Trade Liberalisation and Agglomeration with Firm Heterogeneity----Forward and Backward Linkages

Toshihiro Okubo
University of Manchester

May 2007

ABSTRACT

This paper studies the impact of trade costs reduction on geographical manufacturing concentration in the presence of intermediate inputs linkage, firm heterogeneity and export fixed costs. Firm heterogeneity with export fixed costs hampers full agglomeration through weakening forward and backward linkages and fortifying market crowding effects. Rather than catastrophic agglomeration that the standard new economic geography models such as the Core-Periphery Model of Krugman (1991) have long suggested, trade liberalisation causes gradual agglomeration. Also, trade liberalisation produces a divergence in welfare for the periphery, which loses, and the core, which gains. Even free trade never equalises welfare between core and periphery, i.e. trade liberalisation does not eliminate inequality among nations due to the fact that the periphery’s needs cannot all be sourced from the core.

JEL F15, F23.

Keywords: heterogeneous firms, economic geography, spatial sorting effect, non-full agglomeration, non-catastrophic agglomeration, trade liberalisation.

1. INTRODUCTION

The last few decades have witnessed a remarkably large reduction of trade costs and trade barriers through trade liberalisation and also a high degree of spatial concentration in industry. Krugman and Venables (1995) suggested that trade liberalisation first creates core-periphery structure and divergence in real income across nations but finally could at even lower trade costs produce convergence in real income over the world. This tells us that trade liberalisation ultimately leads to the elimination of inequality between nations.

However, this prediction may not be secure. Regardless of trade liberalisation, the income gap between the richest and poorest countries in the world is still vast. Per-capita income convergence can be found only among the OECD countries and divergence can be found within the developing world. UNCTAD (1997) said, “since the early 1980s the world economy has been characterized by rising inequality and slow growth. Income gaps between North and South have continued to widen. In 1965, the average per capita income of the G7 countries was 20 times that of the world's poorest seven countries. By 1995 it was 39 times as much.” In other words, as Baldwin and Martin (1999) and Jones (1997) suggested, in the process of trade liberalisation, the post-war globalisation has a twin-peaks economic convergence with the rich nations getting richer and the poor nations remaining poor. In detail, some developing countries...
fall farther behind due to trade liberalisation (Dollar and Kraay, 2004). Along with these evidences, one often sees small developing nations opposing trade and trade-related liberalisations and the liberalisation process reaching a deadlock. The trade liberalisation in the real world seems to involve much more pessimistic outcomes and even more bitter struggles than we have thought.

This paper studies how trade liberalisation affects industrial location and produces divergent trends in real income and welfare in the world, taking into account economic geography. To demonstrate this, we incorporate the vertical linkage (VL) model of Krugman and Venables (1995) in NEG literature with the heterogeneous-firms trade model of Melitz (2003) and study the impact of trade and technical barriers to trade (TBTs) liberalisation on firm relocation and agglomeration and on welfare.

1.1. Literature and Our Paper

**Liberalisation and Inequality**

Recent heterogeneous firms trade models have sought to demonstrate how trade liberalisation leads to effects that vary across firms (Melitz, 2003). Trade liberalisation raises only exporters’ profits but reduces local producers’ profits (share shifting effect) and forces the exit of the least efficient local producers from the market, accommodating the most efficient local firms into the export market (selection effect). Shifting our attention to international inequality, Baldwin (1999) found divergence in real income due to capital accumulation in neo-classical growth model, and Baldwin, Martin and Ottaviano (2001) found that the difference of technological spillovers matters for divergence. Furthermore, Krugman and Venables (1995) modelled vertical linkage of intermediate inputs as peculiar externalities and suggested divergence in real income between core and periphery at intermediate trade costs and finally convergence at still lower trade costs. Integrating both of international and intra-national inequality literatures, this paper studies the inequality of nations in industrial location due to trade liberalisation. The New Economic Geography model of Krugman and Venables (1995) is incorporated into the heterogeneous firms trade model of Melitz (2003).

**Core-periphery Structure and Vertical Linkage in New Economic Geography**

The New Economic Geography (NEG) literature, our viewpoint, has seen the dramatic development over decade, whose main undertaking was to endogenise the choice of industrial location. The most familiar NEG models are the Core-Periphery (CP) model by Krugman (1991) and the vertical linkage (VL) model by Venables (1994, 1996) and Krugman and Venables (1995) (See also Fujita, Krugman and Venables, 1999; Baldwin, Forsslid, Martin, Ottaviano and Robert-Nicoud, 2003; Fujita and Thisse, 2002). The agglomeration is generated by labour migration in the CP model and on the other hand by input-output linkage in the VL model. Importantly as a consequence both CP and VL models are isomorphic in equilibrium properties. The most important forces of work in them are backward and forward linkages as agglomeration force and market crowding effect as dispersion force. Due to these working forces, a symmetric two nations model result in manufacturing concentration on one nation as a consequence of trade liberalisation (endogenous asymmetry), which occurs suddenly (catastrophic agglomeration).

---

One of the current research avenues in NEG is to overcome the intractability of the CP and VL models and check the robustness of the existing NEG models. One of the common difficulties in VL and CP models is their tendency to produce analytically unsolvable equilibria. To tackle this difficulty and to check the robustness of the VL model, the footloose entrepreneur (FE) model of Forslid and Ottaviano (2003), the footloose capital vertical linkage (FCVL) model of Robert-Nicoud (2004, Appendix 2) and Baldwin et al. (2003, Ch.8) and the footloose entrepreneur vertical linkage (FEVL) model of Ottaviano and Robert-Nicoud (2006) have been successful in providing more tractable and substitutable theoretical frameworks. They demonstrate that all of these models are isomorphic with the standard CP and VL models in equilibrium features without losing key properties. In particular, the FCVL model proposed by Robert-Nicoud (2004) is the most tractable and simplest in a family of VL models, in which VL model marries with footloose capital (FC) model of Martin and Rogers (1995). The FCVL model does not depart from the CP and VL models: it has backward and forward linkages and market crowding effect with circular causality and has the same equilibrium path as they do, along with endogenous asymmetry and catastrophic agglomeration.

**Firm Heterogeneity and New Economic Geography**

Indeed, the VL model has reflected the important phenomena of the vertical input-output linkage in creating industrial agglomeration, but the NEG models did not yet capture one important aspect revealed by a number of remarkable recent empirical studies, i.e. firm heterogeneity. The current econometric analysis in international trade and industrial organisation literature sheds light on different impacts of trade liberalisation across firms. Along with the development of econometric technique and the improvement of firm level data sets, a good deal of empirical evidence has been provided (Aw et al. 2000; Bernard and Jenson 1995, 1999; Bernard, Jenson and Schott 2003; Pavcnik, 2002; Tybout and Westbrook, 1995). The evidence has highlighted how trade liberalisation differently affects productivity across firms. As a response to this empirical evidence, Melitz (2003) has demonstrated that trade liberalisation raises industry productivity via a selection effect, as well as via a production reallocation effect (Helpman, Melitz and Yeaple, 2004; Melitz and Ottaviano, 2005; Falvey, Greenaway and Yu, 2004; Baldwin and Robert-Nicoud, 2004; Baldwin and Forslid, 2004). We call these studies heterogeneous-firms trade (HFT) models. Although the HFT models’ main focus has been the trade liberalisation driven selection effect through entry and exit processes and the change in firm behaviour as well as in their productivity, these above-mentioned previous theoretical studies were not directly linked to the NEG literature: They did not perfectly deal with the issues of economic geography, namely firm migration, agglomeration/ dispersion forces, and equilibrium stability, in spite of provocative intentions of delocation in Europe and offshoring in North America.

To analyse these issues, Baldwin and Okubo (2006a) incorporated the Melitz literature into the footloose capital (FC) model of Martin and Rogers (1995) in the NEG literature, which is in the line of this paper. To highlight firm location and economic geography rather than productivity, they assumed away free entry and exit and instead modelled interregional firm relocation. They found that the presence of marginal cost discrepancies results in qualitatively different characteristics from the existing NEG literature, in which firm heterogeneity dampens the agglomeration process in terms of the number of firms. However, the model does not exhibit

---

2 The FEVL and FCVL models have the feature of agglomeration without labour migration, while sustaining all of the equilibrium features of the VL model (See Baldwin et al. 2003).

3 Baldwin and Okubo (2006a) suggested spatial selection and sorting mechanisms different from the selection effect in Melitz-type models (heterogeneous firms trade model). A positive delocation cost allows the most efficient firms in the periphery to be more attracted towards the core region, i.e. there is a spatial selection effect. On the other hand, delocation subsidies would induce efficient firms to move to the core and inefficient firms to move to the periphery via a sorting effect.
the full features of NEG models because it depends on only the simplest NEG model of Martin and Rogers (1995). Baldwin and Okubo (2006a) as well as the FC model lacks many key features of the other NEG models: 1) no forward and backward linkages nor circular causality 2) no endogenous asymmetry, nor catastrophic agglomeration, in which symmetric countries with symmetric trade costs reduction never lead to agglomeration of manufacturing in one country and 3) no overlap and self-fulfilling expectations. In their models, agglomeration in one country is created by (exogenously given) asymmetric market size without self-reinforcing. Trade liberalisation gradually creates agglomeration (gradual agglomeration). These shortcomings inherent of the FC model definitely degenerate from the CP and VL models.

