# Why Capital does not Migrate to the South: A New Economic Geography Perspective

Jang Ping Thia*

November 2006 (First draft)
March 2007 (This draft)

**Abstract**

This paper explains why capital does not flow from the North to the South - the Lucas Paradox - with a New Economic Geography model that incorporates mobile capital, immobile labour, and productively heterogeneous firms. In contrast to neoclassical theories, the results show that even a small difference in the ex-ante productivity distribution between North and South can a have huge impact on the location of firms. Despite differences in aggregate capital to labour ratios, wage and rental rates continue to be the same in both locations. The paper also analyses the effects of risk on industrial locations, and shows why 'low tech' industries tend to migrate to the South, while 'high tech' industries continue to locate in the North.

## I Introduction

It has long been a source of consternation among economists why there has been much less capital flow from the capital rich industrialised economies to the capital poor developing

*Email: j.p.thia@lse.ac.uk, London School of Economics, Centre of Economic Performance, Houghton Street London WC2A 2AE. Updates to this paper will be posted on http://personal.lse.ac.uk/thia/ . I would like to thank my supervisor, Stephen Redding, for his advice, particularly on the numerical solutions which he has done in his previous research. I gratefully acknowledge the PhD funding from the Ministry of Trade and Industry (Singapore).

economies compared to what neoclassical theory predicts. Using a standard neoclassical growth model, Robert E. Lucas Jr. (1990) shows that the implied marginal productivity of capital in India is an astounding 58 times that of US. It is therefore a puzzle among economists why traditional theories cannot explain the capital flow (or lack of) from the developed to the developing economies.

There has been much research dedicated to explaining the 'Lucas Paradox'. Some economists have used differences in fundamentals (production structure, technology, policies, institutions) as explanations for the paradox. For example, Lucas cited the differences in human capital as the key reason why capital does not move to the South. On the other hand, other economists have mainly relied on capital market failures (expropriation risks, sovereign risks, asymmetric information) to resolve the paradox[1].

Even more interestingly, not all economists agree that there is a paradox in the first place. Carmen M. Reinhart and Kenneth S. Rogoff (2004) even suggest that the real paradox is that there is in fact too much capital flow to developing countries, considering the history and incidence of default in these economies. Aaron Tornell and Andres Velasco (1992) highlight the theoretical possibility that poor property rights may even result in capital flight from a capital poor country to a capital rich country with better protection (or private access). It is therefore not abnormal that capital stays in the North as it offers better property rights. On a separate note, Abhijit Banerjee and Esther Duflo (2004) document comprehensive evidence which suggests that differences between the rates of return within some economies are larger than those across countries, which of course brings to the question of whether it is relevant to focus on Lucas Paradox.

Nevertheless, it is also clear from empirical research that it is often difficult to distinguish one theory from another. Countries with weak institutions tend to have lower human capital, and weak institutions tend to be associated with greater information asymmetry and expropriation risks. There can be too much or too little capital to the South, depending on which benchmark model is used, what instruments are used, what

[1] See Alfaro et al (2005) for a brief discussion on the various competing hypotheses.

is defined as capital, and what kind of growth accounting is used[2].

Notwithstanding the various arguments presented, development over the past decade has necessitated a new understanding to the Lucas Paradox. The opening of China, India and other major emerging economies has resulted in increased flow of Foreign Direct Investment (FDI) to what is loosely termed as the South. The flow of capital is however highly uneven. In a recent working paper by Stephany Griffith-Jones and Jonathan Leape (2002), the authors highlight the huge differences in the capital flows to emerging economies. China has attracted a fifth of all private capital flows to developing countries in the 1990s, peaking at $60 billion a year in 1997. India's share has been paltry by contrast, with a peak of $7 billion only in 1994. The latest figures show that China took in $72 billion in FDI in 2005, while India only received $6.6 billion[3]. Despite recent headline grabbing growth rates from India, the FDI gap with China has not closed, although this might change in the near future.

If the Lucas paradox exists for India, it is on the face of it much less of a paradox for China. Is it therefore correct to conclude that China somehow has better fundamentals - institutions, technology, human capital, and/or less capital market imperfections? Given the fact that India is a stable parliamentary democracy, has a deeply entrenched English legal system with the associated emphasis on property rights, and a largely free press, it is difficult to turn the argument around and conclude that China has better institutions, or that better functioning markets result in the huge difference in observed investment

[2] Francesco Caselli and James Feyrer (2006) offer a similar insight by making a distinction between reproducible and non-reproducible capital. The authors argue that the reward to reproducible capital is in fact rather low in the South once proper accounting is done. There is therefore no parodox that capital does not move there.

[3] China's cumulative inward FDI stands at $318 billion compared to $45 billion for India (UNCTAD). The difference in the levels of FDI is not due to differences in domestic investment. Inward FDI makes up 11.3 per cent of China's gross capital formation between 1990-2000, but only 1.9 per cent compared to India. One of the explanations for the big difference is the effect of 'round-tripping' - domestic investment by Chinese firms disguised as FDI due to gain a tax advantage. A look at foreign investment position from the US however recorded the following difference: US cumulative investments in China and India (historical price) stand at $16.9B and $8.5B respectively. For manufacturing, the respective figures are $8.8B and $2.4B (Bureau of Economic Analysis). While 'round-tripping' may well account for some of the difference between the FDI China and India have received, it is clear that China continues to receive signficantly more bona fide FDI than India.

flows[4].

The puzzle is therefore not only why relatively little capital has flowed to the developing economies but also the distribution of the flow of capital to these economies. The objective of this paper is to synthesize the New Economic Geography (NEG) understanding of location of industries with more recent firm-heterogeneity trade model, in order to bring about a new understanding to an old puzzle as well as answer some of these new questions posed.

NEG researchers have had more than a decade of success in demonstrating how industrial agglomeration can result. These models demonstrate how a symmetric fall in trade cost can result in highly asymmetric outcomes (catastrophic agglomeration). The first NEG model, popularly known as the Core-Periphery (CP) model, by Paul R Krugman (1991) demonstrates how the migration of industrial workers can result in industry concentration in a location. Subsequent work by Anthony J. Venables (1996) shows how vertical linkages (VL) between industries can result in firm migration with a similar agglomeration effect. These two models exhibit 'cumulative causation'[5]. An example of the mechanics is that firms locate where there are workers, and workers locate where there are firms (to reduce cost of living), giving rise to a feedback effect. These models tend to be highly intractable as a result. A more tractable model of industrial location is the 'Footloose Capital' (FC) model due to Philippe Martin and Carol Ann Rogers (1995). The key assumption of the model is that only capital is mobile, thereby breaking the feedback effect.

This paper has chosen to adopt the FC assumption. First, international economics

[4] The problem with looking at historical data for defaults to explain current allocation, or predict future capital flows, becomes evident here. Historical data do not account for regime changes, changes investor confidence and perception about the future. Reinhart and Rogoff (2004) duly note that India has never defaulted while China has defaulted on two occasions between 1901-2002. Yet, it is China that has taken in a lion share of FDI.

[5] Some NEG models exhibit agglomeration - namely a feedback effect that generates ever increasing pressure for firms to locate in any one location. Other NEG models exhibit concentration effect, why one observes more firms in one location even though there is no feedback effect. Both these classes of NEG models explain the spatial locations of firms without appealing to production externalities, which are captured by 'spillover' models.

continues to be dominated by high capital mobility. Though there is international migration of labour, the speed at which adjustments take place is far slower, and its magnitude much smaller, compared to the movement of international capital. CP models therefore are more useful in explaining regional adjustments within national economies than across countries. Second, the VL models rest on the onerous assumption that firms require differentiated inputs or Either type production function - specifically, all downstream firms using all upstream firms inputs. While this offers a theoretical benchmark, its stylised assumption that all firms use all inputs is not often observed in real life.

In essence, the model in this paper assumes mobile capital, immobile labour, and firms with heterogeneous productivity. There are two locations North and South. Differences between the two regions are characterised not by the aggregate production functions, but by differences in the productivity (pareto) distributions of firms. The shares of manufacturing firms in each location are then solved for the equilibrium by equalising the ex-ante value of entry in both locations. Several interesting results emerge from the exercise.

