RIETI Discussion Paper Series 17-E-001

Early Agglomeration or Late Agglomeration? Two phases of development with spatial sorting

Rikard FORSLID
Stockholm University

OKUBO Toshihiro
Keio University

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 17-E-001
January 2017

Early Agglomeration or Late Agglomeration? Two phases of development with spatial sorting[1]

Rikard FORSLID
Stockholm University

OKUBO Toshihiro
Keio University

Abstract

This paper analyzes different development paths. Developing countries that limit the geographical movement of human capital (and firms) may end up on a different equilibrium path than countries that allow for geographical mobility. At the early stages of development (when transportation costs are high), the model has an equilibrium where low productive firms concentrate in the large market with abundant human capital, whereas the most productive firms agglomerate to the smaller region with a relatively high endowment of labor. We relate this type of equilibrium to countries in an early stage of development, where industrial productivity in periphery or small suburban cities is far higher than in capital mega-cities. As economies develop and transportation costs fall, the model switches to an equilibrium where productive firms concentrate in the larger and human capital-rich region. This corresponds to a modern equilibrium where highly productive firms concentrate in the largest and most human capital-rich regions as often seen in many developed countries.

Keywords: Firm heterogeneity, Agglomeration
JEL classification: F12, F15, F21, R12

[1]This study is conducted as a part of the Project "International trade and investment" undertaken at Research Institute of Economy, Trade and Industry (RIETI). The author is grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI.

1 Introduction

It is well established that agglomerations produce rents in the form of higher productivity, wages etc. (for surveys of this literature, see Rosenthal and Strange, 2004, Melo et al., 2009 and Puga, 2010). However, agglomerations in rich and poor countries typically seem very different. Clusters in advanced economies, such as Boston, New York or London, attract high-skilled individuals and entrepreneurs, and they are characterized by high productivity and specialization in the advanced service industry. By contrast, clusters in the developing world do, to a larger extent, attract low-educated individuals, and these clusters are characterized by manufacturing or, in some cases, by urbanization without much industrialization (see e.g. Gollin et al., 2015).

In this paper, we develop a model of agglomeration and spatial sorting that is consistent with the different agglomeration processes in developed and developing countries. The model has two different development paths: The first involves agglomeration or urbanization at an early stage of development, whereas the second implies agglomeration at a later stage of development. Our analysis shows that these two agglomeration paths have very different implications for the spatial sorting pattern of firms. Whereas late agglomeration (or agglomeration in a developed economy) leads to sorting of the most productive individuals to the core region, this is not the case for early agglomeration in a development economy, where instead low-productivity individuals are attracted by the core.

In order to generate an early or development economy type of agglomeration process, we use a new theoretical framework that combines heterogeneous firms à la Melitz (2003) with the trade and geography model by Forslid (1999) and Forslid and Ottaviano (2003). The geographically mobile factor in this model is entrepreneurs, and they migrate in response to geographical differences in real returns.[1]

We will use the level of transportation costs as a proxy for the development level of a country. The trade and geography literature highlights how lower transportation costs may lead to the concentration or agglomeration of firms to larger regions (for surveys, see Baldwin et al., 2003, Fujita et al., 1999 and Fujita and Thisse, 2002). The agglomeration results are qualified once we allow for heterogenous firms. Baldwin and Okubo (2006), who introduce heterogeneous firms à la Melitz (2003) in a trade and geography model, show a pattern of spatial sorting where it is the most productive firms that tend to agglomerate to the core. The present paper, in contrast, develops a model with a richer patter of spatial sorting where in one case (agglomeration in a developed economy), the most productive entrepreneurs sort to the core, but where in another

[1] This is in contrast to the model of spatial sorting by Baldwin and Okubo (2006) and other related models (Forslid and Okubo, 2012; Baldwin and Okubo, 2014), which use the 'footloose capital' trade and geography model where physical capital moves geographically and in response to regional differences in nominal return.

case (agglomeration in a developing economy), it is the least productive entrepreneurs that sort to the core.

Our analysis is related to several strands of the development literature. Our early agglomeration case could be viewed as a formalization of so-called proto-industrialization, as proposed by Franklin Mendels (Mendels, 1972). Manufacturing in agricultural areas outside of core cities started to produce light manufacturing products to be sold in cities or exported. This process is described by Saito (1985) for Japan and by Ogilvie and Cerman (1996) for Europe.