By contrast, this paper uses FCVL model with symmetric size of two countries to sustain the most important features in the CP and VL models. The FCVL model involves the most important key features, which are missing in Baldwin and Okubo (2006a): forward and backward linkages and market crowding effect with circular causality. Thus, our model can discuss whether agglomeration could occur in symmetric-region framework with self-reinforcing and circular causality triggered by intermediate demand with firm heterogeneity.

The basic value added in the present paper lies in showing that 1) Firm heterogeneity per se cannot alter the main equilibrium features the VL models, 2) Firm heterogeneity in the presence of export fixed costs a la Melitz (2003), by contrast, entirely changes the equilibrium features in the standard NEG model. Unlike CP and VL (including FCVL and FEVL) models, no catastrophic agglomeration happens and instead gradual agglomeration emerges as trade gets freer in spite of backward and forward linkages. There is the possibility of non-full agglomeration and 3) Trade and TBTs liberalisations lead to divergence in welfare. Even at free trade, defined as zero iceberg trade costs, the international welfare never converges due to the presence of local producers. Hence, all firms might prefer to locate in the core even with free trade and the inequality of nations dissolves.

Plan of Paper
Reflecting the development of the NEG models and HFT models, this paper is to promote the development of the VL models. We aim at examining the impact of firm heterogeneity on the VL model and discuss equilibrium features. The rest of the paper is organised in 4 sections. Section 2 presents the basic model, Section 3 examines equilibrium in the absence of export fixed costs. Section 4 explores equilibrium and its features in the presence of export fixed costs. Then Section 5 presents the impact of trade and TBTs liberalisation on welfare. Finally, concluding remarks are provided in Section 6.

2. THE BASIC MODEL
First, we introduce one application of the Vertical Linkages model, known as the footloose capital vertical linkage (FCVL) model, (Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud, 2003, Chapter 8 for a thorough analysis of this model and Robert-Nicoud, 2004, Appendix 2). Then, we add an extra assumption to generate firm heterogeneity à la Melitz (2003).

2.1. Footloose Capital Vertical Linkage model
To build intuition, we start from the standard FCVL model of Robert-Nicoud (2004), which is the marriage of the footloose capital (FC) model of Martin and Rogers (1995) with the Vertical Linkage (VL) model. The FCVL model works with the same setup as the simplest FC model, i.e. two nations (the North and the South), two sectors (manufacturing and agriculture) and two factors (Labour and Capital). However, different from the standard FC model, two symmetric
nations are referred to as initial equilibrium, and thus two market sizes are initially identical. The two nations are symmetric in terms of tastes, technology, openness to trade, size, and relative factor endowments of labour and capital.\textsuperscript{4} Thus the two nations have the same share of world labour ($s_{l}=0.5$) and capital owners ($s_{k}=0.5$), initially equal to the share of manufacturing ‘$s_{n}$’(=0.5). The two sectors will be referred to as manufacturing (M) a monopolistic competition sector and agriculture (A) as numeraire sector. Manufacturing requires intermediate input demand in input-output linkage as in the original VL model, sustaining the richness of the model: forward and backward linkages and market crowding effect.

The standard Dixit-Stiglitz monopolistic competition model assumes a two tier utility function with the upper tier being the Cobb-Douglas function and the lower tier for manufactured goods being the CES function. Thus we assume:

\[ U = C_{M}^{\sigma} C_{A}^{1-\mu}, \quad C_{M} = \left( \int \Theta^{1-\sigma} \, d\Theta \right)^{1/(1-\sigma),} \quad 0 < \mu < 1 < \sigma \]

where $C_{M}$ and $C_{A}$ are, respectively, consumption of the M-sector varieties composite and consumption of the A-sector good, $\sigma$ is the constant elasticity of substitution between any two M-sector varieties, $\mu$ measures the share of M-sector varieties in total consumption and $\Theta$ is the set of varieties available in one particular nation.

The numeraire sector, agriculture, is characterised by perfect competition, constant returns to scale and zero trade costs. Manufacturing is marked by increasing returns, Dixit-Stiglitz monopolistic competition and iceberg trade costs. Unlike the FC model, the FCVL model supposes intermediate goods are required to produce M-good. As in the original VL model by Krugman and Venables (1995), manufactured goods are used as intermediate goods as well as consumer goods (final). Each manufacturing firm requires one unit of K as fixed costs and ‘a’ units of L per unit of output and intermediate goods (all varieties of manufacturing goods) as variable costs. In the variable cost part, labour and intermediate goods are used by the Cobb-Douglas cost function. For simplicity, we assume $\mu$ to be the share of intermediate goods in the variable cost function. Therefore a typical firm’s cost function can be represented as:

\[ \pi_{i} + a_{i}x_{i}P_{p}, \quad P_{p} = w^{1-\mu} \bar{P}^{1-\sigma} \]

where $\pi$ are $w$ are capital’s and labour’s reward, $x$ is firm-level output and $P_{p}$ refers to the producer price of the Cobb-Douglas aggregate of labour and the bundle of intermediate goods. The second term represents variable cost; i.e. the payment for labour and intermediate goods.\textsuperscript{5} \(\bar{P}\) is the (weighted) average of varieties prices sold in the North, as defined below.

As in the standard FC and FCVL models, all capital is owned by capital owners and is mobile between nations, but capital owners and labour are inter-nationally immobile. Thus, capital rewards are repatriated across nations. Note that labour and capital owners are equally distributed, i.e. $s_{l}=s_{k}=0.5$ is fixed regardless of firm location at any equilibrium. Instead, the Northern firm share, $s_{n}$, is changeable to reflect firm relocation, viz capital movement across nations.

\subsection*{2.2. Intermediate Results}

The presence of the numeraire sector helps to simplify the results: constant returns, perfect competition and zero trade costs equalise nominal wage rates across nations ($w=w^{*}=1$). What

\textsuperscript{4} In the FC model, the symmetric two nations are in stable equilibrium for any positive level of trade costs and thus full agglomeration never occurs, unlike in the FCVL model.

\textsuperscript{5} The cost function of a typical manufacturing firm in the FCVL model is non-homothetic; the fixed cost involves only capital and the variable cost involves not only labour but also intermediate goods.
this means is 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 subject to all the usual Dixit-Stiglitz results, where prices of good M 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 includes the iceberg costs, $t \geq 1$, associated with the constant Dixit-Stiglitz mark-up. As a result of utility maximisation, we can work out a demand function for each variety of good M as

$$c_i = \frac{(p_j)^{\sigma/\mu}}{\overline{p}} E; \quad \overline{p} = \int_{j \in \Theta} p_j^{1-\sigma} dj + \int_{h \in \Theta^*} \phi p_h^{1-\sigma} dh, \quad \sigma > 1 \geq \phi \equiv t^{1-\sigma} \geq 0$$

where $p_j$ is variety-i’s producer price (which equals its consumer price since it is produced locally), and $\overline{p}$ is the (weighted) average of consumer prices of varieties sold in the North. $E$ denotes total expenditure in the North. The first term in the definition of $\overline{p}$ represents the prices of goods that are produced in the North (and so bear no iceberg trade costs). The second term stands for the imported varieties whose producer prices are $p_h$; $\Theta$ and $\Theta^*$ refer to sets of consumed goods that are produced in the North and the South, respectively. Note that the geometric weights are negative due to the regularity condition that $\sigma > 1$, and thus the weighted average price, $\overline{p}$, falls as individual prices rise. A critical parameter in our paper is $\phi$, namely the freeness of trade. Note that $\phi$ ranges from zero when trade is prohibitive in autarchy ($t=\infty$) to unity when trade is perfectly free ($t=0$). Southern demand functions are isomorphic.