## A Explaining the Lack of Capital Flow to South

First, while neoclassical models suggest that the productivity difference between North and South has to be very large to explain the lack of capital flow, this paper shows that a small improvement in North's productivity (by changing the mean of the pareto distribution) can have a dramatic impact on the share of firms, while keeping the returns to factors equal in both locations. This therefore resolves the Lucas paradox. Admittedly, this paper does not explain why the small difference in productivity arises in the first place. This question is better left to development or political-economy researchers [see James R. Tybout (2000) for a brief discussion].

## B Resolving The Paradox of Risk

The second key result concerns the effect of risk. A well known property of the profit function is its convexity. Consider the example of the constant elasticity of substitution (CES) preference function. For whatever the cost of production (inverse of productivity), the firm's revenue is bounded from below by zero - that is, revenue is always positive no matter how high the cost (and price) is. However, there is no upper bound to revenue. Therefore, a means-preserving spread of productivity actually increases expected profits because of the very convexity of the profit function[6].

If the South were to have greater aggregate productivity risks while keeping its mean productivity equal to the North, this would imply that expected profit is higher there, and mobile capital will flow to the South until the expected returns to capital is once again equalised for both locations. This is the 'paradox of risk' for it contradicts commonplace intuition that firms shun locations perceived to have high risks to production. But in principle, the firm is a risk neutral entity. As long as the firm maximises expected profits, why does it care about risk?

It turns out that there is a good reason for this if one thinks of risk as in a firm-specific productive risk in Marc J. Melitz (2003). Each firm will have to pay a sunk cost to attempt entry into a market. Upon the payment of this cost, the firm draws a level of productivity specific to itself, from an ex-ante distribution. The firm then makes the decision whether to continue production based on the level of realised productivity. If productivity is high enough, the firm will sink in fixed cost and produce. Otherwise, the firm 'let bygones be bygones' and loses the sunk cost.

It turns out that in equilibrium, the level of sunk cost will have significant impact on the location of industries. High sunk cost industries prefer less risky locations because they offer them a higher probability of entry. Coupled with the home market effect, the model can explain why 'hi-tech' industries - characterised by high sunk cost - cluster in

[6] A mean-preserving spread of expenditure will have no such effect since it will still result in the same expected profits.

the less risky North while 'low-tech' industries move to the risky South.

### C The Importance of Trade Cost

Finally in the standard FC model, if expenditures are equal in both locations, the distribution of firms will continue to be symmetric at all levels of iceberg trade costs except zero. FC models can only achieve asymmetric concentration of industries through the home market effect (that is, different expenditure shares) whereby the location with the larger expenditure share has more than its proportionate share of firms. Trade cost is impotent otherwise.

The introduction of firm heterogeneity restores the potency of trade cost. Different levels of trade costs will result in different concentration of industries, even if expenditures are the same at both locations.

**Outline of Paper**

Section II will provide the setup of the model. Section III will solve and provide the equilibrium conditions to the model. Section IV will give the results of the numerical solutions. Section V concludes.

## II The Model Setup

### A Endowment and Regions

There are two primary factors of production - capital and labour. There are two regions - North and South, subscripted by $N$ and $S$ respectively. The North has $K_N$ units of capital and $L_N$ units of labour while the South has $K_S$ and $L_S$, all factors in fixed and known quantities. Capital is completely mobile between regions, and capital returns can be costlessly remitted to owners for consumption. Workers (who are also owners of capital) are completely immobile between regions, and their labour is supplied inelastically to the local market.

## B Preferences

There are two types of goods - agriculture ($a$) and manufacturing ($m$). The motivation is similar to most NEG models. The consumer's utility is given as

$$u_j = c_{mj}^{\mu} c_{aj}^{1-\mu}$$

where $c_{mj} = \int_{\Omega} c_i^{\frac{\sigma-1}{\sigma}} di$ is the consumption of the set of $\Omega$ composite of manufactured goods, $\sigma > 1 > \mu$.

## C Technology and Firms

### i Agriculture

The agricultural sector has a constant returns to scale production function. For simplicity, 1 unit of labour will produce 1 unit of output. As per the usual assumption for NEG models, the agricultural good is costlessly traded between countries. This assumption equalises the price of agricultural good between North and South, and also equalises the wage per unit of effective labour because of the perfectly competitive, constant returns to scale production.

### ii Manufacturing

The manufacturing sector requires a composite factor production $\kappa$ which is produced by the primary factors - capital and labour - with a constant returns to scale Cobb-Douglas production technology

$$\kappa = AK^{\alpha}L^{1-\alpha}$$

where $A$ is the aggregate technology parameter.

There is a large number of firms, each producing one variety. The firm's technology is homothetic and represented by the familiar increasing returns function

$$C_i = \left[f + \frac{q_i}{\varphi_i}\right]\kappa$$

where $f$ is the fixed production cost, $q$ the output and $C$ is the total input requirement of the $i^{th}$ firm in terms of $\kappa$. All firms have the same fixed cost but different levels of productivity $\varphi$.

Traditionally, the FC model has a disembodied technology - capital inputs for fixed cost and labour inputs for variable cost. Using a standard FC model but incorporating firm heterogeneity, Richard E. Baldwin and Toshihiro Okubo (2005) show how the home market effect can induce for productive firms to relocate to the larger market. That paper takes the ex-post productivity distribution of firms as given and ignores the entry or exit decision of firms. In a subsequent paper, Baldwin and Okubo (2006) introduce the entry and exit process. In this paper, the author again highlight the home market effect, but further show how instantaneous entry and exit is a perfect substitute for re-location.

To achieve more analytical tractability, Baldwin and Okubo (2006) make some simplifying assumptions. Sunk cost, fixed export cost (beachhead cost), and variable production cost is borne by labour inputs only. Fixed production cost consists of capital only. The production technology is therefore a disembodied one, much like the standard FC model. In a firm heterogeneity setup however, there are many types of cost. Though it is not a criticism of the Baldwin and Okubo setup, it is not clear (at least theoretically) why the production technology should be so. Realistically, one could also think of sunk or beachhead cost to consist of capital only, or a combination of capital and labour.

A further complication arises from a non-homothetic production function. Melitz (2003) requires homotheticity in order not to introduce systematic differences between the cutoffs of large or small markets. Market size determines only the mass of firms, not the cutoffs, in equilibrium. If the production function is non-homothetic, larger markets will have higher productivity even if the ex-ante production distribution is the same. To give an example, US firms will be systematically more productive than Switzerland, even if firm specific productivity is drawn from the same distribution. This does not appear to be well supported empirically. This paper therefore adopts a more uniform approach towards the various types of costs by assuming a homothetic production technology that is more similar to Bernard, Redding and Schott (2004) - known henceforth as BRS - where all costs require the same composition of inputs. Though it sacrifices some analytical tractability, it brings several advantages.

First, it is more realistic in that all costs will require capital and labour. The homotheticity of inputs towards manufacturing allows the model to be solved easily as in Melitz (2003) even in the presence of firm heterogeneity by making use of the 'Zero Cutoff Profits' and 'Free Entry' conditions. With a disembodied technology, this cannot be done. Second, changes in absolute or relative endowment do not have an impact on firm level aggregates. The presence of a costlessly traded homogeneous good (agriculture) equalises wages between two locations. Capital mobility ensures that rental rates are also equalised. In equilibrium, atomistic firms therefore see a infinitely elastic labour and capital supply at fixed prices. Changes in endowment affect only the number of firms, relative returns of primary factors, and associated welfare, with no additional effect on firm level aggregates. The effect of changing endowment is just like changing market size[7]. Third, if the productivity distributions are the same, cutoffs will be the same from the previous point. Though the large country has more firms, its firms are therefore no more productive on average compared to small country firms if the ex-ante productivity distributions are the same. This therefore also avoids some irregularities where the larger the market, the higher the average productivity - a result demonstrated in Baldwin and Okubo (2005) but yet is supported empirically in terms of international comparison.

Finally, though this paper draws inspiration from BRS (2004), there is a key difference.

[7] Consider the opposite case with a disembodied technology, where suppose only capital is used for the sunk cost $f_e$. An increase in capital endowment, relative to labour, will mean that there will be relatively more resources for sunk cost compared to production. In equilibrium, it has to be more difficult to gain entry, and cutoff productivity has to increase. In other words, with a disembodied technology, changes in relative endowment will affect firm level aggregates.

In BRS (2004) both factors of production - skilled and unskilled labour - are immobile. In this paper however, one of the factors - capital - is completely mobile. In essence, the technology function in this paper is a hybrid, combining elements of various research [Martins and Roger; Melitz; BRS] to incorporate various useful properties.

