/\rho}, \qquad n = \frac{E(\beta-1)}{f\beta} \quad where \quad \beta \equiv \frac{\rho}{\sigma-1}$$

Here we introduce the collection of parameters, β, 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.[5]

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: (1) Firms can physically migrate, moving their unit of capital from the small South to the big North; or (2) Firm births can exceed firm deaths in the North, while the opposite occurs in the South.
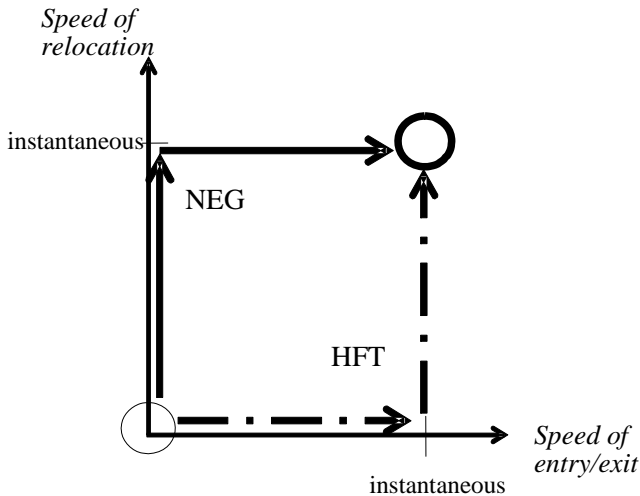
Figure 1: NEG and HFT solution to calculation difficulties.

The fact that relocation 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, <u>where</u> these varieties will be manufactured. The crux of the difficulty is that determining production location requires knowledge of the degree

---

of local competition, namely the $\overline{m}$'s, but defining these requires knowledge of where each variety is manufactured.[6]

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

First we assume entry/exit is instantaneous, and then we allow relocation (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 relocation 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 relocation*

As we shall see, when entry/exit is instantaneous but relocation is slow, all adjustment to the HME pressure occurs via entry/exit. The solution therefore is very similar to that of the well-known Helpman, Melitz and Yeaple (2004) equilibrium. The cut-off conditions for local and export sales are:[7]

(7) $$a_D{}^{1-\sigma} B = f, \quad \phi a_X{}^{1-\sigma} B = f$$

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

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

Here n and n* are the masses of firms producing in the North and the South.[8]

The North and South free entry conditions are:

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

Since we have normalised the size of the world economy such that world expenditure on M-varieties is unity (i.e. μ(L+L*) equals 1), the four equilibrium conditions in **(7)** and **(8)** can be solved analytically to yield:

---

[6] To highlight two contrasting cases, i.e. relocation versus entry/exit, we study instantaneous relocation and entry/exit in the static model. Chaney (2005) models the dynamics and finds overshooting of productivity when a country opens trade. Since his model is dynamic, the model successfully depicts asymmetric speed of entry and exit. Firms that had entered but found production unprofitable in autarchy would start production. The slow exit of firms could make more severe competition when trade opens but in the long-run it will be mitigated as firms eventually die.

[7] 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)..

[8] The first integral in the expression for $\overline{m}$ 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 its '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 $\overline{m}*$ is isomorphic, but φ applies to sales in the Northern market. Recall that the geometric weights are negative, so $\overline{m}$ and $\overline{m}*$ fall as prices rise.

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

where

$$s_E \equiv \frac{E}{E+E^*}, \qquad s_n \equiv \frac{n}{n+n^*}, \qquad n^w \equiv n+n^*$$

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 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 **(9)**, the expression of the equilibrium s_n. By inspection, we see that as trade gets freer (dϕ>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 ds_n/ds_E equals $(1+\phi)/(1-\phi)$,[9] 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 relocation

When firms are allowed to change locations (i.e. relocate), 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 relocate, i.e. instantaneous entry/exit is a perfect substitute for relocation, so the instant-entry-then-relocation case is identical to the standard HFT model.

In the next section, Section 3.2, we consider the diametrically opposed case, where we move from autarky to trade assuming that relocation 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-relocation-then-entry case in Section 3.2.

3.1.2   Trade, production and productivity effects

The effects of greater openness are well known. The rise in a_X means more firms in both nations export, so exports rise and this, in turn, heightens competition in both markets, so the sales of D-

---

[9] See Martin and Rogers (1995) for instance.

type firms fall in both markets; this is the well-known share-shifting effect. Additionally, the marginal D-types (those with a's near a_D) are driven out; this is the well-known selection effect. Both the selection and the share-shifting effects 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 is:

(10)
$$\bar{a} = \int_0^{a_X} a\left(\frac{a^{1-\sigma}B(1+\phi)}{TP}\right)ndG[a|a_D] + \int_{a_X}^{a_D} a\left(\frac{a^{1-\sigma}B}{TP}\right)ndG[a|a_D];$$

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:

$$TP \equiv \int_0^{a_X} a^{1-\sigma}B(1+\phi)\, n\, dG[a|a_D] + \int_{a_X}^{a_D} a^{1-\sigma}B\, n\, dG[a|a_D].$$

Note that the ratios in large parentheses are the production shares 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 average productivity.