Utility maximisation generates the familiar CES demand functions. These, together with the standard Dixit-Stiglitz monopolistic competition assumptions on market structure imply ‘mill pricing’ optimality and this in turn implies a constant operating profit margin, $1/\sigma$, where $\sigma > 1$ is the constant elasticity of substitution among varieties. Thus the operating profit that a firm earns – denoted as $\pi$ – equals the value of the firm’s sales times $1/\sigma$. The Dixit-Stiglitz market share is a simple function of the firm’s price relative to the average price of all its competitors. Firm i’s prices in its local market and export market can be written as

$$p_j = \frac{a_i p_{p_j}}{1-1/\sigma}, \quad p_j^* = \frac{1a_i p_{p_j}}{1-1/\sigma}$$

where $t \geq 1$ represents the usual iceberg trade costs and $t$ units must be shipped in order to sell one unit in the export market (i.e. ‘$a$’ for local sales and ‘ta’ for export sales). Next, we characterise the northern expenditure on manufactures:

$$\mu E = \mu \left( Y + \int_{i \in \theta} P_x a_i x_i \right)$$

where $Y$ is total income of northern workers and capital owners, $Y = s_L L^W + s_K K^W$, ($s_L$ and $s_K$ are equal to 0.5), and the second term represents domestic intermediate demand. The northern total income $Y$ is composed of workers’ income and capital owners’ rewards. All capital owners in both nations are assumed to equally receive as average operating profits, $\pi$, regardless of country of residence. We assume that capital owners supply their own capital endowments (one unit per capita) and hold an equal amount of stocks in all firms in the world so as to share and equally split total profits of all firms in the world. The half of total capital rewards from all firms in the North belongs to northern owners and the rest belongs to southern owners,

---

6 Demand for a typical variety $j$ is $c(j) = p(j)^{-\sigma} \mu/\Delta$, where $\Delta = \int p(j)^{1-\sigma} dj$ and the integral is over all available varieties, while $\mu$ is expenditure on all varieties as usual.

7 A typical first order condition is $p(1-1/\sigma) = wa$; rearranging, the operating profit, $(p-wa) c$, equals $pc/\sigma$. 
and vice versa. As a result, the capital owners’ rewards between nations are equal. For simplicity’s sake, $E^w (=E+E^*)$ can be normalised to one by assuming $L^w=1-\mu$: $E^w = \frac{L^w}{1-\mu} = 1$

### 2.3. Firm Heterogeneity and Export Fixed Costs

Then, we add firm heterogeneity à la Melitz (2003) to the FCVL model. Here, our model assumes away free entry and exit of Melitz (2003), instead of allowing for firm relocation (international capital movement). Therefore, without free entry and exit, we employ marginal costs difference across firms and one type of beachhead cost, i.e. export fixed costs. Due to no free entry and exit, the mass of varieties is exogenously given as $n$ and $n^* (=0.5)$, which is determined by initial symmetric capital endowments. We allow firms to have different $a$’s and thus different marginal production costs subject to the symmetric probability distribution function. Since each firm is associated with a particular unit of capital, it is natural to assign the source of heterogeneity to capital, i.e. each unit of capital in each nation is associated with a particular level of marginal cost as measured by the unit labour requirement, ‘$a$’. The ‘$a$’s are assumed to be distributed by a Pareto probability distribution, whose cumulative density function can be defined as:

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

where is $\rho$ and $a_0$ respectively stand for the ‘shape’ and ‘scale’ parameters; without loss of generality, we choose units such that $a_0=1$. (See Figure 1)

---

**Figure 1: “$a$” distribution and firm types at initial equilibrium.**

---

8 This definitely stems from the finding of Baldwin and Okubo (2006b) that firm free relocation and free entry/exit are not compatible in specifying equilibrium solution. The equilibrium is not uniquely determined when firms are internationally mobile allowing for free entry and exit. For this reason, the standard HFT models except Baldwin and Okubo (2006a), ruling out firm relocation, assume that all firms are required to produce the nation in which they are born.
Next, we look into one beachhead cost more in depth. This fixed export market-entry costs, beachhead cost, represent the cost of exporting and selling the variety in the market so as to satisfy the market-specific standards and regulations. The export cost, denoted as $F$, is the same across nations. This is overhead-type fixed cost. Relying on the standard logic of beachhead costs, we can already anticipate that only firms with sufficiently low marginal costs will enjoy sales that are high enough to justify the market-entry costs. The beachhead cost, $F$, creates one cut-off level for marginal costs, labelled $a_X$ (Figure 1). The export fixed costs involve labour with its equal units in dispatched country and destination country to keep symmetry of income and factor endowments. Firms that have $a_i$'s above $a_X$ will not find it worthwhile to sell in their export market. Under regularity conditions discussed below, the cut-off point identifies two types of firms: D firms that produce and sell only locally (D-types, short for domestic firms); firms that sell locally and also export (X-types, short for export firms).

2.4. Initial Symmetric Equilibrium

We start with symmetric equilibrium in autarchy and then trade costs reduce gradually. $\Delta$ is a mnemonic for denominator and it is related to the CES price index $P = \frac{\mu}{\mu} = (\Delta)^{\mu(1-\sigma)/(1-1/\sigma)}$. Here, $\Delta$ and $\Delta^*$ at the initial symmetric equilibrium (sE=0.5), proportional to the denominator of a standard CES demand function, can be defined as:

$$
\Delta = n\Delta^\mu \int_0^\infty a^{1-\sigma} dG(a) + n^* \phi \Delta^\mu \int_0^\infty a^{1-\sigma} dG(a)
$$

where the mass of varieties, $n$ and $n^*$, are given as 0.5 and * denotes the South. Note that these equations are recursive in $\Delta$ and $\Delta^*$.

2.5. Location Choice and Order of Delocation

Each firm can choose its location in search of the highest pure-profit, i.e. capital rewards. Since firm productivity is dissimilar, the location decisions are idiosyncratic across firms. The pure profits by X type and D type firms in each country can be written as

$$
\pi_X[a] = b B_X a^{1-\sigma} - F \quad \text{and} \quad \pi_X^*[a] = b B_X^* a^{1-\sigma} - F
$$

where $b \equiv \frac{\mu}{\sigma}$, $B_X \equiv \Delta^\mu \left( s_E + \phi \frac{1-s_E}{\Delta^*} \right)$, $B_X^* \equiv \Delta^* \left( \phi \frac{s_E}{\Delta} + \frac{1-s_E}{\Delta^*} \right)$ and $s_E \equiv \frac{E}{E^\mu}$.

$$
\pi_D[a] = b B_D a^{1-\sigma} \quad \text{and} \quad \pi_D^*[a] = b B_D^* a^{1-\sigma}
$$

where $B_D \equiv \frac{s_E}{\Delta^{1-\mu}}$, $B_D^* \equiv \frac{1-s_E}{\Delta^{n-\mu}}$ and $s_E \equiv \frac{E}{E^\mu}$.

Profits are a function of the expenditure ratio, $s_E$, and the marginal costs, $a$. Small marginal costs (small ‘$a$’s) types, i.e. high productivity firms, can make more pure profits.

Cut-off conditions

A firm exports only if its operating profit exceeds a given level of the beachhead cost, $F$. The cut-off levels of marginal cost $a_X$ in Figure 1 are determined by:

---

9 The income from the engagement for export fixed costs is equal between nations even if all firms concentrate in one.
where \( a^*_X \) and \( a_X \) are the cut-off marginal costs for entering export markets, respectively, \( s_E \) (or \((1- s_E)\)) is expenditure in the North (or South).

**Two phases of the firm migration process**

At an initial symmetric equilibrium, profits for all firms should be equal between the two nations, caused by the symmetric size of demand \((s_E=0.5)\), the equal size of factor endowments (labour and capital) and the symmetry in firm distribution, bringing about the identical value of the \( \Delta s \), \((7)\). But the symmetric equilibrium is not always stable. This is the same conjecture as the standard NEG model. Below a critical level of trade costs (above the break point in freeness of trade), once a single firm migrates from one to the other nation (from the South to the North in this paper for convenience), the \( \pi[a]-\pi^*[a] \) becomes positive.\(^{10}\) As seen in \((8)\) and \((9)\) the profit gap can be divided by two part expression: market potential gap, \( B-B^* \), and productivity terms, \( a^{1-\sigma} \), i.e. \( \pi(a) - \pi^*(a) = b \left( B-B^* \right) a^{1-\sigma} \). Since the productivity part is exogenously given, relocation solely alters the market potential part with a smaller “\( a \)”, higher productivity firms are likely to receive a larger impact from market potential gap due to \( a^{1-\sigma} \).

When a single firm deviates from the North to the South, the gap in the pure profits of the deviated firm, \( \tilde{a} \), is represented as a function of each firm’s productivity and its type:

\[
\begin{align*}
\nu_X[\tilde{a}] &= \pi_X[\tilde{a}] - \pi^*_X[\tilde{a}] = b(B_X - B^*_X)\tilde{a}^{1-\sigma} \geq 0; \quad 0 \leq \tilde{a} \leq a_X \\
\nu_D[\tilde{a}] &= \pi_D[\tilde{a}] - \pi^*_D[\tilde{a}] = b(B_D - B^*_D)\tilde{a}^{1-\sigma} \geq 0; \quad a_D \leq \tilde{a} \leq 1
\end{align*}
\]

As in the standard VL models and CP model, the solutions are also analytically unsolvable due to the recursive forms of \( \Delta \) and \( \Delta^* \) in \( B_X, B^*_X, B_D \) and \( B^*_D \) as shown below.\(^{11}\) For this reason, we employ numerical solution following the standard NEG technique. As a result of numerical simulation, Figure 2 graphs \((11)\) which shows that when trade costs are small enough to become unstable symmetric equilibrium, the firms that have the most to gain from relocation are the most efficient firms, thus it is the most efficient firms that will tend to move first (See Appendix 2 for the mathematical proof). This result is intuitively obvious, since the home market effect (HME) is driven by firms’ 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 marginally larger market (the North). Note that by the definition of \( a_X \), firms with \( \tilde{a}=a_X \) are just indifferent to exporting so \( \nu_X[a] \) and \( \nu_D[a] \) touch at \( a_X \) as shown in Figure 2.