### iii Capital Market

This paper abstracts any capital market considerations by assuming there is a well functioning capital market such that capital is transferred from owners to firms, and rewards are transferred costlessly back to owners for consumption.

### iv Normalisation of Prices

As $a$ good is costlessly traded across the two regions, wage rate $w$ is therefore equalised. Moreover, as capital is freely mobile across, the rental rate $r$ is also equalised. The cost of the composite input $\kappa$ - which depends on $r$ and $w$ - will therefore be also equalised between the two regions (in an interior solution). Cost minimisation, together with the normalisation of the price of $\kappa$ to 1, gives the following identity

$$\frac{w^{1-\alpha}r^{\alpha}}{A}\left[\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}+\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}\right]\equiv 1 \tag{1a}$$

This identity allows interest rate to be expressed in terms of wage rate and parameters (or vice versa). The advantage of normalising the price of $\kappa$ (rather than wages) is that it allows all equilibrium conditions for the manufacturing firms to be written in terms of $\kappa$ only, without having to deal with the cost minimising price function of $\kappa$.

Furthermore, cost minimisation implies that the rental to wage ratio is given as

$$\frac{r}{w}=\left(\frac{\alpha}{1-\alpha}\right)\frac{L_m}{K_N+K_S} \tag{1b}$$

where $L_m$ is the total labour used in manufacturing. Equation (1a) allows $r$ to be expressed as a function of $w$ and parameters. Substituting this into equation (1b), one can express the labour to capital ratio as a function of $w$ only.

## D Pareto Productivity Distributions

All manufacturing firms face a similar ex-ante distribution of productivity in each location. This paper assumes pareto distributions for productivities in both North and South [Helpman, Melitz and Yeaple (2005); BRS (2004); Baldwin and Okubo (2006)][8]. The parameters for the North are $\varphi_{N\min}$ and $k_N$, where $\varphi_{N\min}$ specifies the minimum support and $k_N$ the shape of the distribution. The corresponding parameters for the South are $\varphi_{S\min}$ and $k_S$.

## E Fixed Cost

Firms trying to enter the manufactured goods market are required to pay a sunk cost of $f_e$ (again in terms of $\kappa$) to draw the firm specific productivity $\varphi$. As capital is completely mobile, a firm can choose to pay this cost either at North or at South, upon which its productivity will be drawn from the respective distribution. The paper assumes that firms are not allowed to relocate their investment once they have selected on the initial location[9].

[8] The relevant cumulative density, probability density, mean and variance are given as

$$G(\varphi) = 1 - \left(\frac{\varphi_m}{\varphi}\right)^k \qquad g(\varphi) = \frac{k\varphi_m^k}{\varphi^{k+1}}$$

$$E(\varphi) = \frac{k\varphi_m}{k-1} \qquad Var(\varphi) = \frac{\varphi_m^2 k}{(k-1)^2(k-2)}$$

where $k > 2$ and $\varphi_m > 0$. For a pareto distribution, both mean and variance is decreasing in $k$.

[9] If both locations have the same ex-ante productivity distribution, no firms will relocate in equilibrium since the cutoffs are the same. An atomistic firm will have the same expected profits in both locations. If the productivity distributions are different, it becomes conceptually problematic to consider the effects of relocation since it becomes unclear which productivity should be assumed - the initial location or the new location.

## F Trade Cost

Trade in the manufacturing sector is costly. There is a $\tau > 1$ iceberg trade cost for every unit shipped. In addition, exporters will have to incur a beachhead, or a fixed export cost $f_X$ in order to export. Both costs are in terms of $\kappa$, paid in the home country. To achieve the partition of domestic producers and exporters, assume that the partition condition always holds $f < \tau^{\sigma-1} f_X$ [see Melitz (2003) for detailed derivation].

# III Trade Equilibrium

As usual, the $a$ sector equalises wages between the two locations

$$w = p_a = p_a^* = w^*$$

where Southern variables are denoted with the asterisk.

## A Average Profits in Manufacturing

As in Melitz (2003) and BRS (2004), the average profits in the North can be written as

$$\bar{\pi}_N = \left[ \left( \frac{\tilde{\varphi}_N}{\varphi_N^*} \right)^{\sigma-1} - 1 \right] f + \left( \frac{\varphi_N^*}{\varphi_{NX}^*} \right)^{k_N} \left[ \left( \frac{\tilde{\varphi}_{NX}}{\varphi_{NX}^*} \right)^{\sigma-1} - 1 \right] f_X \tag{2N}$$

where $\varphi_N^*$ is the cutoff productivity for entry, $\tilde{\varphi}_N$ the average productivity of all Northern firms above the cutoff, $\varphi_{NX}^*$ the cutoff productivity into export, and $\tilde{\varphi}_{NX}$ is the average productivity of Northern exporters. Defining $PNX = \left( \frac{\varphi_N^*}{\varphi_{NX}^*} \right)^{k_N}$, this gives the conditional probability of having a high enough productivity to export. The analogous expression for the South is

$$\bar{\pi}_S = \left[ \left( \frac{\tilde{\varphi}_S}{\varphi_S^*} \right)^{\sigma-1} - 1 \right] f + \left( \frac{\varphi_S^*}{\varphi_{SX}^*} \right)^{k_S} \left[ \left( \frac{\tilde{\varphi}_{SX}}{\varphi_{SX}^*} \right)^{\sigma-1} - 1 \right] f_X \tag{2S}$$

where $PSX = \left(\frac{\varphi_S^*}{\varphi_{SX}^*}\right)^{k_S}$ is the conditional probability of export in the South.

The marginal firms in the North and South, with productivities $\varphi_N^*$ and $\varphi_S^*$, recover only fixed cost in equilibrium. This gives the following relationship

$$\left(\frac{\sigma}{\sigma-1}\frac{1}{\varphi_N^*}\right)^{1-\sigma}\left[\frac{\mu E_N}{P^{1-\sigma}}\right] = \sigma f = \left(\frac{\sigma}{\sigma-1}\frac{1}{\varphi_S^*}\right)^{1-\sigma}\left[\frac{\mu E_S}{P^{*1-\sigma}}\right] \quad (3)$$

## B Productivities of Northern and Southern Firms

As with the usual derivation in such models, average productivities of Northern and Southern firms - $\tilde{\varphi}_N$ and $\tilde{\varphi}_S$ - are functions of the respective cutoffs only[10]. The pareto productivity distributions allow the ratios between the average productivities and their respective cutoffs to be written as a function of parameters only

$$\left(\frac{\tilde{\varphi}_N}{\varphi_N^*}\right)^{\sigma-1} = \left[\frac{k_N}{k_N+1-\sigma}\right] \qquad \left(\frac{\tilde{\varphi}_S}{\varphi_S^*}\right)^{\sigma-1} = \left[\frac{k_S}{k_S+1-\sigma}\right] \quad (4a)$$

It is also clear from this derivation that further parameter restrictions need to be in place, namely $k_S + 1 - \sigma > 0$ and $k_N + 1 - \sigma > 0$.

Furthermore, the average productivities of exporters are the conditional expectations above the respective export cutoffs. With the pareto distribution, the following relationships can also be established

$$\left(\frac{\tilde{\varphi}_{NX}}{\varphi_{NX}^*}\right)^{\sigma-1} = \left[\frac{k_N}{k_N+1-\sigma}\right] \qquad \left(\frac{\tilde{\varphi}_{SX}}{\varphi_{SX}^*}\right)^{\sigma-1} = \left[\frac{k_S}{k_S+1-\sigma}\right] \quad (4b)$$

Together, equations (4a) and (4b) give the extremely useful result that $\frac{\tilde{\varphi}_N}{\varphi_N^*} = \frac{\tilde{\varphi}_{NX}}{\varphi_{NX}^*}$ with the pareto distributions. Though exporters have a higher average productivity, the ratio of average productivity of all producers to the entry cutoff is exactly the same as the ratio

[10] These are $\tilde{\varphi}_N = \left[\frac{1}{1-G_N(\varphi_N^*)}\int_{\varphi_N^*}^{\infty}\varphi^{\sigma-1}g_N(\varphi)d\varphi\right]^{\frac{1}{\sigma-1}}$ and $\tilde{\varphi}_S = \left[\frac{1}{1-G_S(\varphi_S^*)}\int_{\varphi_S^*}^{\infty}\varphi^{\sigma-1}g_S(\varphi)d\varphi\right]^{\frac{1}{\sigma-1}}$. With the pareto distributions, these can be further simplified to $\tilde{\varphi}_N = \left[\frac{k_N}{k_N+1-\sigma}\right]^{\frac{1}{\sigma-1}}\varphi_N^*$ and $\tilde{\varphi}_S = \left[\frac{k_S}{k_S+1-\sigma}\right]^{\frac{1}{\sigma-1}}\varphi_S^*$.

of average productivity of all exporters to the export cutoff. A similar expression holds for the South as well. The equations in (4a) and (4b) can then be used to simplify the profits expressions in equations (2N) and (2S) greatly.