As we shall see below, the productivity effects are substantially more complex when relocation occurs. To set the stage for the more complex analysis, we replicate Melitz's well-known pro-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 TP integrals and the equilibrium expression for a_D in **(9)**, we can re-write (10) as:[10]

$$\bar{a} = \int_0^{a_X} W_X(a^{1-\sigma+\rho})da + \int_{a_X}^{a_D} W_D(a^{1-\sigma+\rho})da$$

where W_X and W_D are the weights on X and D type firms:

$$W_X \equiv \frac{(1+\phi)}{(1+\phi^\beta)^{1/\beta}}\left(\frac{\rho(1-1/\beta)}{(f_I(\beta-1)/f)^{1-1/\beta}}\right), \qquad W_D \equiv \frac{1}{(1+\phi^\beta)^{1/\beta}}\left(\frac{\rho(1-1/\beta)}{(f_I(\beta-1)/f)^{1-1/\beta}}\right)$$

Plainly, W_X rises and W_D falls as trade gets freer (dϕ>0).[11]

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

(11)
$$\frac{d\bar{a}}{d\phi} = \int_0^{a_X}\frac{dW_X}{d\phi}a^{1-\sigma+\rho}da + \int_{a_X}^{a_D}\frac{dW_D}{d\phi}a^{1-\sigma+\rho}da + \frac{da_X}{d\phi}\left(W_X a_X^{1-\sigma+\rho} - W_D a_X^{1-\sigma+\rho}\right) + \frac{da_D}{d\phi}W_D a_D^{1-\sigma+\rho}$$

When d$\bar{a}$/dϕ is written in this way, we see that the productivity impact of freer trade depends upon the impact of dϕ on the 'weighting' terms – the dW/dϕ's – under the integrals – and the impact on the limits of integration – i.e. the last two terms. We know that the derivative (11) 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.

---

[10] See Baldwin and Okubo (2006a) for details.
[11] Note d ln{W_X}/dϕ is $(1-\phi^{\beta-1})/\{(1+\phi)(1+\phi^\beta)\}$; this is positive since β>1 (regularity condition).

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, "Weights (pre)". After the opening, the weighting function shifts to "Weights (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 third and fourth terms in (11). The first and second terms in (11) 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 term 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. 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'. Freer trade is pro-productivity in the South for the same reasons.
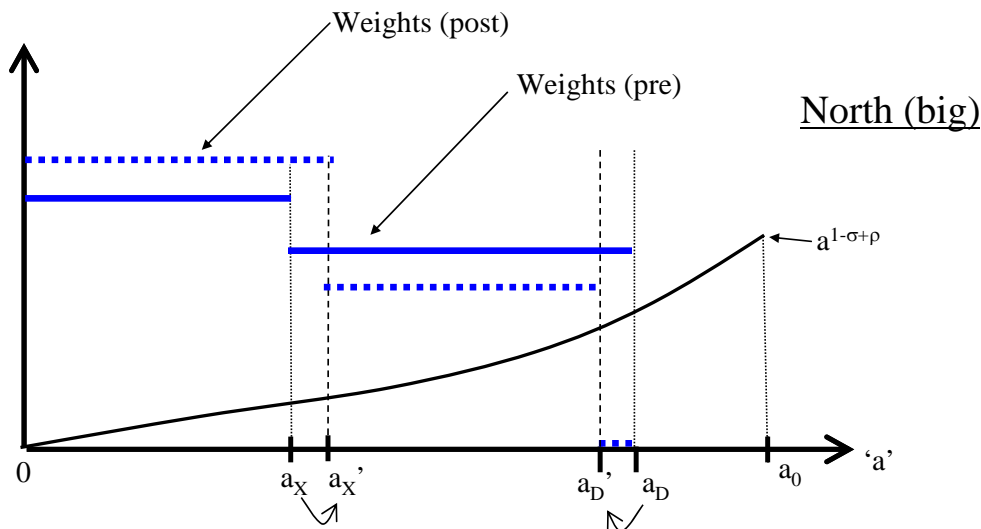

Figure 2: Trade liberalisation's productivity impact HFT model, free entry case.

3.2. *Instantaneous relocation followed by entry/exit*

The assumption of instantaneous birth-and-death of firms is clearly unrealistic. If entry/exit takes time, 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 relocation 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 relocation is instantaneous and then considers possibility of entry/exit. 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 firms that were active in autarky remain active initially when trade and firm-migration become possible.