---

\(^{10}\) Conversely, even if the delocation from the North to the South, all of the qualitative results can be kept as the same shown in this paper.

\(^{11}\) (12), (13), (14) and (15) are used for plotting of (11) shown as Figure 2.
Figure 2: Relocation gains.

**Result 1:** The symmetric equilibrium becomes unstable at sufficiently low iceberg trade costs; the first firms to mover are the efficient export firms. Small local firms have less incentive to move to the other nation and so do not move at first.

When the symmetric equilibrium is unstable due to small trade costs (above the break point in freeness of trade), the most efficient Southern firms have the most to gain from moving and thus move first. The delocation *per se*, however, raises the degree of competition in the North whilst lowering than in the South, but additionally raises $s_E$ through the intermediate input linkage in our model, which cannot arise in the FC model of Baldwin and Okubo (2006a). Therefore, although the value of $v[a]$ to any given firms of moving declines as the range of firms that have moved expands, the decline is slower than Baldwin and Okubo (2006a) due to the intermediate demand towards the North. The process continues until the gain from moving is zero (i.e. firms move until all firms are indifferent to location which happens when $v[a]=0$).

Inspection of (11) reveals that $\bar{a}$ starts out at zero but gets progressively closer to $a_\lambda$ as relocation goes on the small trade costs (above the break point, $\phi^{\text{break}}$). This implies that the migrating firms are initially X-types. Given $\phi$, the production cost linkage through delocation cumulatively proceeds firm relocation. $\Delta$ and $\Delta^*$ in the X-type relocation phase can be written as

$$
\Delta = n\Delta^* \int_0^a a^{1-\sigma} dG(a) + n^* \phi \Delta^* \int_{aX}^{a_\lambda} a^{1-\sigma} dG(a) + n^* \Delta^* \int_0^{a_\lambda} a^{1-\sigma} dG(a)
$$

$$
\Delta^* = n^* \Delta^{\ast\ast} \int_0^a a^{1-\sigma} dG(a) + n \phi \Delta^{\ast\ast} \int_{aX}^{a_\lambda} a^{1-\sigma} dG(a) + n^* \phi \Delta^{\ast\ast} \int_0^{a_\lambda} a^{1-\sigma}
$$

Note that both are recursive and thus cannot be solved as before, solving to $\Delta = 0.5\bar{\lambda}(\Delta^* (1 + \bar{a}^\alpha) + \phi \Delta^{\ast\ast} (a_\lambda^\alpha - \bar{a}^\alpha))$ and $\Delta^* = 0.5\bar{\lambda}(\phi \Delta^* (a_\lambda^\alpha + \bar{a}^\alpha) + \Delta^{\ast\ast} (1 - \bar{a}^\alpha))$ where $\alpha \equiv 1 - \sigma + \rho$ and $\bar{\lambda} \equiv \rho/(1 - \sigma + \rho)$. Note that $1 - \sigma + \rho > 0$ is a regularity condition, which ensures the integral convergence.

Using Result 1, the share of expenditure in this phase, $\bar{a} < a_\lambda$, can be written as
where $b \equiv \mu / \sigma$ and $\tilde{\alpha}$ refers to the cut-off level of efficiency. The first term is northern workers’ income ratio, the second term represents northern intermediate input demand ratio. Note that since we assume that the southern firms with marginal costs from 0 to $\tilde{\alpha}$ relocate to the North, $s_E > 0.5$ can always arise.

However, all southern X-types have moved to the North, as seen in Figure 2, the profit gap is still positive, i.e. $v(\alpha_X) > 0$. Accordingly, a second phase of relocation begins where all the migrating firms are D-types. (We refer to these two phases of migration as X-type relocation and D-type relocation.) After $\tilde{\alpha}$ equals $\alpha_X$, migration affects the $\Delta$ and $\Delta^*$ differently because the south no longer exports manufactures to the North and the migrating firms no longer sell back into the southern market. Now the $\Delta$ and $\Delta^*$ can be expressed as:

$$\Delta = n \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a) + n^* \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a)$$

$$\Delta^* = n^* \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a) + (n + n^*) \phi \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a)$$

which solve to $\Delta = 0.5 \lambda \Delta^{\mu} (1 + \tilde{\alpha}^{\alpha})$ and $\Delta^* = 0.5 \lambda [\Delta^{\mu} (1 - \tilde{\alpha}^{\alpha}) + 2 \phi \Delta^{\mu} \tilde{\alpha}^{\alpha}]$. The share of expenditure can be given as,

$$s_e = \frac{1 - \mu + b}{2} + \frac{(\sigma - 1) \int \left( n \int_0^{\tilde{\alpha}} B_x a^{1-\sigma} dG(a) + n \int_0^{\tilde{\alpha}} B_\mu a^{1-\sigma} dG(a) + n^* \int_0^{\tilde{\alpha}} B_x a^{1-\sigma} dG(a) + n^* \int_0^{\tilde{\alpha}} B_\mu a^{1-\sigma} dG(a) \right)}{\left( n \int_0^{\tilde{\alpha}} B_x a^{1-\sigma} dG(a) + n \int_0^{\tilde{\alpha}} B_\mu a^{1-\sigma} dG(a) + n^* \int_0^{\tilde{\alpha}} B_x a^{1-\sigma} dG(a) + n^* \int_0^{\tilde{\alpha}} B_\mu a^{1-\sigma} dG(a) \right)}$$

As shown in Figure 2, the relocation continues to raise $\tilde{\alpha}$ gradually up to $B_D = B_D^*$, in which the equilibrium relocation cutoff is determined as $\tilde{\alpha} = \alpha_R$ (‘R’ is mnemonic for relocation).

Note that the slope of X type relocation diagram in Figure 2 is flatter than the D-type relocation diagram and never cut the horizontal axes, $v(\alpha_X) > 0$ for any ‘$\alpha$’s. The steeper slope of D-type phase diagram implies that relocated D type firms cannot sell the original market by relocation, and thus much more sensitive to the gap of market size and its market crowding effect, which is more severe than X-type firms. The decline of the gain from relocation is faster than X type firms. By contrast, since relocated X firms can sell the original market, exporting expanded by trade liberalisation mitigates the market crowding effect in their own location for X types. Thus, since the forward and backward linkages are always stronger than the market crowding effect for X type, all X-type firms move to one nation or the other above the break point.\footnote{Compared to Baldwin and Okubo (2006b), the X type relocation phase looks flatter. The intermediate input linkage increases $s_e$ in firm delocation. This relatively increased market size through delocation makes the decline of $v(\alpha)$ slower in terms of ‘$\alpha$’ and thus more southern firms can move to the North. If $\mu$ is close to 0 and intermediate input linkage becomes weaker, X-type relocation phase shift down and has steeper slope, and closer to the one in Baldwin and Okubo’s heterogeneous firms FC model. Note that $\mu$ should be still kept as positive: $\mu > 0$ means not only no manufacturing intermediate input demand but also no consumption of the M goods as final goods, and thus the discussion would be invalid. Note that the relocation cutoff $\alpha_R$ in the FC model with firm heterogeneity of Baldwin and Okubo (2006b) can be either D type or X type firms in two phase relocation. Since the FC model has neither demand-linked nor cost-linked circular causality, agglomeration forces are weaker than our model and thus the equilibrium yields in X-type relocation phase as well as in D-type phase.}

\begin{align}
\Delta &= n \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a) + n^* \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a) \\
\Delta^* &= n^* \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a) + (n + n^*) \phi \Delta^{\mu} \int_0^{\tilde{\alpha}} a^{1-\sigma} dG(a)
\end{align}
To summarise:

Result 2: (Two phases of firm delocation). Once trade costs are small (above the break point in freeness of trade), relocation can begin. This is marked by two phases. Migration of Southern X-types that remain X-types after their relocation, takes place first. In the second phase, which starts once all X-type firms have left the small region, there is a migration of Southern D-type firms that remain D-types in their new location. The cost linkage through intermediate inputs can promote firm relocation in our model in a sense of its causing the North to attract more firms than in the FC model in the presence of firm heterogeneity. As a consequence, the equilibrium cutoff level, $a_R$, is always determined in the second relocation phase (D-type relocation phase).