Although proto-industrialization is one possible story, there is also more current evidence from developing countries. For example, Ghani et al. (2012) and Desmet et al. (2015) find evidence that the periphery has a stronger development than core regions in India: Desmet et al. (2015) find that areas with a high manufacturing concentration have a slower economic growth than low concentration areas between 2000 and 2005, and Ghani et al. (2012) find that the urban share of production and employment has declined for the larger firms that dominate national outputs after 1995.

Our paper is also related to the theoretical literature on skill sorting across cities. This literature shows how the sorting of individuals with different skills to different cities gives rise to a hierarchy of cities (see e.g. Nocke, 2006, Behrens et. al., 2014, and Eckhout et al., 2014). In contrast, our paper focuses on how different levels of development, proxied by the level of transportation costs, can give rise to very different processes of agglomeration and spatial sorting. Before developing the theoretical model, we present stylized empirical facts showing an example of early agglomeration using Japanese data from the interwar period.

2 Stylized evidence: Agglomeration in Japan in the interwar period

We here show stylized evidence of the agglomeration process in Japan between WWI and WWII. This is a case that displays many of the properties of agglomeration in a developing economy where, on average, relatively low-productivity activities agglomerate to the core regions (Tokyo and Osaka). The interwar period in Japan was characterized by falling transport costs and strong agglomeration. We use the extent of the railroads as a proxy for transport costs, and the extent of the railroad net doubled during this period as shown in Figure 1. There was also a strong tendency for agglomeration to the largest cities during this period.[2] Figure 2 shows how the population share of the two largest cities, Tokyo and Osaka, grew rapidly in the interwar period. However, contrary to agglomerations in advanced nations today, the concentration of activity to the largest cities in Japan, in the interwar period, did not lead to a corresponding productivity increase in these regions. Productivity measured as output per employee does not show any clear pattern of increasing more in the core (Tokyo and Osaka) than in the periphery, as shown in Figure 3, and comparing output per worker in all prefectures (in 1930), Figure 4

[2] See Settsu (2015) for detailed data from this period.

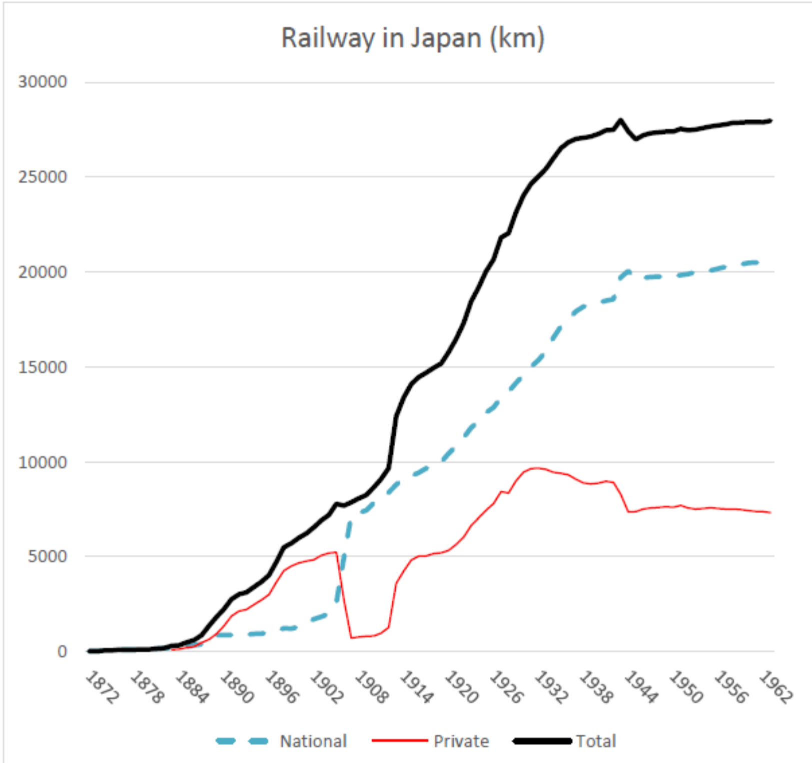

Figure 1: Railways in Japan 1872-1960. Source Minami (1965).

does not show any clear pattern of higher productivity in the core regions such as Tokyo and Osaka as compared to the smaller regions. Finally, the same picture emerges when comparing the average establishment size (employees per firm) in the core as compared to the periphery. The relative establishment size in the core declined during this period and the average firm size was actually lower in the core than in the periphery during the later years in our sample, as shown in Figure 5.