The initial distribution of active firms under autarky are characterised in Section 2.2. Using (6)

(12) $$n^w = \frac{\beta-1}{f\beta}, \qquad a_D = \left(\frac{f_I(\beta-1)}{f}\right)^{1/\rho}, \qquad s_n = s_E$$

When we allow relocation, 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 relocation has yet occurred.[12] The present-value gain from relocation for X-type firms would be:

(13) $$v_X[a] = \frac{a^{1-\sigma}}{\delta\sigma}(1-\phi)(B-B^*) > 0; \qquad 0 \le a \le a_X$$

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.[13] For Southern D-type firms migrating to the North, the gain would be:

(14) $$v_D[a] = \frac{a^{1-\sigma}}{\delta\sigma}(B-B^*) > 0; \qquad a_X \le a \le a_D$$

Figure 3 graphs **(13)** and **(14)** 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.[14] This result is intuitively obvious, since the HME is driven by firm's desire to minimise trade costs by locating near the big market; 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.
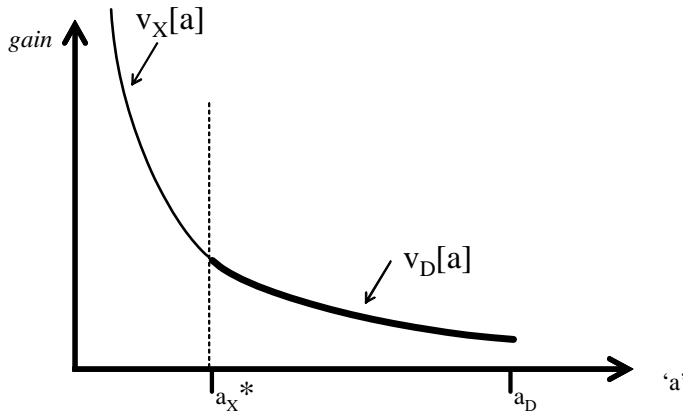
Figure 3: Relocation gains in initial, no-relocation situation.

---

[12] Our paper follows the relocation process of Baldwin and Okubo (2006b). See Appendix for more details and formal characterisation. The locational equilibrium line is assumed to be a sequential equilibrium rather than a Nash equilibrium. In the sequential equilibrium, firms move in order of productivity.
[13] See our first version, Baldwin and Okubo (2006a), for calculations.
[14] See Appendix for a more formal proof of these involving quadratic migration costs.

The order of relocation

Since the most efficient firms have the most to gain from moving, the most efficient Southern firms move first. The relocation 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 $\overline{m}$'s are:

(15)
$$\overline{m} = n\int_0^{a_D} a^{1-\sigma}dG[a|a_D] + n*\int_0^{a_R} a^{1-\sigma}dG[a|a_D] + n*\phi\int_{a_R}^{a_X} a^{1-\sigma}dG[a|a_D]$$

$$\overline{m}* = \phi n\int_0^{a_X} a^{1-\sigma}dG[a|a_D] + \phi n*\int_0^{a_R} a^{1-\sigma}dG[a|a_D] + n*\int_{a_R}^{a_D} a^{1-\sigma}dG[a|a_D]$$

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:

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

$$\alpha \equiv 1 - \sigma + \rho > 0$$

Solving the location condition, B = B*, yields the 'relocation threshold' a_R as a function of the export threshold a_X:

(16)
$$a_R^\alpha = \frac{s_n + s_E - 1}{(1-s_n)(1-\phi)}\phi a_X^\alpha$$

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

(17)
$$a_R = \left(\frac{\phi(2s_E - 1)}{(1-\phi)(1-s_E)}\right)^{\frac{1}{\alpha}} a_X, \qquad \frac{\phi a_X^{1-\sigma}(\beta-1)f_I}{\left(f_I(\beta-1)/f\right)^{1-1/\beta} + \phi a_X^\alpha} = f$$

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 α and $1-\sigma$ are different non-integer powers, but the expression 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 **(17)** reveals that a_R starts out at zero but gets progressively closer to a_X as the level of trade freeness (i.e. ϕ) rises. This implies that the migrating firms are initially X-types as discussed above. When trade is sufficiently free, 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 relocation 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 ϕ where DD migration begins is simple to calculate. Setting $a_R = a_X$ and solving for ϕ yields:

(18)
$$\phi^P = \frac{1 - s_E}{s_E}$$

where ϕ^P is the 'partitioning' level of trade freeness.