3. THE IMPACT OF FIRM HETEROGENEITY ON THE LOCATION EQUILIBRIUM IN THE SIMPLEST MODEL

To analyse the impact of firm heterogeneity per se on the VL model equilibrium features, we first simplify the model, stepping back to exclude the export fixed costs ($F=0$) case and thus no D-type firms ($a_X=1$). All heterogeneous firms are X type. The equilibrium is determined by profit equalisation. This simplification characterises only X-type relocation phase in Figure 2. Due to $v_X[a]>0$ for all $a$, once trade freeness is above the break point, all firms move instantaneously and then catastrophic agglomeration occurs. Since the equilibrium path cannot be solved analytically as in the standard NEG models because “$a$” has different powers of $\alpha$ in $\Delta$ and $1-\sigma$, this section diagrammatically shows equilibrium using the Tomahawk Diagram and the Wiggle Diagram, following the standard NEG numerical simulation approach.

3.1. Wiggle Diagram

First, this section evaluates local stability in equilibria diagrammatically using the so-called “wiggle diagram” as in the earliest CP model analysis. Using (8) subject to (12) and (13) under given $a_X=1$, our wiggle diagram depicts market potential gap, $B_X-B_X^*$, in terms of the Northern production share, $s_P$, as Figure 3, which represents market potential difference: a driver of relocation, because the product of the difference, $B_X-B_X^*$, and productivity terms are profit gap between two nations for each firm, i.e. $\pi(a) - \pi^*(a) = b \left( B_X - B_X^* \right) a^{1-\sigma}$. Note that $s_P$ is defined as

$$s_P = \left( \frac{\text{market potential gap}}{\text{productivity}} \right) \left( \int B_X a^{1-\sigma} d\sigma(a) + n \int B_X a^{1-\sigma} d\sigma(a) \right).$$

(At equilibrium, $\tilde{a} = a_R$)

The diagram has upward slope with low trade costs while it is downward sloping with high trade costs. This suggests that the symmetric equilibrium is stable with high trade costs but low trade costs make symmetric equilibrium unstable and a core-periphery equilibrium stable. With intermediate trade costs, two asymmetric equilibria are unstable and symmetric equilibrium and core-periphery equilibrium are both stable. These findings are the same as in the CP, VL and FCVL models in the absence of firm heterogeneity.

---

13 Note that all results are the same if northern firms move to south instead of northern firms moving into the North.
14 At $B_X=B_X^*$ all firms are indifferent in location between two nations, which is an equilibrium.
3.2. **The Tomahawk Diagram**

Next, the equilibria are summarized in a so called “tomahawk diagram”; such a diagram shows the share of the number of firms, $s_n$, as well as production shares, $s_P$, in terms of $\phi$.

$s_n$ is defined as $s_n = n + n^* \tilde{a}'' = 0.5(1 + \tilde{a}'')$ for the relocation from the South to the North, and vice versa. (At equilibrium, $\tilde{a} = a_n$)

The relationship between $s_n$ and $\phi$ for each equilibrium point is plotted in Figure 4. The sustain and break points are not affected by firm heterogeneity. Firm heterogeneity only influences asymmetric unstable equilibria. Firm heterogeneity works as a dispersion force. That is, as firms are more heterogeneous (smaller $\rho$), the delocation process moderates the agglomeration in terms of firm shares in asymmetric equilibria. Intuitively, the most efficient firms have the biggest incentive to locate in the bigger market, because they have the highest sales and intend to save trade costs as seen in the last section. But coincidentally, the delocation of the most efficient firms (i.e. the largest-production firms) causes more local competition in the big market, resulting in lowering the number of delocated firms.
Next, in terms of production shares, $s_P$, firm heterogeneity never affects any of the equilibrium features. The heterogeneity never plays the role of a dispersion force even for asymmetric equilibria in production shares, in which the number of delocated efficient firms is more limited to keep the production shares at a certain constant level. In firm distribution, the increased firm heterogeneity (smaller $\rho$) raises relatively the number of high-productivity firms that can produce more, and thus it is more limited to the delocation in firm shares so as to compensate the constant production shares. Total demand is invariant even at asymmetric equilibria. Accordingly, this keeps constant total production level in each market. What this means is that production share is independent of firm heterogeneity.

Result 3: Firm heterogeneity never affects break points and sustain points. Firm heterogeneity affects asymmetric equilibria in firm shares, and moderates the delocation process as dispersion forces. In production shares, however, firm heterogeneity has no influence on the delocation process even in asymmetric equilibria.
3.3. **Equilibrium Features in non-export fixed cost model**

As in the standard FCVL model, because the firms buy their intermediate inputs, they care about the local cost of production, i.e. price index and $\Delta$. Thus, firm relocation is determined by the real reward to production as shown in (8). Cost linkage plays crucial role via production-cost linkage in this model. Therefore, we can say that the cost-linked circular causality operates (circular causality).

Next, to confirm the result that firm heterogeneity never influences break and sustain points, analytical solutions are provided. Differentiating $B_x-B_x^*$ with respect to $\tilde{a}$ around $\tilde{a} = 0$ using (8), (12) and (13) (given $a_x = 1$), the break point can be derived as

$$\phi^{\text{Break}} = \frac{(1-\mu)(1-\mu+b)}{(1+\mu)(1+\mu-b)}.$$

The sustain point is found to satisfy the following equation:

$$1 - (\phi^{\text{sustain}})^{\mu-1}(1-\mu+b) = 0; s = \frac{1 + \mu - b}{2}.$$

These points are just correspondent to those of the standard (homogeneous-firms) models (FCVL and VL models). In other words, we claim that both points are not a function of firm heterogeneity, $\rho$ and thus we can confirm the previously shown result: firm heterogeneity does not affect the sustain and break points (the neutrality of firm heterogeneity).

Due to $0 < \phi^{\text{sustain}} < \phi^{\text{break}} < 1$, as shown in Baldwin et al. (2003), the stable equilibria are overlapped, and thus shocks to expectations may result in large spatial reallocation either symmetric or full agglomeration (the overlap and self-fulfilling expectations). Moreover, sudden and massive agglomeration could occur as a response of a small reduction of trade costs from $\phi_B$. This is called catastrophic agglomeration. Overall, trade costs reduction ends up to full agglomeration. Also, partial agglomeration is never a stable equilibrium, in which the location of industry is either symmetric or all firms concentrated in one nation or the other (endogenous asymmetry).

Starting from a core-periphery situation, firms are not indifferent to their location. Plainly, core-periphery outcome can be measured by agglomeration rents, i.e. the loss incurred by the relocation from the core to the periphery, the agglomeration rents are a function of $\phi$:

$$\frac{b}{\lambda}(1-\phi^{\mu-1}(1-s+\phi^2 s))a^{1-\sigma}; s = \frac{1 + \mu - b}{2}$$

which is parallel to the ones in the standard NEG models (Baldwin et al. 2003). The agglomeration rent curves look hump shaped. This means that as trade gets freer, the agglomeration rent curve first rises and then falls.

**Result 4:** In the absence of export-fixed costs, where all firms are exporting, firm heterogeneity never affects the sustain and break points. Firm heterogeneity per se has no impact on equilibrium features: circular causality, endogenous asymmetry, catastrophic agglomeration, the overlap and self-fulfilling expectations and hump-shaped agglomeration rent curve.

---

15 See Appendix 1 for the derivation.

16 Although we cannot derive an analytical solution for the sustain point, this equation can be derived by $\pi[a]-\pi^*[a]=-b(B_x-B_x^*) a^{1-\sigma}=0$ with full agglomeration. Since all firms locate in the North (core), and so the total income in the core (north) is given as $(1-\mu+b)/2$ and the intermediate input demand is $(1-\sigma)b$. Thus, $s$ in $B_x$ and $B_x^*$ can be given as $s=(1+\mu-b)/2$. Also $\Delta$ and $\Delta^*$ in $B_x$ and $B_x^*$ can be given as $\Delta = 2^{(1-\rho)}$ and $\Delta^* = \phi_B^{2(1-\rho)}$.

17 See Baldwin et al. (2003) about agglomeration rent curves for detail.
4. THE EQUILIBRIUM IN THE GENERAL MODEL

This section turns to the general model with full features by reviving the assumption of export fixed costs \( F > 0 \) and explores the equilibrium starting with symmetric equilibrium. As discussed in section 2, the efficient firms are X type, while the inefficient firms are D type. Importantly, since firm productivity is dissimilar, the location decisions are firm-dependent. The order of delocation is X type and then after all X type move to the other and then D type starts to move in order of their efficiency. This relocation process continues until the gain from relocation is zero. As seen in Figure 2, X type relocation phase is always positive. Thus, the equilibrium always yields in D type relocation phase rather than in X phase, i.e. \( B_D = B_D^* \) and \( B_X > B_X^* \) for any ‘a’s: All X type firms and some D type firms locate in the core and some D firms remain at periphery. Hence, the cutoff level of relocation \( a_R \) is D type firms, which means that the relocation cutoff level is always more than that of the export market, i.e. \( a_R > a_X \) (Result 2).