The development in Japan in the interwar period is consistent with an agglomeration process where many low skilled individuals and smaller entrepreneurs move to the core, and some relatively advanced production locates in the periphery. In the following, we will present a model that generates this type of agglomeration pattern at an early stage of development.

3 The Model

3.1 Basics

There are two regions, core and periphery (denoted by *), and two factors, human capital H and labor L. We will assume that the core region is large and rich in human capital. Human capital or entrepreneurs move between regions and bring with them their business. However, moving entails a fixed moving cost χ. In contrast, workers are not mobile between regions but can move

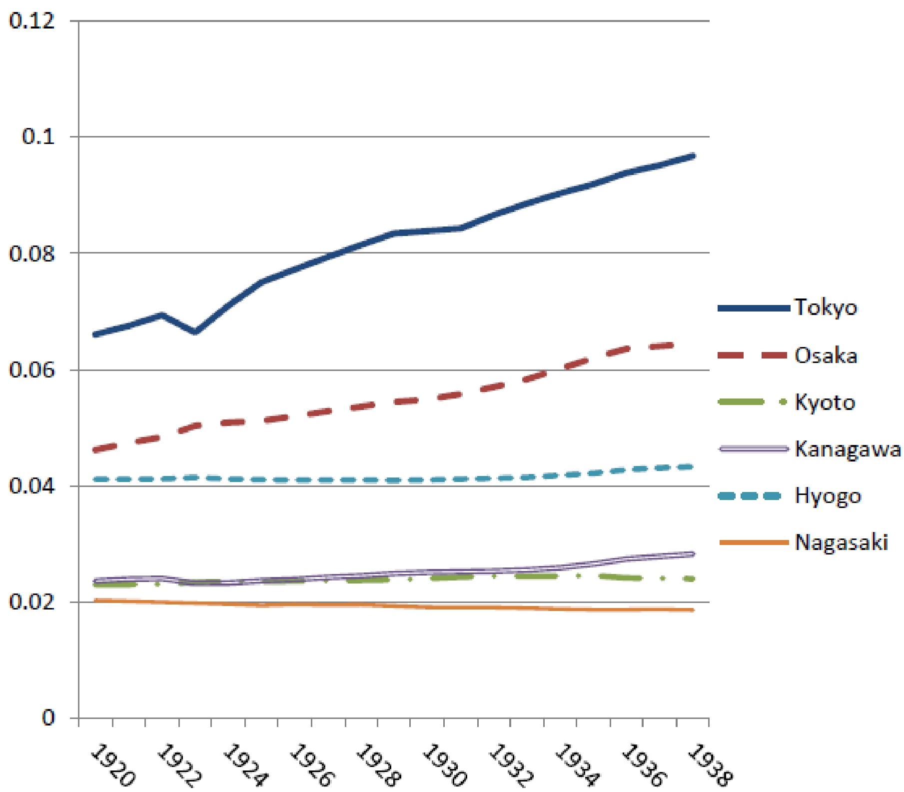

Figure 2: Share of the Janapese population living in the lagest cities. Source: Manufacturing census.