When ϕ falls below ϕ^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 $\overline{m}$'s are now:

$$\overline{m} = (n + n^*)\int_0^{a_R} a^{1-\sigma}dG[a|a_D] + n\int_{a_R}^{a_D} a^{1-\sigma}dG[a|a_D], \qquad \overline{m}^* = (n + n^*)\phi\int_0^{a_X} a^{1-\sigma}dG[a|a_D] + n^*\int_{a_R}^{a_D} a^{1-\sigma}dG[a|a_D]$$

which solve to:

(19)
$$\overline{m} = \frac{a_R{}^\alpha + s_n(a_D{}^\alpha - a_R{}^\alpha)}{(1 - 1/\beta)a_D{}^\rho / n^w}, \qquad \overline{m}^* = \frac{\phi a_X{}^\alpha + (1 - s_n)(a_D{}^\alpha - a_R{}^\alpha)}{(1 - 1/\beta)a_D{}^\rho / n^w}$$

Using **(19)** to define the B's and $s_n = s_E$, and solving the relocation condition for a_R yields:

(20)
$$a_R = a_X \left(\frac{\phi s_E}{1 - s_E}\right)^{1/\alpha}$$

The export cut-off condition for sales from the North to the South is the same as in the XX migration case given. From **(9)**, we see that as before, a_R continues to rise gradually as trade gets freer.

Using the ratio of the cut-off conditions **(20)** can be written in terms of a_D. Solving for the ϕ where $a_R = a_D$ tells us the level of openness at which all firms have left the small nation – what is known as the 'sustain point' in economic geography, ϕ^S. It is:

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

Note that since $\beta > 1$, and $s_E > \frac{1}{2}$ the $\phi^P < \phi^S < 1$. This means that as trade freeness, ϕ, limits to 1, it eventually passes the sustain point ϕ^S and so full agglomeration occurs. However, as β increases, full agglomeration is less likely. At extreme, when $\beta = \infty$, full agglomeration never happens except for costless trade in which case any spatial location of production is an equilibrium, including full agglomeration. For β to approach infinity, ρ must approach infinity (this means that the firm marginal cost distribution is heavily skewed to the left, i.e. towards the highest marginal cost), or σ must approach unity.

To summarise:

Result 3: (Two phases of firm relocation). The gradual reduction of iceberg trade costs produces a gradual relocation of firms from the small nation to the big nation as in NEG models, but firms the most efficient firms leave first. This process is marked by two phases: migration of small-nation X-types (who remain X-types after their relocation), and – once all X-type firms have left the small nation – migration of small-nation D-type firms (who remain D-types in their new location). The 'partitioning' threshold between the two phases of migration is defined by the level of trade freeness, ϕ^P, which equals $(1-s_E)/s_E$; all firms are in the big region when trade freeness ϕ surpasses the sustain point $((1-s_E)/s_E)^{1/\beta}$.

Result 4: The sustain point rises (implying less agglomeration for any given level of trade freeness), as Pareto distribution of firm marginal costs is skewed rightward. 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, ϕ, where the relocation is still in the XX phase.
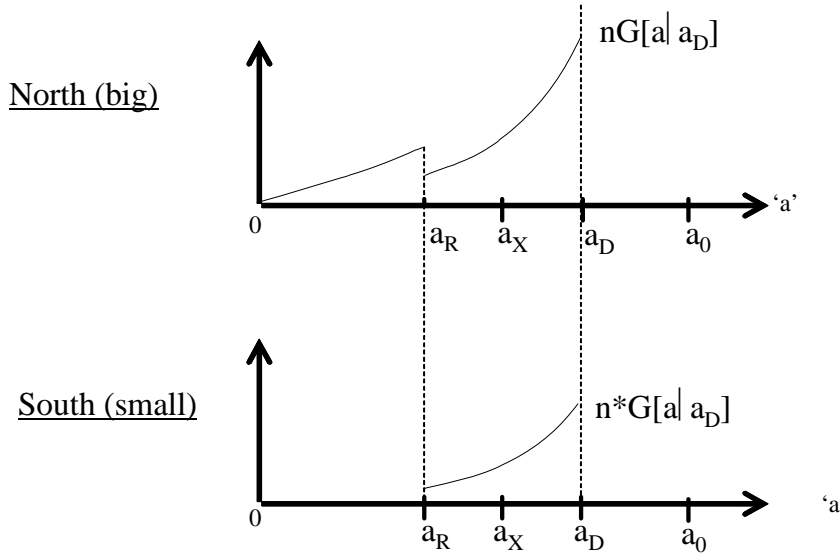
Figure 4: Geographic distribution of firms with free relocation.