Finally, a firm with $\varphi_N^*$ makes zero profits in the domestic market, while a firm with $\varphi_{NX}^*$ will make zero profits on export (holds for South as well). This allows the export cutoffs to be written as a function of domestic cutoffs only[11].

## C Aggregate Productivity and Prices

The aggregate productivity and price level in a location depend not only on domestic firms, but also on foreign firms selling there. Define the total number of active firms in the North by $M_N = n + PSX.n^*$. This says that the number of varieties in the North is made up of $n$ domestic firms and $PSX.n^*$ of Southern firms that are successful in exporting to the North. The corresponding expression for the South is $M_S = PNX.n + n^*$.

The average productivity of the North becomes the weighted average of productivities of Northern firms and Southern exporters

$$\tilde{\varphi} = \left[ \frac{1}{M_N} \left( n\tilde{\varphi}_N^{\sigma-1} + PSX.n^* \phi \tilde{\varphi}_{SX}^{\sigma-1} \right) \right]^{\frac{1}{\sigma-1}} \tag{5N}$$

where $\phi = \tau^{1-\sigma}$ is the freedom of trade index. The corresponding equation for the South can be written as

$$\tilde{\varphi}^* = \left[ \frac{1}{M_S} \left( PNX.n\phi \tilde{\varphi}_{NX}^{\sigma-1} + n^* \tilde{\varphi}_S^{\sigma-1} \right) \right]^{\frac{1}{\sigma-1}} \tag{5S}$$

With the definitions of productivities, the aggregate price levels in the North and South are given as

$$P = M_N^{\frac{1}{1-\sigma}} \frac{\sigma}{\sigma-1} \frac{1}{\tilde{\varphi}} \qquad\qquad P^* = M_S^{\frac{1}{1-\sigma}} \frac{\sigma}{\sigma-1} \frac{1}{\tilde{\varphi}^*} \tag{6}$$

This completes the characterisation of the aggregate price levels for both locations. The

[11] The respective export cutoffs are a function of production cutoffs and parameters only, with $\varphi_{NX}^* = \varphi_N^* \tau \left(\frac{f_x}{f}\right)^{\frac{1}{\sigma-1}}$ and $\varphi_{SX}^* = \varphi_S^* \tau \left(\frac{f_x}{f}\right)^{\frac{1}{\sigma-1}}$.

aggregate prices $P$ and $P^*$ in equation (6) can also be substituted into the marginal firm conditions in equation (3).

## D Equilibrium Conditions

### i Equalisation of Expected Values of Entry in North and South

Free entry ensures that the ex-ante value of entry must be equal for both locations if there is to be an interior solution (with manufacturing firms in both locations)[12]

$$\left(\frac{\varphi_{\min N}}{\varphi_N^*}\right)^{k_N}\bar{\pi}_N = f_e \qquad\qquad \left(\frac{\varphi_{\min S}}{\varphi_S^*}\right)^{k_S}\bar{\pi}_S = fe$$

With the appropriate substitutions of (4a) and (4b) into equations (2N) and (2S), the above expression can be explicitly written as

$$\left(\frac{\varphi_{\min N}}{\varphi_N^*}\right)^{k_N}\left(\frac{\sigma-1}{k_N+1-\sigma}\right)\left[f+\left(\frac{\varphi_N^*}{\varphi_{NX}^*}\right)^{k_N}f_X\right]=f_e$$

$$\left(\frac{\varphi_{\min S}}{\varphi_S^*}\right)^{k_S}\left(\frac{\sigma-1}{k_S+1-\sigma}\right)\left[f+\left(\frac{\varphi_S^*}{\varphi_{SX}^*}\right)^{k_S}f_X\right]=f_e \tag{7}$$

### ii Market Clearing

There are in equilibrium $n$ successful entry firms in the North and $n^*$ in the South. But due to the cutoffs, the number of firms that attempt entry has to be higher. The total number of firms that attempt entry, including those below the cutoffs, are

$$n_e = \frac{n}{1-G_N(\varphi_N^*)} \qquad\qquad n_e^* = \frac{n^*}{1-G_S(\varphi_S^*)}$$

where $n_e$ and $n_e^*$ are the total number of entry attempts in the North and South respectively.

The composite input $\kappa$ is used for four purposes - sunk cost ($f_e$), fixed production

[12] In a corner solution (that is, full agglomeration), one of the equality will not hold.

cost ($f$), marginal production cost, and export costs (this is incurred by exporters only). The key to note here is that even unsuccessful entrants will use up industrial inputs. The marginal cost for each firm is $\frac{1}{\varphi}$, a firm-specific variable.

The aggregate production cost (or total cost) in the North can be written as $n\left(\frac{k_N}{k_N+1-\sigma}\right)(\sigma-1)f$ [see Appendix: Deriving total resource cost]. Aggregate composite input used in the North becomes

$$\kappa_N = n\left\{\begin{array}{c} f+\left(\frac{k_N}{k_N+1-\sigma}\right)(\sigma-1)f+\frac{f_e}{1-G_N(\varphi_N^*)} \\ +PNX\left[f_X+\left(\frac{k_N}{k_N+1-\sigma}\right)(\sigma-1)\frac{f_X}{\tau}\right] \end{array}\right\}$$

Multiplied by the number of firms, the first term within the brackets on the right hand side is the total fixed production cost. The second term on the right is the aggregate cost of all firms. The third term is the total sunk cost incurred, including that of the unsuccessful firms. Finally, the terms inside the square brackets are the beachhead cost and export production cost, which are incurred by exporters only. An analogous term can be written for the South

$$\kappa_S = n^*\left\{\begin{array}{c} f+\left(\frac{k_S}{k_S+1-\sigma}\right)(\sigma-1)f+\frac{f_e}{1-G_S(\varphi_S^*)} \\ +PSX\left[f_X+\left(\frac{k_S}{k_S+1-\sigma}\right)(\sigma-1)\frac{f_X}{\tau}\right] \end{array}\right\}$$

The above two expressions therefore give the quantity of the composite input $\kappa$ demanded in the North and South respectively.

Due to the cost minimisation property, the conditional demand for capital is $K = \frac{1}{A}\left(\frac{w}{r}\right)^{1-\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\kappa$. By substituting the demands of $\kappa$ into the appropriate conditional demands, one can derive the demands of the primary factors capital and labour. Since the total demand of capital in the world must be equal to the endowment, the capital clearing condition can be written as

$$\bar{K}_W = \frac{1}{A}\left(\frac{w}{r}\right)^{1-\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}(\kappa_N+\kappa_S) \tag{8}$$

Equation (8) converts the industrial inputs into capital by substituting $\kappa$ into the appropriate cost minimising function. Similarly, since the conditional demand for labour (for manufacturing) can be written as $L = \frac{1}{A}\left(\frac{r}{w}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}\kappa$, total labour requirement for manufacturing becomes

$$L_m = \frac{1}{A}\left(\frac{r}{w}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}(\kappa_N + \kappa_S) \tag{9}$$

As labour is also used for agriculture, the total manufacturing labour does not equal to total labour endowment. Instead, the amount of labour available for agriculture is whatever labour not used in manufacturing. This has to be equal to the real demand for agricultural goods (nominal expenditure divided by the price of agriculture goods, which is $w$), giving the agricultural market (or labour market) clearing condition

$$\bar{L}_W - L_m = \frac{(1-\mu)\left[E_N + E_S\right]}{w} \tag{10}$$

Equations (8) and (10) therefore provide the two market clearing conditions for capital and agriculture goods respectively. Since these two markets clear, the local labour markets will also clear.