4.1. Wiggle Diagram

As mentioned in section 2.4, only the market potential difference, \( B - B^* \), in the profit gap promotes firm relocation and determine equilibrium. \( B_X - B_X^* \) for X-type relocation phase and \( B_D - B_D^* \) for D-type relocation phase, using (8), (12) and (13), and (9), (14) and (15) respectively, are plotted in terms of sp (Figures 6, 7 and 8). A big difference from the non-export fixed costs model is the presence of D-type relocation phase. After all X-firms move to the other, the diagram curve switches from X-type relocation phase to D-type relocation phase. With high trade costs, given the level of F, the diagram of X and D D-type relocation phases is downward sloped (\( sp = 0.5 \) is stable) (Figure 6). Since D-type relocation phase is downward sloped, overall the diagram is down sloped with only one stable symmetric equilibrium. Then, with intermediate trade costs above the break point (\( \phi > \phi^{break} \), X-type relocation phase is upward sloped but D-type relocation phase is downward sloped (Figure 7). Since the symmetric equilibrium is unstable, firms move to the other but after all firms move to the other and next movers are D types. But the D type delocation reduces the southern competition and causes congestion in the North and then the delocation process stops. Partial agglomeration emerges. As trade costs reduce, D-type relocation phase is still downward sloped but steadily shifts up: 1) The increased number of X-type firms increases intermediate input demand ratio for export productions in the bigger North (increased \( s_E \)), and in addition 2) lower trade costs decrease the import prices in the smaller South (decreased \( \Delta^* \)). These two changes shift up \( B_D - B_D^* \) as trade costs fall. At last, with still lower trade costs, down-sloped D-type relocation phase reaches positive value at \( s_P = 1 \), and thus full agglomeration occurs as stable equilibrium, involving positive agglomeration rents (Figure 8).

\[ \text{Note that all X firms are in the bigger North in the D firm relocation phase. Only the North exports M goods.} \]
Figure 6: The Wiggle Diagram in the Presence of Export Fixed Costs with High Trade Costs.

Figure 7: The Wiggle Diagram in the Presence of Export Fixed Costs with Intermediate Trade Costs.
Figure 8: The Wiggle Diagram in the Presence of Export Fixed Costs with Low Trade Costs.

Note that the D-type relocation phase is everywhere downward sloped for any $s_1 > 0.5$. As D firms move to the North, their varieties unavailable in the South increase, and thus some of the Southern consumed varieties (as final goods and as intermediate inputs) are lost and thus drastically falls $\Delta^*$ and rise $\Delta$. This boosts market crowding effect in the core. Since D type firms cannot export to the other, their gain from the location in the core is highly dependent on market competition in its own location. For this reason, this congestion for D-type firms is much more severe than X-type firms and works as severe dispersion forces and reduces their gain from the core location as firm relocation proceeds irrespective of trade costs. What this means is that the firms are deviated from the symmetric equilibrium above break point but when all X type moves to the north and some high productivity D type firms move and some other low productivity D type firms remain at the South, we can reach a stable asymmetric equilibrium. There exist stable asymmetric stable equilibria above the break point. In other words, gradual agglomeration happens without catastrophic agglomeration with trade costs reduction.

4.2. Tomahawk Diagram

Summarising the equilibrium, as seen in Figure 9, the Tomahawk Diagram looks like a pitchfork rather than a tomahawk, in which the symmetric equilibrium turns from stable to unstable and asymmetric equilibria becomes stable before the core-periphery outcome: there is no overlap between break and sustain points and gradual agglomeration occurs as trade gets freer.

Our model contains basic agglomeration forces similar to those in the standard NEG models: forward/backward linkages serve as an agglomeration force and a market crowding effect as a dispersion force. The backward (demand) linkage is caused by the firms buying each other’s output as intermediate inputs. The forward (cost) linkage is caused by the benefits from the proximity of intermediate suppliers. The market crowding effect is caused by firm delocation reducing each firm’s market share in the bigger market.

Nevertheless, there is also a sharp contrast with the standard NEG model results. In the standard CP and VL (also all kinds of VL models including FCVL) models as well as our non-export-
fixed cost-heterogeneous firms model in Section 3, the tomahawk diagram can be always observed but the gradual agglomeration and pitchfork diagram does not appear. The stable equilibria in intermediate trade costs are multiple: stable symmetric equilibrium and core-periphery equilibrium and asymmetric unstable equilibrium.

The reason for the contrast outcomes in our model comes from the fact that D-type firms created by the export fixed costs cannot export to the other market. The delocated D-type firms cannot sell in the original market. This reduction in available varieties reduces access to intermediate inputs, which weakens backward and forward linkages as can agglomeration force for X and D type firms, compared with the standard models. The market crowding effect for D-type firms is more severe than the X type firms: X type firms even reduce market share in the bigger North but they can mitigate the loss in the northern share through exporting and expanding market share in the smaller South. By contrast, D type firms cannot benefit from this kind of compensation by export sales. This severe market-crowding effect for D type firms works as much stronger dispersion force. Furthermore, the impossibility of exporting of D types firms leads to reduce the availability of import varieties in both nations, lowering backward/forward linkages. Therefore, in the vertical linkage model with firm heterogeneity in the presence of export fixed costs, the presence of D-type firms weakens agglomeration forces for all firms and strengthens dispersion force for D type firms, resulting in gradual agglomeration.

Result 5: In the presence of positive export fixed costs, gradual agglomeration occurs above the break point and catastrophic agglomeration never happens. The “tomahawk” diagram evolves into a “pitchfork”.

Result 6: The impossibility of exporting by the D-type local firms reduces agglomeration forces and increases dispersion force. This leads to gradual agglomeration.

Moreover, there is the possibility of non-full agglomeration, in which full agglomeration never occurs even at $\phi=1$ and asymmetric equilibrium ($0 < s_p < 1$) is stable. Full agglomeration never happens if
Higher export fixed costs and firm heterogeneity (higher $\lambda$, i.e. lower $\rho$) are likely to cause non-full agglomeration. Intuitively, since most firms are D type and a few firms are X type due to high $F$ or low $\rho$, firm relocation immediately involves D type relocation phase relocation. But this phase faces severe market crowding effect and then relocation stops at partial agglomeration. To summarise, higher export fixed costs or a higher degree of firm heterogeneity makes non-full agglomeration more likely.

### 4.3. Discussion of the equilibrium

**Circular Causality and Endogenous Asymmetry**

As in the non-export fixed costs model, there exists cost linked circular causality, because the firms buy their intermediate goods. But demand linkage and cost linkage are weaker than non-export fixed cost model and the NEG models because of less varieties in intermediate inputs due to unavailability of D type firms’ varieties in foreign market. Furthermore, the market crowding effects are different across firms, in particular the effects for D-firms are more severe than that of X-types as they have no opportunity to sell into foreign markets.

A gradual trade costs reduction eventually leads to full or partial agglomeration as a stable equilibrium. Different from our previous model and the other NEG models, the equilibrium path looks like “pitch fork”. Thus, trade costs reduction in a perfectly symmetric equilibrium results becoming endogenously asymmetric. But, it might not involve all firms in one nation or the other. There is the possibility of non-full agglomeration, i.e. non core-periphery structure.

**Non-catastrophic Agglomeration and no-overlapping and self-fulfilling expectations**

While the sustain point was not analytically solvable as reduced form in the standard VL model and our non-export fixed costs model, our model, solving $B_D = B_D^*$ at full agglomeration with respect to $\phi$, can explicitly be written as

\[
\phi_{\text{sustain}} = \left(\frac{1-s}{s}\right)^{1-\mu} \frac{1}{a_X^\sigma}; \quad a_X = \left(\frac{1-s}{\lambda f}\right)^{1-\mu}, \quad s = \frac{1+\mu-b}{2} \tag{20}
\]

Note that $a_X$ is determined by the export cutoff conditions (10), independent of $\phi$.\(^{20}\) The sustain point is substantially different from the non-export fixed costs case.\(^{21}\)

However, the break point is intractable due to the complexity of a marginal shift of $a_X$ in response of marginal firm relocation deviated from the symmetric equilibrium. This makes impossible to derive sustain point analytically. Thus, attempted to reveal a good comparison with non-export fixed costs model, the break point is approximately derived by assuming a fixed

\[\text{Solving } B_D - B_D^* = 0 , \text{ where } \Delta = \lambda^{1/(1-\mu)}, \Delta^* = \lambda \phi \Delta^\sigma a_X^\sigma = \lambda^{1/(1-\mu)} \phi a_X^\sigma \text{ and } s=(1+\mu-b)/2, \text{ we can get } s = (1-s)/(\phi^{1-\mu} a_X^\sigma (1-\mu)) .\]

\[\text{From the export cutoff condition, } \phi a_X^{\lambda^\sigma \lambda^{(1-\mu)}} \frac{1-s}{\lambda^{1/(1-\mu)} \phi a_X^\sigma} = f \Rightarrow \frac{1-s}{\lambda a_X^\sigma} = f . \text{ Note that } a_X \text{ is independent of } \phi .\]

\(^{20}\) If $a_X$ is thought to be as 1 due to $F=0$ and sales and profits from the foreign market are added in the formulation of $B_D$, i.e. transformed to $B$ function, the sustain point converges to the one in the non-export fixed costs case.
$a_x$ (assuming no response from the change of $a_b$). In sum, the approximate break point, $\phi^{\text{Break}}$, regardless of not reduced form, can be implicitly expressed as

\[
(21) \quad \phi^{\text{Break}} = \frac{(1 - \mu)(1 - \mu + b)}{(1 + \mu)(1 + (\mu - b)a_x^\mu)}. \]

The approximate break point is always larger than in non-export fixed costs (16) because of $0 < a_x < 1$. (higher $\phi^{\text{Break}}$ reduces $a_x$, and decreased $a_x$ increases $\phi^{\text{Break}}$.) Accordingly, $\phi^{\text{Break}}$ is larger than those of the standard CP and VL model.