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 ϕ 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 **(16)** to eliminate a_R and the ratio of cut-offs to eliminate a_X, the domestic cut-off condition is:

$$a_D^{\,1-\sigma} B = f, \qquad B = \frac{s_E(1-1/\beta)a_D^{\,\rho}/n^w}{s_n a_D^{\,\alpha} + s_E \phi^\beta a_D^{\,\alpha}}$$

Solving this for n^w:

(21) $$n^w = \frac{(\beta-1)/f\beta}{s_n/s_E \ - \ \phi^\beta}$$

Since $s_n = s_E$ in the case at hand, the comparison of **(21)** and **(12)** makes it clear that allowing entry and exit after instantaneous relocation results in a reduction in the number of active firms. In other words, relocation does not fully relieve the HME pressure; global mass of firms must rise to restore expected pure profits to zero. This of course, requires the firm-birth rate to exceed the firm-death 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 rise 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:[15]

$$\frac{1}{n^w} = f + f(\frac{a_X}{a_D})^\rho + f_I(\frac{1}{a_D})^\rho$$

Using **(21)** to solve this for a_D and using the ratio of cut-off conditions to get a_X, we have:

(22) $$a_D = (\frac{f_I(\beta-1)/f}{\beta(s_n/s_E - 1) + (1+\phi^\beta)})^{\frac{1}{\rho}}, \qquad a_X = (\frac{\phi^\beta f_I(\beta-1)/f}{\beta(s_n/s_E - 1) + (1+\phi^\beta)})^{\frac{1}{\rho}}$$

(21), **(22)** and **(16)** characterise the equilibrium. Comparing **(22)** to **(12)**, we see that the impact of trade and relocation 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-relocation-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 relocation, 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.
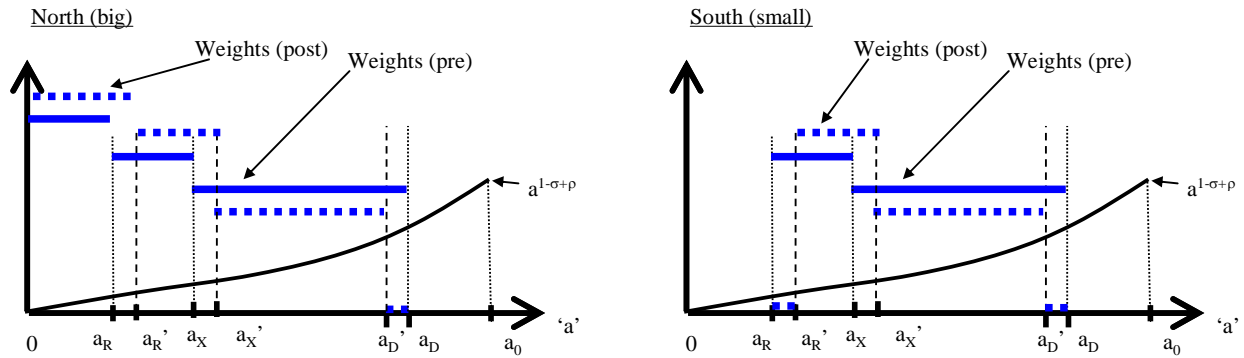

Figure 5: Productivity effects, free relocation case.

---

[15] While (8) 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^w 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: $\bar{F} \equiv F + F\frac{G[a_X]}{G[a_D]} + F_I\frac{1}{G[a_D]}$. 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: $\frac{E^w}{n^w} - \bar{f} = 0$

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 (relocating firms are X-types). For the North, freer trade pivots the density around a_X' and so unambiguously shifts weight to lower values of $a^{1-\sigma+\rho}$. The North's weighted-average 'a' thus 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_R' 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.

In the XX-case, our productivity measure is:

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

where

$$TP* = \int_{a_X}^{a_D} a^{1-\sigma}Bn*dG[a|a_D;a_R] + \int_{a_R}^{a_X} a^{1-\sigma}B(1+\phi)n*dG[a|a_D;a_R], \quad dG[a|a_R;a_D] \equiv \frac{\rho a^{\rho-1}}{a_D^\rho - a_R^\rho}da$$

Solving the integral, we get:

$$\bar{a}* = (\frac{\alpha}{1+\alpha})\frac{(1+\phi)a_X^{1+\alpha} - (1+\phi)a_R^{1+\alpha} + a_D^{1+\alpha} - a_R^{1+\alpha}}{(1+\phi)a_X^\alpha - (1+\phi)a_R^\alpha + a_D^\alpha - a_R^\alpha}$$

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

$$\bar{a}* = \int_{a_R}^{a_D} a\left(\frac{a^{1-\sigma}B}{TP*}\right) n*dG[a|a_D;a_R]; \qquad TP* = \int_{a_R}^{a_D} a^{1-\sigma}Bn*dG[a|a_D;a_R]$$