### iii Aggregate Expenditure

As owners of capital are immobile, all capital returns are remitted to the owners and consumed locally. The aggregate expenditures for the North and South are simply their respective factor endowments multiplied by the rental and wage rates

$$E_N = rK_N + wL_N \qquad\qquad E_S = rK_S + wL_S$$

# IV Numerical Solution[13]

The endogenous variables for equilibrium are $\{w, \varphi_N^*, \varphi_S^*, n, n^*\}$ - although interest rate is endogenous, it can be recovered by the identity in equation (1). For the five endogenous variables, the equilibrium is pinned down by (i) two free entry conditions in equation (7); (ii) zero profit condition in equation (3); and two market clearing conditions in equation (8) and (10).

This paper does not make any empirical estimates on any parameters. Instead, parameters on preferences and pareto distribution are taken from existing research. The choice of endowment is arbitrary. However, the same level of endowment is chosen for the North and South in order not to introduce the home market effect that would otherwise be evident in an economic geography model. This assumption will be relaxed later in sub-section c to bring out the home market effect. The list of parameters is provided in the Appendix. The set of cost parameters is $\{f, f_e, f_X, \tau\}$. These parameters will also be varied in various numerical solutions to highlight the effects of changes in them.

Finally, note that there is a distinction between the share of firms in a location and the share of capital deployed. In the traditional FC model, there is no difference between the two since the deployment of capital leads to firm formation in a direct manner depending on fixed production cost. Due to firm heterogeneity here, the cutoff conditions may be different in both locations.

## A The Impact of Higher Firm Productivity

In the first set of numerical solutions, North and South have the same distribution shape $k_N = k_S = 3.6$. However, North is given a better productivity compared to the baseline scenario, $\varphi_{N\,\min} = 0.21 > \varphi_{S\,\min} = 0.2$. This shifts the North's distribution rightwards

[13] Numerical solutions are obtained through MATLAB. An initial guess is provided for all the variables. The endogeneous variables are then solved through the equilibrium conditions, and incremental updates in each round is carried out by taking the weighted average between the 'old' and 'new' solutions, until there are no further changes (convergence). The solution method is similar to BRS (2004). I am grateful to Stephen Redding for sharing the MATLAB codes.

(first degree stochastic dominance). With this specification, the unconditional mean productivity implied by the parameters are $c_N = 0.291 > c_S = 0.277$. The North is only around 5 per cent more productive than the South on the basis of the unconditional mean.

Even though North and South have the same level of expenditure (given the same level of endowment), the slight perturbation of the pareto distributions resulted in dramatic differences in industry location. The equilibrium effects on industrial concentration are presented in Table 1A for three different levels of trade cost.

Table 1A: Share of Firms and Capital in the North

| *f*= 10 *fe*= 10 *fx*= 10 | τ = 1.3 | τ = 1.2 | τ = 1.1 |
|---|---|---|---|
| Share of firms in the North | 0.806 | 0.964 | 1 |
| Share of firms in the South | 0.806 | 0.964 | 1 |

The full equilibrium results with the medium level of trade cost $\tau = 1.2$ are presented in Table 1B.

Table 1B: Aggregate Variables in Equilibrium

| τ = 1.2 | North | South |
|---|---|---|
| Cutoff productivity | 0.304 | 0.290 |
| Probability of successful entry | 0.294 | 0.294 |
| Average firms' productivity, conditional on entry | 0.520 | 0.496 |
| Aggregate price level | 0.648 | 0.680 |

The results show that a small increase in the mean of the productivity distribution in the North can have a dramatic impact on the location of firms. A 5 per cent increase in the unconditional mean of the productivity distribution creates 'catastrophic' concentration of industrial activity in the North at low levels of trade cost (Table 1A). Even with high trade cost, industries are still heavily concentrated in the North. The intuition becomes clear in Table 1B. The better productivity distribution in the North means that firms there are more productive and profitable. More firms need to move there until the effects of local market competition cancel out any productivity advantages.

Another striking feature of this equilibrium is that in an interior equilibrium $r$ and $w$ are in fact the same in both locations despite a higher level of capital in the North. The South continues to have a lower aggregate $\frac{K}{L}$ ratio compared to the North, but the marginal returns to capital is the same in both North and South. The Lucas paradox disappears. The superiority of the North lies not in the aggregate production function, but is due to an improvement in firm-specific productivity draws.

Third, the fall in trade cost will accentuate the advantages of locating in the North even though the levels of expenditure are the same in each location. In the traditional FC model, if the expenditures of both locations are the same, location of firms will be symmetric at all positive levels of trade cost. The concentration of industry depends on the home market effect. In other words, trade cost is completely 'impotent' in creating asymmetric concentration when the two markets are of equal size.

This is however not the case here. Expenditure is the same in both locations, but the fall in trade cost brings about increasing concentration of industry to the North (Table 1A). The key to understanding this lies in the inspection of equations (5N) and (5S). The fall in trade cost (resulting in higher $\phi$) creates a greater increase in weighted average productivity in the South $\tilde{\varphi}^*$ compared with $\tilde{\varphi}$. Competitive pressure intensifies more quickly in the South with a fall in trade cost, thereby accentuating the advantages of locating in the North. Conversely, Northern firms are less affected by the effects of increased competition as a result of freer trade since they are more productive than their Southern counterparts.

Finally, as established by Melitz (2003) with a single factor of production, the size of the market affects only the number of firms in equilibrium, but not firm level aggregates. The same reasoning applies here. Though North and South have different cutoffs due the difference in the productivity distributions, changes in endowments do not affect respective cutoff productivities or average profits. To the atomistic firm, the supply of factors is completely elastic. The size of the endowment will determine only the number of firms in equilibrium. Relative endowment will affect the $\frac{r}{w}$ ratio required to clear the respective markets, but otherwise will also have no impact on firm level aggregates.

## B The Impact of Risk

In the previous sub-section, the North is more attractive due to its better productivity distribution. The result in equilibrium is that firm size distribution in the North dominates the South. However, suppose the South is not less productive but riskier, how will this change the distribution of capital and firms?

It is important that the impact of risk is clearly understood since one of the competing hypotheses on why relatively little capital flows to the South is the inherent riskiness in investing there (expropriation risk, political risk etc). Tybout (2000) for example notes that it is common to see very large plants existing side by side with very small ones in developing countries, even though there is little evidence to suggest plants in developing countries are inherently less productive. The author therefore suggests that this may be a result of 'uncertainty about policies . . . poor rule of law'. In this set of numerical simulation, it is precisely this effect that is being modelled by allowing the same mean productive draws but greater dispersion in the South.

In this set of numerical solutions, the South continues to be the baseline case. However, the North has the following minimum support $\varphi_{N\min} = 0.204 > \varphi_{S\min} = 0.2$. Moreover, the shape of the North's distribution is tighter with $k_N = 3.8 > k_S = 3.6$. The result of this is that the unconditional productivity means in both locations are the same with $c_N = c_S = 0.277$. However, the variance in the North is 16 per cent smaller than the South. The set of parameters in fact creates a 'means preserving spread' of the productivity distribution in the South. The South is not less productive on average, but has higher risk as characterised by the higher variance.

The numerical solution to the equilibrium with a high level of trade cost $\tau = 1.3$ and different levels of sunk cost $f_e$ are presented in Table 2A.

Table 2A: Distribution of Firms and Capital with Different Sunk Cost

| τ = 1.3 *f*= 10 *fx*= 10 | *fe*= 10 | *fe*=2 0 | *fe*=3 0 |
|---|---|---|---|
| Share of firms in the North | 0.135 | 0.207 | 0.247 |
| Share of firms in the North | 0.115 | 0.179 | 0.215 |

The full equilibrium with $\tau = 1.3$ and $f_e = 20$ is presented in the table below.

Table 2B: Aggregate Variables in Equilibrium

| τ = 1.3 | North | South |
|---|---|---|
| Cutoff productivity | 0.227 | 0.239 |
| Probability of successful entry | 0.663 | 0.531 |
| Average firms' productivity, conditional on entry | 0.366 | 0.408 |
| Aggregate price level | 0.865 | 0.824 |

The results of this sub-section show the effects of greater variance in the productivity distribution. There is more industrial concentration in the South. The higher variance in the South implies that there is a fatter right side tail for the pareto distribution. As can be seen from Table 2B, the effect of this is that though the probability of entry is lower in the South, the average productivity upon successful entry is in fact higher in the South due to the fatter right end tail.