Next, to rigorously check the non-overlapping of the sustain $\phi^{\text{sustain}}$ and “exact” break point $\phi^{\text{Break}}$ in a sense of assuming away no response of $a_x$ through delocation, Figure 10 plots sustain and ‘exact’ break points in terms of $F$ using numerical simulation. The minimum effective level, $F^{\text{min}}$, represents the least efficient firms are D-type firms.\(^{23}\) (The right hand side of the $F^{\text{min}}$ is feasible.) The figure illustrates no overlap, i.e. $\phi^{\text{Break}} > \phi^{\text{sustain}}$, which is parallel to the diagrammatic outcomes from the Wiggle diagram. Note that sustain point increases as $F$ rises, which indicates the increased $F$ is less likely to occur full agglomeration. The expansion of the gradual agglomeration range indicates more likely happened gradual agglomeration.

---

\(^{22}\) See Appendix 1 for derivation.

\(^{23}\) $F^{\text{min}}$ is the condition for at least one D type firm operating at full agglomeration. The export cutoff condition can be written as $F = ba_x^{\lambda - 1}(1 - s)$. Due to $\Delta^{1 - \mu} \leq \lambda$ and $a_x < 1$, we can derive the condition: $F^{\text{min}} = b\lambda \frac{1 - \mu + b}{2} < F$. 

---

**Figure 10: The sustain/break points and export fixed costs**

In sum, since the model does not have the overlapping range between break and sustain points, the model does not feature catastrophic agglomeration by marginal trade costs reduction from break point. Due to the presence of D-type firms, gradual agglomeration occurs (No catastrophic agglomeration). That is to say, there is no possibility of large sudden spatial reallocation involving full agglomeration caused by shocks to expectations (No overlap and self-fulfilling expectation). To summarise,
Result 7: No overlap between the break and sustain points implies neither catastrophic agglomeration nor self-fulfilling expectation.

Note that although this finding of non-multiple equilibria and gradual agglomeration is entirely contrast to the widely spread findings of the standard NEG models like Core-Periphery Model, empirical studies could support our findings (Davis and Weinstein, 2002 and 2004).  

**Agglomeration Rent Curve**

Different from the non-export fixed costs model and the standard NEG models, agglomeration rents curves are for D-type firms, because the least efficient D-type firms move to periphery first. That is, the agglomeration rents for D-type firms are given as

\[
\frac{b}{\lambda} \left( s - \frac{1 - s}{\phi^{1/\mu}} \left( (1 - s) / (1 - \mu) / \lambda f \right)^{(1 - \mu)/\lambda} \right) a^{1-\sigma}
\]

which is increasing in \( \phi \). Agglomeration rent curve has monotonically positive slope (Figure 11). Specifically, agglomeration rents at free trade, \( \phi=1 \), are represented as

\[
\frac{b}{\lambda} \left( s - \frac{1 - s}{(1 - s) / \lambda f} \right) a^{1-\sigma}.
\]

---

24 Their review of the recovery of Japanese cities after the Second World War, Davis and Weinstein (2004) did not find any empirical evidence to verify the existence of multiple stable equilibria and catastrophic agglomeration as the standard NEG model suggests.

25 With the value being less than zero, non-full agglomeration can be derived as (19)
varieties. Free trade with small $F$ allows for so many X-firms, exporting costlessly to the periphery, which seeks to reduce the discrepancy of the market congestion effect between two nations. On the other hand, forward and backward linkages in the North are stronger than in the South because of all D-type firms in the North. For this reason, all firms prefer to the core even at free trade in case of small $F$ (full agglomeration case).\textsuperscript{26}

**Result 8:** In the presence of export fixed costs, the agglomeration rent curve is upward sloped with respect of $\phi$. Even at free trade, agglomeration rents could be positive. Free trade never equalises firm location preference between core and periphery.

The standard NEG model has hump-shaped agglomeration rent curve. On the other hand, D-firms in our model cannot export to the other in our model, and thus the agglomeration rent curves are increasing function. This outcome could be contrast to corporate tax competition in NEG literature (Baldwin et al. 2003). When we come to think about agglomeration rents as taxable resource, as suggested by Baldwin and Krugman (2004), taxable rents increase. This is because D-firms always prefer to locate bigger market due to the impossibility of export for any $\phi$. Thus, tax competition in our model is race to the top over trade liberalisation rather than race to bottom.

5. **INEQUALITY--TRADE AND TBTs LIBERALISATION**

5.1. **Export Fixed Cost (TBTs) liberalisation**

Trade liberalisation has long focussed on lowering trade costs, but many kinds of trade barriers still remains among the developed countries, related to market specific regulations and policies. These kinds of barriers can be called technical barriers to trade (TBTs), currently discussed in the WTO. In our model, $F$ can be interpreted as foreign market specific costs for regulation and market policies difference. Then, the reduced $F$ owing to the TBTs liberalisation allows more D-type firms to enter the export market, promoting demand and cost linkages due to more availability of foreign varieties. As seen in Figure 10, while both sustain and break points decrease in TBTs liberalisation, the range of the gradual agglomeration, i.e. between $\phi_{sustain}$ and $\phi_{break}$, reduces. Full agglomeration is more likely to occur.

**Result 9:** TBTs liberalisation promotes agglomeration process: Decreasing the break and sustain points and reducing the gradual agglomeration range.

5.2. **Inequality of Nations**

Now, turning to the discussion of Krugman and Venables (1995) that there is convergence of real incomes and welfare between core and periphery at free trade, this part studies whether trade and TBTs liberalisation ends up to the convergence in welfare by starting from the core and periphery structure above sustain point for simplicity.\textsuperscript{27} The total utility of each nation is given by $V = \frac{Y}{D^\mu} = \frac{Y}{\Delta_{1-\sigma}}$ in the North (core) and $V^* = \frac{Y^*}{P^\mu} = \frac{Y^*}{\Delta_{1-\sigma}}$ in the South (periphery), where $Y$ is total workers’ and capital owners’ income in each nation, $Y=Y^*=(1-\mu+b)/2$ due to symmetric endowments and portfolio assumption of capital return. Note that two symmetric initial endowments and capital owners’ portfolio keeps each nation’s total nominal income constant. Since all firms are in the North and have no incentive for deviation above

\textsuperscript{26} Note that the agglomeration rents are always negative in case of non-full agglomeration case

\textsuperscript{27} In this part, trade liberalisation is to gradually reduce trade costs from the sustain point and TBTs liberalisation is to reduce $F$ up to $F_{min}$ (at least the least efficient firms are D-type).
sustain point, Y and Δ are not affected by trade (increased φ) and TBTs (decreased F) liberalisations. Hence, $V$ is constant. On the other hand, trade liberalisation directly increases $\Delta^* (= \lambda \phi \Delta' a_x')$ without changing $a_x$ in $\Delta^*$, while TBTs liberalisation (the fall of F), ceteris paribus, increases $\Delta^*$ by increasing $a_x$, namely expanding the consumed varieties in periphery. This is because $a_x$, determined by the cutoff point condition, $\frac{(1-s_x)}{\lambda^2 a_x'} = F / b$, is independent of φ but decreasing in F (footnote 20). It follows that $V^*$ decreases. Both liberalisations never produce convergence in both nations’ welfares due to the presence of unavailable local core producers’ varieties in the consumption in periphery. This implies that the small country would take negative attitude towards trade liberalisation as well as TBTs liberalisation due to the divergence of welfare. These outcomes are entirely different from Krugman and Venables’ outcomes. Our outcomes might more or less reflect the current strenuous liberalisation process.

**Result 10: Trade and TBTs liberalisations never equalise welfares between core and periphery. On the contrary, the liberalisations produce a divergence in nations’ welfare.**

Peripheral nations are likely to object to trade and TBTs liberalisation. Importantly, this result tells us how important it is that the economy should take a core position and create agglomeration. Different from Krugman and Venables (1995), the economies in trade liberalisation never converge but diverge between core and periphery once trade costs reach the break point and gradual agglomeration occurs. Thus, it is more crucial and important for nations to involve full agglomeration and become an agglomerated area. Subsidies and other kinds of industrial policies aiming at attracting more firms at an initial stage (symmetric equilibrium at high trade costs) crucially affect the future core-periphery structure. These policies are much more important than we have predicted from existing NEG models and are based on the creation of divergence in the inequality of nations in a liberalisation process.