Solving the integral, we get:

$$\bar{a}* = (\frac{\alpha}{1+\alpha})\frac{a_D^{1+\alpha} - a_R^{1+\alpha}}{a_D^\alpha - a_R^\alpha}$$

Since a_D falls with freer trade and a_R rises, the productivity effect is ambiguous. In particular, using **(20)**, the ratio of cut-offs and **(22)**, we have

(23) $$\bar{a}* = \left((1+\phi^\beta)^{\frac{-1}{\rho}}\right)\left(\frac{1-\left(\phi^\beta s/(1-s)\right)^{1+\frac{1}{1-\sigma+\rho}}}{1-\phi^\beta s/(1-s)}\right)\frac{1-\sigma+\rho}{2-\sigma+\rho}(\frac{f_I(1-\sigma+\rho)}{f(\sigma-1)})$$

where $s = s_n = s_E$ is given by the initial conditions. Note that the first right-hand term is decreasing in ϕ and so tends to imply that freer trade is pro-productivity; the second term however is increasing in ϕ and thus anti-productivity.[16] We can vary the relative importance to the two conflicting terms by varying ρ. When ρ is small (the regularity condition only requires ρ 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 ρ is large, freer trade tends to be anti-

---

[16] Expressing the second term as $(1-x^a)/(1-x)$ which is increasing in x if and only if a>1.

productivity. Heuristically speaking, when ρ is high, then the line $a^{1+\rho-\sigma}$ is very steep and this amplifies the productivity impact of the loss of probability weight on low values of a's that occurs due to a rise in a_R.

For example, in the special case of σ = ρ = 2, the productivity impact of freer trade, $d\bar{a}*/d\phi$, equals $\phi\{s(2+\phi^2)/(1-s) - 1\}/2(1+\phi^2)^{3/2}$, which is positive since s > ½. In other words, freer trade lowers productivity in the small nation in this case.

To summarise, we write

Result 5: (New productivity effects). Allowing relocation introduces a new productivity effect; since the most efficient firms migrate first and freer trade encourages such relocation, 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. During the second phase of migration (see Result 3), freer trade always boosts productivity in the large nation, but may raise or lower productivity in the small nation. If ρ is large enough, freer trade lowers the small nation's productivity.

3.3. *General model*

Modelling the entry/exit and relocation 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 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:

(24)
$$a_R{}^\alpha = \frac{s_n + s_E - 1}{(1-s_n)(1-\phi)}\phi a_X{}^\alpha, \qquad n^w = \frac{(1-1/\beta)/f}{s_n/s_E + \phi^\beta}$$

where the cut-offs are defined as in **(22)**. The standard HFT model resolves the indeterminacy by ignoring relocation, i.e. by assuming a_R = 0. The standard NEG model resolves it by ignoring firm heterogeneity. The relocation-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 relocation 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 relocation 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 relocation and some via entry/exit.

4. SMALL NATION WELFARE

Section 3.2.2 showed that relocation 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 relocation was forbidden from freer trade. This raises the question of whether the small nation might, under certain parameter conditions, benefit from forbidding relocation 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:

$$U* = E - \mu + \mu \ln(\mu \overline{m} *^{\frac{1}{\sigma-1}})$$

Where $\overline{m}*$ is the weighted average of marginal selling costs in the small market, see **(15)**, which covers both imports and locally produced varieties.

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

Since the small-nation domestic cut-off condition, namely $a_D^{1-\sigma} B* = f$, holds in both cases and $B* = (1-s_E)/\overline{m}*$, we have that $\overline{m}* = a_D^{1-\sigma}(1-s_E)/f$. The comparison of $\overline{m}*$'s in the two cases boils down to a comparison of a_D's in the two cases. In particular, $\overline{m}*$ in the no-relocation case exceeds $\overline{m}*$ in the relocation case if and only if, $a_D^{1-\sigma}$ is higher in the no-relocation case. Given our welfare criteria, the no-relocation case is preferred if and only if $a_D^{1-\sigma}$ is higher in the no-relocation case. From **(9)** and **(22)**, the a_D's in the two cases are:

(25) $(a_D^{reloc})^{1-\sigma} = (\dfrac{\beta(s_n/s_E - 1) + 1 + \phi^{\beta}}{f_I(\beta-1)/f})^{\frac{\sigma-1}{\rho}}$, $(a_D^{no\,reloc})^{1-\sigma} = (\dfrac{1 + \phi^{\beta}}{f_I(\beta-1)/f})^{\frac{\sigma-1}{\rho}}$

In both cases, falling trade costs always decrease a_D; this enhances the small nation's welfare as it increases m. Intuitively, the result turns on the fact regardless of relocation, trade liberalisation promotes competition and pushes out the least efficient firms. As a result, trade liberalisation always raises the small nation's welfare.

Result 6: Trade liberalisation always promotes small nation's welfare.