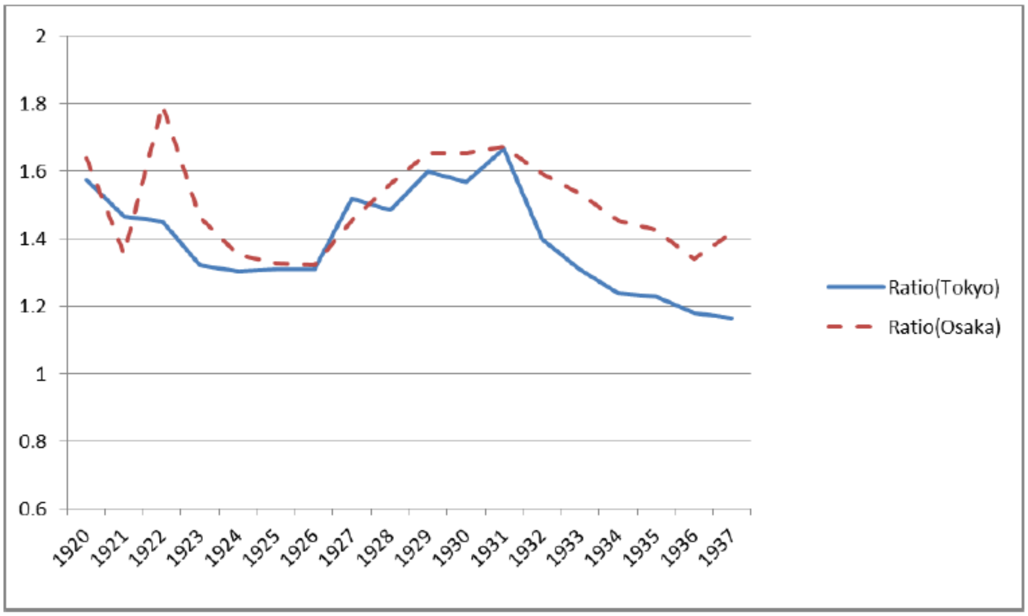

Figure 3: The ratio of average productivity (output per worker) in the core regions Tokyo and Osaka relative to the periphery.

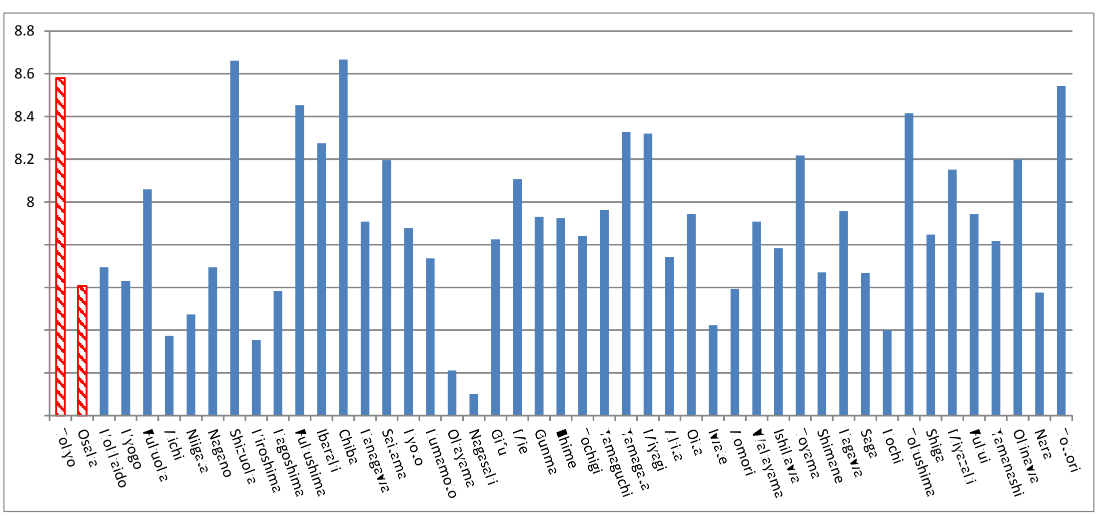

Figure 4: Output per worker 1930 (logs)

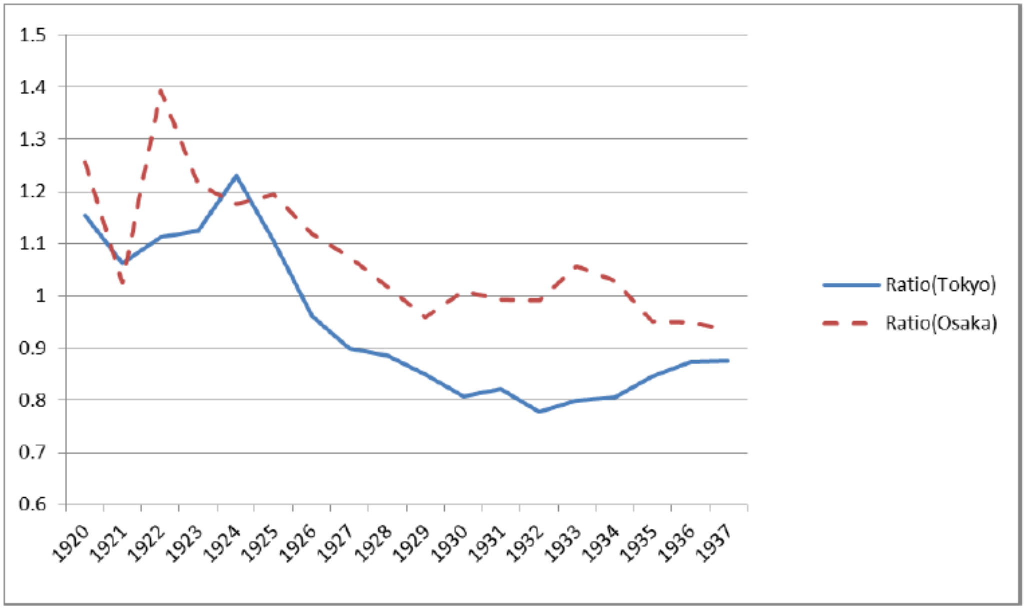

Figure 5: The ratio of averge firm size in the core regions Tokyo and Osaka relative to the periphery. Source: Manufacturing census.