However, the South's advantage diminishes with high sunk cost $f_e$ (see Appendix: Effects of Increasing $f_e$). After a firm invested in the sunk cost and discovers its productivity, it can decide whether to sink in the fixed production cost $f$. The sunk cost effectively becomes an option as a firm has a choice of whether or not to carry out production. At low sunk cost, the South is more attractive since it offers a greater probability of a high productivity draw (and higher average productivity). At higher sunk cost however, this option effectively becomes more expensive and reduces the attraction of the South. It has become more expensive to take a punt with the South, so to speak.

## C The Home Market Effect

The previous sub-section shows how a higher sunk cost might induce firms to the relatively less risky North. However, the results also clearly show if the mean productivities are the same, the increase in South's productivity variance alone is not the reason why capital stays in the North. All things equal, the higher variance in the South will in fact attract more capital there.

Nevertheless, numerical solutions thus far have kept the assumption that North and South have similar expenditure size and hence ignored the home market effect. By varying this and supposing the North is a larger market (while keeping the assumption that the South has the same mean but higher variance), the results show the there can be a higher concentration of industries in the North above what FC model predicts.

In the standard FC model, the share of firms in a location (reference North) is given by

$$\left(s_N - \frac{1}{2}\right) = \frac{1+\phi}{1-\phi}\left(s_E - \frac{1}{2}\right) \tag{11}$$

where $s_N$ denotes the share of firms in the North while $s_E$ denotes the share of expenditure. The home market amplification or effect is given by $\frac{1+\phi}{1-\phi}$. Using $\tau = 1.2$ and $\sigma = 3.8$ to arrive at $\phi = 0.6$ will give an amplification factor of 4. A 1 per cent increase in North's expenditure share brings about a 4 per cent increase in the share of firms.

In this sub-section, the numerical solutions retain the parameter assumptions of pareto distributions used in the previous sub-section (that is, a means preserving spread $\varphi_{N\min} = 0.204 > \varphi_{S\min} = 0.2; k_N = 3.8 > k_S = 3.6$). Investment costs are $f = f_e = f_X = 10$.

In each successive solution, the North will be allowed higher endowments, the factor of increment is the same for capital and labour. The factor is shown in column 1 of Table 3. In equilibrium therefore, the North will have a greater share of expenditure in each successive set of solutions. The expenditure share is then entered into equation (11) to compute the share of firms that would have occurred in the standard FC model without firm heterogeneity. The shares of firms generated by the standard FC model can then be compared to those with firm heterogeneity coupled with means preserving spread in the South.

Admittedly, this is in a sense 'naive' since the two models rest on completely different assumptions - one has heterogeneous productivity while the other does not. Nevertheless, comparison between the two serves to highlight the differences in the magnitude of the home market effect. It will also show how sunk cost can interact with the home market effect under firm heterogeneity and improve on the understanding of industrial locations. The results are presented in the following table.

Table 3: Share of Firm and Capital in Larger North

| Factor | Exp. Share | Capital Share | Firm Share | Firm Share (Standard FC) |
|---|---|---|---|---|
| 1.0 | 0.500 | 0.115 | 0.135 | 0.500 |
| 1.1 | 0.524 | 0.339 | 0.380 | 0.596 |
| 1.2 | 0.545 | 0.530 | 0.574 | 0.684 |
| 1.3 | 0.565 | 0.695 | 0.731 | 0.760 |
| 1.4 | 0.583 | 0.838 | 0.861 | 0.832 |
| 1.5 | 0.600 | 0.964 | 0.970 | 0.900 |

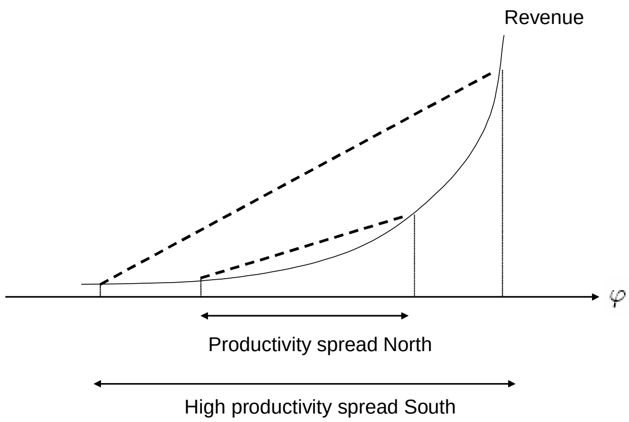
Column 2 gives the expenditure share of the North, column 4 the share of firms in the heterogeneity setup with a means preserving increase in variance in the South, and column 5 the share predicted by the standard FC model in equation (11).

When the market size of the North is the same or slightly bigger than the South, the share of firms in the North is lower than the standard FC model. The South is more attractive because again because of the effect of the fat right end tails. But as the size of the North's economy becomes larger in each successive solution, more firms concentrate in the North and the share eventually becomes larger than what FC predicts.

Furthermore, a higher sunk cost $f_e$ will increase the values of all the entries in column 4. In other words, a higher sunk cost will bring about increased concentration of industries in the larger North. Though it is difficult to solve the model analytically, one can still sketch out the mechanics of the model and distil the insight.

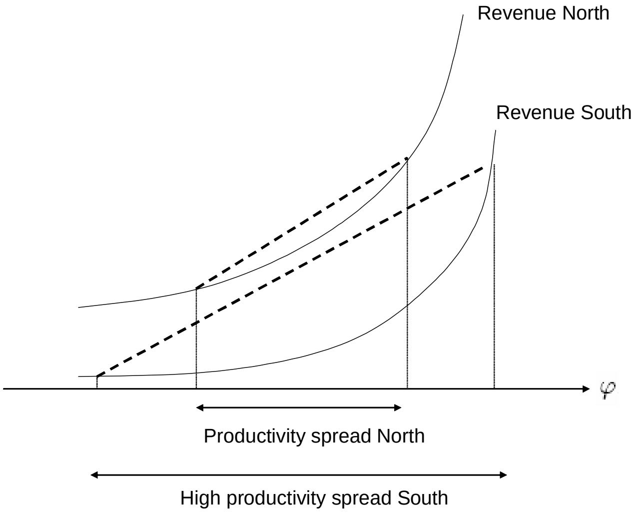
Suppose at first that North and South are of equal size in terms of expenditure, and have the same pareto distributions. In equilibrium, they will attract the same number of firms. But suppose now that South's distribution is perturbed with a higher spread of productivity draws. The convexity of the revenue function then implies that expected profits is higher in the South. In order for North-South profits to be equalised, more firms will have to enter the South in order to equalise profits again. This brings about increased industrial concentration in the South, and is in a sense the paradox of risk. This is the effect seen in Figure 1.

Figure 1: Expected Profits with Higher Productivity Spread



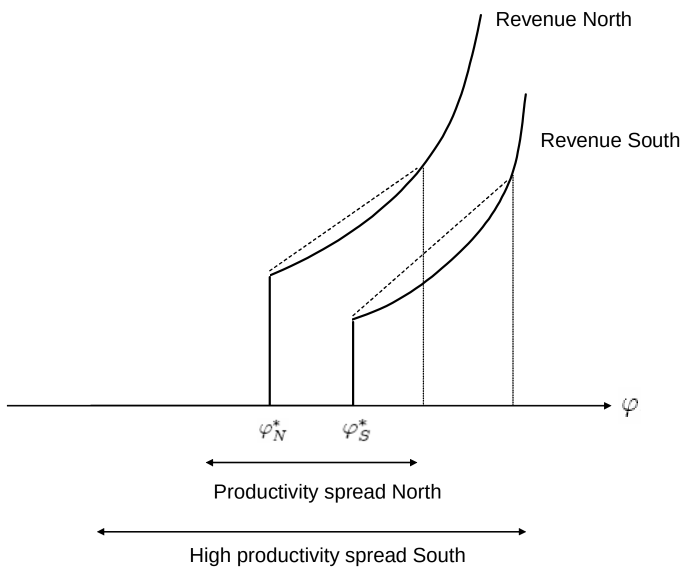

However, if the North is larger in terms of endowment, and hence has higher expenditure, the revenue effect of a larger market begins can offset the effect of South's higher productivity spread. Revenue function for the North shifts up (Figure 2). The result is that more firms will move to the North to equalise profits between the two locations.