While trade liberalisation benefits the small nation in both cases, the size of the gain will different. By inspection of (26), we see that welfare comparison turns entirely on the issue of whether s_n exceeds s_E when relocation is allowed. Formally, our metric for the welfare effects of relocation can be written as:

(26) $$\dfrac{\overline{m}*^{reloc}}{\overline{m}*^{no\,reloc}} = (1 + \beta \dfrac{s_n/s_E - 1}{1 + \phi^{\beta}})^{\frac{1}{\beta}}$$

where relocation is welfare improving if this ratio exceeds unity, i.e. if $s_n > s_E$ when relocation is allowed.

What determines the relative size of s_n and s_E when relocation is allowed? Section 3 considered two extreme cases: instant entry/exit then relocation, and instantaneous relocation then free entry/exit. In the instantaneous relocation case, relocation 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 **(26)**, we see that if relocation is instantaneous, it has no welfare effects. In other words, the small nation's welfare is the same under instantaneous relocation and under forbidden relocation.

In the general case discussed in Section 3.3, however, neither free entry/exit nor relocation 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 relocation is allowed but

occurs slowly, some of the HME pressure will be relived via relocation 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 relocation is forbidden, by contrast, we know that a_D will be characterised by the second expression in **(25)**. Given **(26)**, this tells us that relocation is always welfare improving for the small nation. To summarise:

Result 7: Allowing relocation in the face of trade liberalisation improves the small nation's welfare more when neither free entry/exit nor relocation is instantaneous.

Both cases of instantaneous relocation and instantaneous entry/exit are perfectly substitutable. Relocation and entry/exit eliminates HME and consequently maintains $s_n = s_E$. In both case, trade liberalisation results in relocation to the North or entry in the North and promotes more imports to the South together with lowering import prices in the South. Consequently it raises welfare. On the other hand, the general case restores HME. We have two effects. First, slow firm relocation can deter relocation and thus the small country can keep more firms than instantaneous case. Second, more entry in the North increases imports and trade liberalisation promotes lower import prices. These two effects enhance welfare in small country, which is higher than in the two instantaneous cases.

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 relocation 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 relocation 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 relocation always mitigates the positive productivity effect, and may, in some circumstances, result in an anti-productivity effect.

The paper also shows that relocation is a complex phenomenon in the presence of heterogeneous firms and fixed market-entry costs. When trade costs are high, the firm-migration (relocation) involves small-nation 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 already moved to the big nation so subsequent relocation involves firms that sell only locally. In this phase, relocation reduces the number of varieties available in the small market.

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

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.E. and T. Okubo (2006a), "Agglomeration, Offshoring and Heterogeneous Firms", CEPR Discussion Paper, 5563.

Baldwin, R.E. and T. Okubo (2006b), Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting, *Journal of Economic Geography* 6, pp.323-346.

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 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. (1999a), 'Exporting and Productivity: Importance of Reallocation', *NBER working paper* 7135.

Bernard, A. and Jensen, B. (1999b), 'Exceptional Exporter Performance: Cause, Effect, or Both?' *Journal of International Economics* 47, 1-26

Bernard A. and Jensen B. (2001), 'Why Some Firms Export', *NBER working paper* 8349.

Bernard, A.B., S. Redding, P. K. Schott (2007), "Comparative Advantage and Heterogeneous Firms," *Review of Economic Studies* 74, pp.31-66.

Chaney, T. (2005) "The Dynamic Impact of Trade Opening: Productivity Overshooting with Heterogeneous Firms", mimeo.

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

Demidova, S. (2005). Productivity Improvements and Falling Trade Costs: Boon or Bane?"Pennsylvania State University, Working Paper Series.

Eaton J, and Kortum S.(2002). "Technology, Geography and Trade", *Econometrica*, Vol. 70.

Eaton, J, S. Kortum, and F.Kramarz (2004), "Dissecting Trade: Firms, Industries, and Export Destinations", *American Economic Review*, 94, 150-154.

Falvey, R, D. Greenaway and Z. Yu (2004). "Intra-industry Trade Between Asymmetric Countries with Heterogeneous Firms," *GEP Research Paper* 04/05.

Fujita M., Krugman P. and A. Venables (1999) *The Spatial Economy: Cities, Regions and International Trade* (Cambridge (Mass.): MIT Press).

Fujita, M. and J.-F. Thisse (2002). *Economics of Agglomeration*, (Cambridge: Cambridge University Press).

Head, K and Ries, J., (2003). Heterogeneity and the FDI versus export decision of Japanese manufacturers. *Journal of the Japanese and International Economies* 17, 448–467.

Helpman, E, M. Melitz and S. Yeaple (2004). "Export versus FDI with heterogeneous firms," *American Economic Review*, 94, 1, pp. 300-317.