### 5.3. **Inequality within a Nation**

Finally, drawing attention to heterogeneous firms’ behaviour within the core (North), we discuss the impact of trade and TBTs liberalisation on firms’ profits in the North in core-periphery structure. As seen in the left panel of Figure 12, plotted profit in the core in terms of $a$, trade liberalisation boosts exporters’ profits without altering D firms’ profits (no reallocation effect but profit expansion effect) because of no impact of $\phi$ on $\Delta$ in $BD$ due to no northern imports, and without allowing the most efficient D firms to enter the export market (no selection effect) because of $a_x$’s independence of $\phi$ (Footnote 20). The benefit of trade liberalisation is allocated by only exporters with remaining local producers at the same level of profits in our core-periphery structure. This is different result from Melitz (2003), which has reallocation effect and selection effect through entry and exit. On the other hand, our TBTs liberalisation (fixed $\phi$) has selection effect. As seen in the right panel of Figure 12, the fall of $F$ due to the liberalisation increases $a_x$ (Footnote 20) and allows the most efficient local firms to enter the export market. The increased exporter reduces per firm profit from export market. Accordingly, each exporter’s profit falls. This implies that TBTs liberalisation has a redistribution effect within a nation. Intuitively, our core-periphery structure in NEG model tells us that TBTs liberalisation is more important in a viewpoint of redistribution of the benefit of liberalisation.
Result 11: The benefit of trade liberalisation is allocated only to existing exporter without allowing the local firms to enter the export market. On the contrary, TBTs liberalisation allows them to enter the export market, associated with the reduction of each exporter’s profit.

6. CONCLUSIONS

This paper studies the impact of trade liberalisation on firm location in a heterogeneous-firm economic geography model with intermediate inputs linkage. In the absence of export fixed costs, firm heterogeneity dampens firms’ shares in asymmetric unstable equilibria but has no impact on the other equilibrium properties. By contrast, in the presence of export fixed costs, firm heterogeneity definitely dampens the agglomeration process. Rather than catastrophic agglomeration, we find gradual agglomeration and the possibility of a non-full agglomeration outcome. This is consistent with some empirical findings (Davis and Weinstein, 2002; 2004). Furthermore, trade and TBTs liberalisation increases divergence in welfare between core and periphery and even at free trade both economies never converge. The periphery suffers the loss from liberalisation. This implies that policies for attracting firms and creating agglomeration in the early stage of liberalisation process are much more crucial and important for the future path than in the standard NEG models.

This paper finds the possibility of inequality in welfare across nations and in profits across firms as a consequence of trade and TBTs liberalisation. The current situation of poor countries remaining poor without fully benefiting from globalisation might be reflected by the divergence discussed in our paper. It is worthwhile to state that this divergent outcome in welfare implies that there is still room for intervention by governments’ policies to re-allocate firm location across nations associated with trade and TBTs liberalisation. It might be possible to say that the currently observing highly concentrated industrial agglomeration in some specific nations in the EU, offshoring in North America, delocation in Europe and hollowing-out in Japan would be in scope of coordination in industrial allocation by cross-governmental policies. With respect to developing nations, the development policies in the early stage might be much more crucial to the future firm allocation and discrepant welfare than we have thought. This paper implies that it
is necessary to coordinate each developing country’s interests before their trade liberalisation. But the way of the coordination remains future research. It would be necessary to extend to the framework of multi-countries and sectors.

REFERENCES


Falvey, R., D. Greenaway, and Z. Yu (2004). “Intra-industry Trade between Asymmetric Countries with Heterogeneous Firms”, Research paper series -Globalisation, Productivity and Technology, the University of Nottingham


**APPENDIX 1. THE BREAK POINTS**

Here, we show the ‘exact’ break point. The symmetric equilibrium becomes unstable above the break point. As shown in the Wiggle diagram, the symmetric equilibrium is locally unstable if the slope of $B-B^*$ is positive at $s_p = 0.5$. Because of the symmetry of the model, we derive the break point by $d(B-B^*)/ds_p$ at $s_p = 0.5$ which implies $\pi = \pi^*$, $\Delta = \Delta^*$ and $s_E = 0.5$. Now differentiating $B-B^*$ given as (8), (10), (12) and (13) at the symmetric equilibrium gives,

$$\sum \left( 1 - \frac{\mu(1 + a^*_p - \phi(a^*_r - a^*_X))}{1 + \phi a^*_r} \right) \frac{d\Delta}{\Delta} - \frac{(1 - \phi)a a^*_r}{1 + \phi a^*_r} da_r - \frac{\phi a a^*_r}{1 + \phi a^*_r} da_X = 0$$

$$-a_X (1 + \mu) \frac{d\Delta}{\Delta} + 2a_X ds + (1 - \sigma) da_X = 0$$

$$-d\pi + 2s \left( \mu - \frac{1 - \phi}{1 + \phi} \right) \frac{d\Delta}{\Delta} + 2 \frac{1 - \phi}{1 + \phi} ds = 0$$

$$-ds + \frac{1}{2} b(\sigma - 1) d\pi + \frac{1}{2} b(\sigma - 1) a a^*_r da_r = 0$$

We can get exact break point, $\phi^{\text{break}}$, so as to satisfy $\frac{d(\pi - \pi^*)}{d\tilde{a}} \bigg|_{\tilde{a}=0} = 0$. However, the equations cannot be given analytical form. Instead, by fixing $a_X$, approximate break point, $\tilde{\phi}^{\text{break}}$, can be derived. To work with this, $da_X = 0$ is inserted into (23) and (24), and then solving (23), (24), (25) and (26), solving $\frac{d(\pi - \pi^*)}{d\tilde{a}} \bigg|_{\tilde{a}=0} = 0$ in terms of $\phi$, $\tilde{\phi}^{\text{break}} = \frac{(1 - \mu)(1 - \mu + b)}{(1 + \mu)(1 + (\mu - b)a^*_X)}$. Next, we derive the break point in non-export fixed costs model, i.e. $a_X = 1$ and $da_X = 0$. Using (23), (25) and (26) and solving $\frac{d(\pi - \pi^*)}{d\tilde{a}} \bigg|_{\tilde{a}=0} = 0$ in terms of $\phi$, we get $\phi^{\text{break}} = \frac{(1 - \mu)(1 - \mu + b)}{(1 + \mu)(1 + \mu - b)}$.

**APPENDIX 2. THE ORDER OF RELOCATION AND X-TYPE RELOCATION PHASE (FIGURE 2)**

Here, we show mathematically that the X-type relocation phase has infinite slope around $\tilde{a} = 0$ above the break point, in other words, the highest productivity firms have the largest incentive to move to the other location. We consider a point just below or equal to the break...
point: \( \frac{d\pi_X - \pi^*_X}{d\tilde{\alpha}} \bigg|_{\tilde{\alpha} = 0} \geq 0 \), as seen in Appendix 1, from which it follows that \( B_X - B^*_X \geq 0 \). Using
\[
\pi_X (\tilde{\alpha}) - \pi^*_X (\tilde{\alpha}) = b(B_X - B^*_X )\tilde{\alpha}^{1-\sigma} \]
(X-type relocation phase in Figure 2), we can differentiate the profit gap in terms of \( \tilde{\alpha} \) around \( \varepsilon > 0 \), where \( \varepsilon \) is a tiny positive number close to zero:
\[
\frac{d\pi_X - \pi^*_X}{d\tilde{\alpha}} \bigg|_{\tilde{\alpha} = \varepsilon} = b(1 - \sigma)(B_X - B^*_X )\varepsilon^{-\sigma} + b^2\left( \frac{B_X - B^*_X}{\varepsilon} \right) \varepsilon^{1-\sigma} \]
This corresponds to the slope of the X-type relocation phase in Figure 2. Then we take the limit in terms of \( \varepsilon \) toward zero, consequently getting infinity:
\[
\lim_{\varepsilon \to 0^+} \frac{d\pi_X - \pi^*_X}{d\tilde{\alpha}} \bigg|_{\tilde{\alpha} = \varepsilon} = \lim_{\varepsilon \to 0^+} \left( b\frac{d(B_X - B^*_X)}{d\tilde{\alpha}} \varepsilon^{1-\sigma} \right) = +\infty \]
because of
\[
\lim_{\varepsilon \to 0^+} (B_X - B^*_X) = 0, \quad \lim_{\varepsilon \to 0^+} \varepsilon^{1-\sigma} = +\infty \quad \text{and} \quad \frac{d(B_X - B^*_X)}{d\tilde{\alpha}} \bigg|_{\tilde{\alpha} = \varepsilon} > 0 \]
This means that the slope of the X-type relocation phase goes to infinity around \( \tilde{\alpha} = 0 \), which means that the highest productivity firms have the largest incentive to go to the other location above the break point.