Figure 2: Higher Productivity Spread in South but Larger Market In North




That is not yet the complete story. A larger expenditure in the North will always result in more firms locating there due to the home market effect. This has long been understood in NEG literature. What is new and more interesting, in the firm heterogeneity case, is how sunk cost $f_e$ interacts with the home market effect. The effect of sunk cost can be seen in Figure 3. In the firm heterogeneity setup with sunk cost, productivity cutoffs truncate the revenue functions. The effect of higher $f_e$ is always to push the productivity cutoffs (both North and South) leftwards. Because the pareto distributions are different, the effect is asymmetric.

Figure 3: Effect on Cutoffs on Revenue Functions




The higher the sunk cost, the more attractive the North becomes. The narrower shape of North's pareto distribution ensures that the probability of successful entry in the North increases relatively more quickly than the South when $f_e$ increases. On the other hand, cutoffs shift to the right if $f_e$ falls. This reduces North's expected profits more than the South, thereby shifting industries towards the latter.

The effects of this can be seen in Table 4, where two sets of equilibrium solutions are presented with different $f_e$ and the scenario that the North is 20 per cent larger in terms of market size[14]. Compare the entries in rows 4 and 5. Successive increase in $f_e$ raises the probability of entry in the North more quickly than the South as the North has a tighter productivity distribution.

[14]This corresponds to the third row (factor=1.2) in Table 3. All parameters remain the same as of the rest of the sub-section in this set of numerical solutions.

Table 4: Equilibrium Effects of a 20 Per cent Larger North at Higher Sunk Cost

| | *fe*= 10 | *fe*= 20 | *fe*= 30 |
|---|---|---|---|
| Expenditure share of North | 0.545 | 0.545 | 0.545 |
| Cutoff North | 0.273 | 0.227 | 0.204 |
| Cutoff South | 0.289 | 0.239 | 0.213 |
| Probability of Entry North | 0.332 | 0.663 | 0.995 |
| Probability of Entry South | 0.265 | 0.531 | 0.796 |
| Average Productivity of North Firms | 0.439 | 0.366 | 0.329 |
| Average Productivity of South Firms | 0.495 | 0.408 | 0.365 |
| Firm Share North | 0.574 | 0.625 | 0.654 |
| Capital Share North | 0.53 | 0.583 | 0.613 |

Therefore, the rise in sunk cost increases the concentration of firms and capital in the North. Eventually, the concentration becomes larger than the home market effect of the traditional FC model. The economic intuition is subtle. If the North is a large enough market, industries with high sunk cost continue to stick with the North because it offers them a greater probability of entry compared to the South. The higher the sunk cost, the more attractive the North becomes. Sunk cost and market size then reinforce each other to create a concentration effect greater than the home market effect only.

If manufacturing is 'hi-tech', as characterised by high sunk cost, they will remain sticky in the larger Northern economy even when it is open to trade with the South. In this equilibrium, capital from the South move northwards, because the South offers a lower probability of entry for these high sunk cost industries. On the other hand, if manufacturing is 'low-tech' with low sunk cost, firms will move southwards even though the North remains the larger market. The model can therefore rationalise capital flows in either direction, depending on the level of sunk cost and the size of the markets[15].

The role of trade cost also becomes clear. A higher trade cost $\tau$ makes each market

[15]This is not formally modelled in this paper, but the intuition should be clear. One can for example expand the Cobb Douglas preferences to include more manufacturing industries with different sunk costs. The high sunk cost industries will move to the less risky location while the low sunk cost industries will move to the more risky one.

increasingly segmented, and firm share will be closely related to expenditure share. If trade cost is high, even 'low-tech' industries cannot move South, and have to be based located where the expenditure is. The increase in beachhead cost also has the same effect since it reduces export probability and makes markets more segmented. Obviously, the uneven flow of capital to the South is a highly complex puzzle. Without claiming to be the sole explanation, this model does somewhat provide an understanding for example of why China has received more FDI capital than India. If trade is free enough, 'low-tech' industries and capital leave the larger Northern market, produce in the South, and export products back to the North. As the South has higher productivity variance and a higher cutoff, the few successful entry firms there are in fact highly productive and large. This also fits some stylised fact observed about the uneven flow of capital to different emerging economies.

# V Conclusion

By synthesizing a variant of New Economic Geography model with recent research into the effects of trade equilibrium under firm heterogeneity, this paper shows it is possible to rationalise the highly asymmetric allocation of capital between North and South without stipulating large differences in productivity between the two locations.

Introducing firm heterogeneity allows the differences between North and South to be modelled by way of firm level differences rather than through the aggregate production function. With a slight improvement in the distribution in the ex-ante productivity distribution in the North, this paper demonstrates that it is possible to explain the high concentration of firms (and capital) to the North, even as returns to factors of production and expenditures are completely identical between the regions. The Lucas paradox disappears as a result.

The second key result of the paper demonstrates how the home market effect can interact with sunk cost to resolve the paradox of risk, and create highly asymmetric industrial locations. 'Hi-tech' or high sunk cost industries tend to locate in the less risky North because it offers them a greater probability of successful entry relative to the South. For 'low-tech' industries with low sunk cost, the North is less attractive since the increase in the probability of entry is offset by the potential of higher post-entry productivity in the South. Capital flows in both directions can be rationalised by the level of sunk cost.

The paper also shows how the level of capital flows also depends crucially on the level of trade cost. If trade costs are high, capital will to a large extent be distributed according to expenditure shares. With low trade cost, 'low-tech' industries will locate in the South. This can then explain some stylised differences amongst the flow of capital to different developing economies. Developing countries with less trade restrictions will receive more capital particularly from 'low-tech' industries.

## References


[1] **Alfaro, Laura; Kalemli-Ozcan, Sebnem; Volosovych, Vadym** (2005), "Why Doesn't Capital Flow from Rich to Poor Countries? An Empirical Investigation", NBER Working Paper No.11901.

[2] **Balwin, Richard E.; Forslid, Richard; Martin, Phillippe; Ottaviano, Gianmarco I.P.; Robert-Nicoud, Frederic.**, "Economic Geography & Public Policy", Princeton University Press.

[3] **Baldwin, Richard E. and Okubo, Toshihiro** (2005), "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting", Journal of Economic Geography 6, pp. 323-346.

[4] **Baldwin, Richard E. and Okubo, Toshihiro** (2006), "Agglomeration, Offshoring and Heterogeneous Firms", CEPR Discussion Paper DP5663.

[5] **Banerjee, Abhijit V. and Duflo, Esther** (2004), "Growth Theory through the Lens of Development Economics", Handbook of Development Economics, Vol. 1a, pp. 473-552

[6] **Bernard, Andrew B.; Redding, J. Stephen and Schott, Peter K.** (2004), "Comparative Advantage and Heterogeneous Firms", NBER Working Paper 10668.

[7] **Bernard, Andrew B.; Eaton, Jonathan; Jensen, J. Bradford and Kortum, Samuel** (2003), "Plants and Productivity in International Trade".

[8] **Caselli, Francesco and Feyrer, James** (2006), "The Marginal Product of Capital", London School of Economics, yet to be published.

[9] **Dixit, Avinash K. and Stiglitz, Joseph E.** (1977), "Monopolistic Competition and Optimum Product Diversity", The American Economic Review, 67(3), pp 297 - 308.

[10] **Ghironi, Fabio and Melitz, Marc J.** (2004), "International Trade and Macroeconomic Dynamics with Heterogeneous Firms", CEPR Discussion Paper No. 4595.

[11] **Griffith-Jones, Stephany and Leape, Jonathan** (2002), "Capital Flows to Developing Countries: Does the Emperor Have Clothes?", QEH Working Paper Series, Working Paper Number 89.

[12] **Harrigan, James. and Venables, Anthony J.** (2004), "Timeliness, Trade and Agglomeration", Centre of Economic Performance (LSE).

[13] **Helpman, Elhanan; Melitz, Marc J. and Yeaple. Stephen R.** (2004), "Export versus FDI with Heterogeneous Firms", The American Economic Review, Vol 94 No.1, pp. 300 - 317.

[14] **Krugman, Paul R.**(1979), "Increasing Returns, Monopolistic Competition, and International Trade", Journal of International Trade, 9(4), pp. 469 – 479.

[15] **Krugman, Paul R.**(1980), "Scale Economies, Product Differentiation, and Pattern of Trade", The American Economic Review 70, pp. 950 – 959.

[16] **Krugman, Paul R.**(1991), "Increasing Returns and Economic Geography", Journal of Political Economy 99, pp. 483 – 499.