Hopenhayn, H (1992a). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica* 60:1127-1150.

Hopenhayn, H (1992b). Exit, Selection, and the Value of Firms, *Journal of Economic Dynamics and Control* 16:621-653.

Krugman, P (1991) Increasing Returns and Economic Geography, *Journal of Political Economy* 99, 483-99.

Martin,P and C.A Rogers (1995) `Industrial location and public infrastructure' *Journal of International Economics* 39: pp335-351.

Melitz, M. (2003). "The impact of trade on intraindustry reallocations and aggregate industry productivity," *Econometrica,* 71, pp 1695-1725.

Melitz, M. and G. Ottaviano (2008). "Market Size, Trade and Productivity", *Review of Economic Studies*, 75 (1), pp.295-316.

Okubo, T (2009). Trade Liberalisation and Agglomeration with Firm Heterogeneity -Forward and Backward Linkages. *Regional Science and Urban Economics* 39 (5), pp.530-541.

Okubo, T, Picard, P.M. and J-F. Thisse (2010). "The Spatial Selection of Heterogeneous Firms", *Journal of International Economics*, 82 (2), pp.230-237.

Redding, S.J. (2010). "Theories of Heterogeneous Firms and Trade", NBER working paper 16562.

Tomiura, E. (2007) "Foreign Outsourcing, Exporting, and FDI: A Productivity Comparison at the Firm Level." *Journal of International Economics*, 72: 113-127.

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.

Yeaple, S. (2005), Firm Heterogeneity, International Trade, and Wages, *Journal of International Economics*, 65(1), pp 1-20.

APPENDIX

To determine the order of relocation, however, we must consider the relocation 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:

(27) $\qquad \cos t = \gamma \dot{x}$

where $\dot{x}$ is the flow of relocating firms and γ is the adjustment cost parameter.

5.1.1   Quadratic adjustment costs and the order of relocation

Atomistic firms observe the actual level of adjustment costs and move if the gain from doing so exceeds the adjustment cost. We observe from (**13**) and (**14**) that if the flow of relocation were zero, then all Southern firms would wish to move instantaneously (since B-B*>0), 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 relocate at the first instant. Once these firms have moved, the flow of relocation would drop from infinite to a finite level, bringing down the relocation cost to a high, but finite level. Firms for whom the value of relocating is less than or equal this finite level will move next. Given (**13**) and (**14**), this implies that the relocation 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 relocation 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 relocation and thereby the level of adjustment costs, so the process occurs smoothly. Ultimately, the process stops when the range of relocated firms rises to the point where the two markets are equally attractive, i.e. B=B*. The range of firms that relocate in response to any given level of trade freeness is [0,a_R], where a_R (the 'R' is a mnemonic for relocate) is the threshold level of marginal cost defined by what we call the 'relocation condition':

(28) $$B = B *$$

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

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

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 a_D are fixed by the initial conditions described above).[17]

As mentioned, relocation alters average competitiveness of each market, so the definition of the B's in the relocation condition, (**15**), is more complicated than the definition the Section 3.1. Specifically for near-zero levels of trade freeness, B=s_E/n^w $\overline{m}$ where $\overline{m}$ is now defined to take account of the relocation of firms with a's in the [0,a_R] range:

(29) $$\overline{m} = s_E \int_0^{a_D} a^{1-\sigma} dG[a|a_D] + (1 - s_E) \int_0^{a_R} a^{1-\sigma} dG[a|a_D] + (1 - s_E) \phi \int_{a_R}^{a_X} a^{1-\sigma} dG[a|a_D]$$

where we have set s_n=s_E by the initial conditions. 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. B* equals (1-s_E)/n^w $\overline{m}$ *, where the expression for $\overline{m}$ * is similar to the above expression for $\overline{m}$.

The relocation condition B=B* can be solved analytically to get a_R as a function of a_X:

(30) $$a_R = \left(\frac{\phi(2s_E - 1)}{(1 - \phi)(1 - s_E)} \right)^{\frac{1}{1 - \sigma + \rho}} a_X$$

Using the definition of B with $\overline{m}$ in (**15**) and a_R as defined in

(17), the export cut-off condition implicitly defines a_X as:

(31) $$\frac{\phi a_X^{1-\sigma}}{\phi a_X^{1-\sigma+\rho} + ((\beta - 1) f_I f)^{1-1/\beta}} = \frac{f / f_I}{\beta - 1}$$

This condition cannot be solved analytically since σ and ρ 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.

---

[17] 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 $\int_o^{a'} n * dG[a|a_D]$, which solves to $(n * (a')^{\rho}) / a_D^{\rho}$; taking the time derivative yields the result in the text.