[17] **Krugman, Paul R. and Venables, Anthony J.** (1995), "Globalization and the Inequality of Nations", The Quarterly Journal of Economics, Vol. 110, No.4, pp 857-880.

[18] **Martin, Philippe and Rogers, Carol Ann**, "Industrial Location and Public Infrastructure", Journal of International Economics 39, pp.355 - 351.

[19] **Melitz, Marc J.** (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," Econometrica 71, pp. 1695-1725.

[20] **Melitz, Marc J. and Ottaviano, Gianmarco I.P** (2005), "Market Size, Trade, and Productivity", NBER Working Paper No.11393.

[21] **Redding, Stephen J. and Venables, Anthony J**. (2003), "Economic Geography and International Inequality", Journal of International Economics Vol 62, pp 53-82.

[22] **Reinhart, Carmen and Rogoff, Kenneth S.** (2004), "Serial Default and the "Paradox" of Rich to Poor Capital Flow", NBER Working Paper No. 10296.

[23] **Tornell, Aaron and Velasco Andres** (1992), "The Tragedy of the Commons and Economic Growth: Why Does Capital Flow from Poor to Rich Countries", The Journal of Political Economy, Vol 100, No.6, pp.1208-1231.

[24] **Tybout, James R.** (2000), "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?", Journal of Economic Literature, Vol. 38, No.1, pp 11-44.

[25] **Venables, Anthony J**. (1996), "Equilibrium Locations of Vertically Linked Industries", International Economic Review, Vol 37, No.2, pp. 341-359.

## Appendix

### Calibration of Numerical Simulation

Parameter values are referenced to various research where possible. The list of parameters is given in the table below.

Table A1: List of Parameter Values

| Parameters | Value | Remarks |
|---|---|---|
| **Preferences** | | |
| $\sigma$ | 3.8 | Referenced to Bernard at el (2003), Ghironi and Melitz (2004) and BRS (2004). |
| $\mu$ | 0.8 | Arbitrary, no effect on firm aggregates or distribution of firms between the locations. |
| **Endowment**<br>$K_N$<br>$L_N$<br>$K_S$<br>$L_S$ | 10000, unless otherwise stated | Endowments are symmetric between North and South except for Section IV (C) where the home market effect is modelled by increasing North's endowment. |
| **Pareto Distribution** | | |
| $\varphi_{min}$ | 0.2 | Referenced to BRS (2004). This is the baseline support for any pareto distribution. In Section IV (A), the North's support is increased to 0.21. In Sections IV (B) and (C), the North's support becomes 0.204. |
| k | 3.6 | Referenced to BRS (2004). This is the baseline shape for the distribution. In Section IV (B) and (C), the North is given a tighter shape with 3.8. |
| **Technology** | | |
| A | 1 | Aggregate productivity is normalised to unity for convenience. |
| $\alpha$ | 0.3 | This is the capital share in the production of the composite input. Its effect is only on the wage-rental ratio, and has no effect on distribution of firms. |

**Deriving total resource cost**. Consider a standard total cost ($TC$) function. This is the integration of the resources used by each firm $\frac{q(\varphi)}{\varphi}$ over the entire distribution of active firms

$$\begin{aligned} TC &= \int_{\varphi^*}^{\infty} n \frac{q(\varphi)}{\varphi} \frac{g(\varphi)}{1-G(\varphi^*)} d\varphi \\ &= nq(\tilde{\varphi}) \int_{\varphi^*}^{\infty} \frac{1}{\varphi} \left(\frac{\varphi}{\tilde{\varphi}}\right)^{\sigma} \frac{g(\varphi)}{1-G(\varphi^*)} d\varphi \end{aligned}$$

The second equality makes use of the property that $q(\varphi) = q(\tilde{\varphi}) \left(\frac{\varphi}{\tilde{\varphi}}\right)^{\sigma}$. With the pareto distribution and the definition of $q(\varphi)$, the above equation can then be simplified to

$$\begin{aligned} TC &= nq(\tilde{\varphi}) \frac{k\varphi^{*k}}{\tilde{\varphi}^{\sigma}} \int_{\varphi^*}^{\infty} \varphi^{\sigma-k-2} d\varphi \\ &= n \left[\frac{k}{k+1-\sigma}\right] \frac{q(\varphi^*)}{\varphi^*} \end{aligned}$$

Total production cost is a $n\left[\frac{k}{k+1-\sigma}\right]$ factor of the production cost of the marginal firm $\frac{q(\varphi^*)}{\varphi^*}$. Multiplying the numerator and denominator by $p(\varphi^*)$ will give $\frac{q(\varphi^*)}{\varphi^*} = \frac{p(\varphi^*)q(\varphi^*)}{p(\varphi^*)\varphi^*} = \frac{r(\varphi^*)}{p(\varphi^*)\varphi^*}$. Since the marginal firm's revenue $r(\varphi^*)$ must cover $\sigma f$ in equilibrium, and its optimal price is $p(\varphi^*) = \left(\frac{\sigma}{\sigma-1}\right)\frac{1}{\varphi^*}$, it is possible to simplify the equation further to $\frac{q(\varphi^*)}{\varphi^*} = (\sigma-1)f$. This allows the total cost equation to be written as

$$TC = n\left[\frac{k}{k+1-\sigma}\right](\sigma-1)f$$

Similarly, the total cost of the exporters can be written as

$$TC_X = PNX.n\left[\frac{k}{k+1-\sigma}\right](\sigma-1)\frac{f_X}{\tau}$$

**Effects of Increasing** $f_e$. Ignoring the differences between North and South for the moment. Consider only the marginal impact of an increase in the sunk cost $f_e$. In equilibrium, the expected value of entry into a market must be equal to the cost of entry.

This gives

$$\left(\frac{\varphi_{\min}}{\varphi^*}\right)^k\left(\frac{\sigma-1}{k+1-\sigma}\right)\left[f+\left(\frac{\varphi^*}{\varphi_X^*}\right)^k f_X\right]=f_e$$

This is the generic form of equation (7). Using the relationship $\varphi_X^*=\varphi^*\tau\left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}$, the above equation can be written as

$$\left(\frac{\varphi_{\min}}{\varphi^*}\right)^k\left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f+\left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\}=f_e \tag{12}$$

The means of a pareto distribution is given as $c=\frac{k\varphi_{\min}}{k-1}$. To keep the mean constant at $c$ while allowing $k$ to vary, the minimum support has to be different. The minimum support can be written as

$$\varphi_{\min}=\frac{c(k-1)}{k} \tag{13}$$

This can be substituted into equation (12) to give

$$\varphi^{*-k}\left[\frac{c(k-1)}{k}\right]^k\left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f+\left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\}=f_e \tag{12a}$$

Partially differentiating $\varphi^*$ with respect to $f_e$ gives

$$\begin{aligned}\frac{\partial\varphi^*}{\partial f_e} &= \frac{-\varphi^{*k+1}}{k\left[\frac{c(k-1)}{k}\right]^k\left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f+\left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\}}\\ &= \frac{-\varphi^*}{k\varphi^{*-k}\left[\frac{c(k-1)}{k}\right]^k\left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f+\left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\}}\end{aligned} \tag{14}$$

In equilibrium, equation (12a) will always hold (analogous to some kind of envelope condition). This allows equation (14) to be simplified to

$$\frac{\partial\varphi^*}{\partial f_e}=\frac{-\varphi^*}{kf_e} \tag{15}$$

The result show that an increase in $f_e$ always reduces the cutoffs $\varphi^*$ - this is a standard result. But using the pareto distribution, $\frac{\partial \varphi^*}{\partial f_e}$ is more negative at lower level of $k$ (higher variance). The cutoff therefore falls relatively more quickly for the location with the lower $k$. Probability of $(PE)$ is given as

$$PE = \left[\frac{c(k-1)}{k\varphi^*}\right]^k$$

The effect of increase in $f_e$ on entry probability can be found by the partial derivative

$$\begin{aligned} \frac{\partial PE}{\partial f_e} &= \frac{\partial PE}{\partial \varphi^*}\frac{\partial \varphi^*}{\partial f_e} \\ &= \left[\frac{c(k-1)}{k}\right]^k \frac{\varphi^{*-k}}{f_e} \end{aligned} \tag{16}$$

Since $\left[\frac{c(k-1)}{k}\right]^k$ is increasing in $k$, the increase in $f_e$ therefore increases the probability of entry relatively more quickly for a location with higher $k$. The more risky location becomes less attractive.