freely between sectors. There are two sectors M (manufacturing) and A (agriculture). The A-sector produces a homogeneous good with a constant-returns technology only using labor. The M-sector produces differentiated manufactures with increasing-returns technologies using both human capital and labor. Firms (entrepreneurs) in the M-sector are heterogenous with respect to their marginal cost. We use quasilinear preferences as in Pfluger (2004) and Borck and Pfluger (2006), which simplifies the algebra. All individuals have the utility function

$$U = \mu \ln C_M + C_A, \tag{1}$$

where $\mu \in (0, 1)$ is a constant, and C_A is consumption of the homogenous good. Differentiated goods enter the utility function through the index C_M, defined by

$$C_M = \left[\int_0^N c_i^{(\sigma-1)/\sigma} di \right]^{\sigma/(\sigma-1)}, \tag{2}$$

N being the mass of varieties consumed, c_i the amount of variety i consumed, and $\sigma > 1$ the elasticity of substitution. Each consumer spends μ on manufactures **and** the total demand for a domestically produced variety i is therefore

$$x_i = \frac{p_i^{-\sigma}}{\int_{k=0}^{N} p_k^{1-\sigma} dk} \cdot \mu, \tag{3}$$

where p_i is the price of variety i. The unit factor requirement of the homogeneous good is one unit of labor. This good has no transportation cost, and since it is also chosen as the numeraire, we have

$$p_A = w = 1, \quad (4)$$

w being the wage of workers in all countries.

Each firm has a fixed cost in terms of human capital. We normalize the fixed cost so that one entrepreneur is associated with one firm. Firms (entrepreneurs) are differentiated in terms of their marginal cost, and the firm-specific marginal production costs a_i **are** distributed according to the cumulative distribution function $G(a)$. The total cost of producing x_i units of manufactured commodity i is

$$TC_i = \pi_i + a_i x_i, \quad (5)$$

where π_i is the return to entrepreneur i.

Distance is represented by transportation costs. Shipping the manufactured good involves a frictional transportation cost of the "iceberg" form: for one unit of good from region j to arrive in region k, $\tau_{jk} > 1$ units must be shipped. Transportation costs are also assumed to be equal in both directions so that $\tau_{jk} = \tau_{kj}$.

Profit maximization by manufacturing firms leads to the price

$$p_i = \frac{\sigma}{\sigma - 1} a_i. \quad (6)$$

Using the demand function in (3) gives the nominal return to entrepreneur i in Core:

$$\pi_i = \frac{\mu}{\sigma} \left(\frac{(L+H)}{\Delta} + \frac{\phi(L^* + H^*)}{\Delta^*} \right) a_i^{1-\sigma}, \quad (7)$$

where $a_i^{1-\sigma}$ is a measure of productivity of entrepreneur i, and

$$\Delta \equiv H \int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + \phi H^* \int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a), \quad (8)$$

where $\overline{H} = H + H^*$ is the country-level mass of varieties produced. The object $\phi_{jl} = \tau_{jl}^{1-\sigma}$, ranging between 0 and 1, stands for "freeness" of transportation between j and l (0 implies insurmountable transportation costs and 1 implies zero transportation costs). These equilibrium conditions hold under the condition that the agricultural sector, which pins down the wage, is active in all regions. We assume this to be the case.

3.2 Long-run equilibrium

In the long run, entrepreneurs are mobile between regions and responsive to the incentives provided by the relative real return that can be attained in the two countries. Heterogeneity in labor requirements, a_i, is probabilistically allocated among firms (entrepreneurs). We assume the following cumulative density function of a:

$$G(a) = \frac{a^\rho - \underline{a}^\rho}{a_0^\rho - \underline{a}^\rho}, \quad (9)$$

where ρ is a shape parameter and a_0^ρ is a scaling factor.[3] The distribution is truncated at $\underline{a}$, where $0 < \underline{a} < a_0$, so that the productivity of firms is bounded, and we normalize so that $a_0 = 1$.

Firm location is determined by the regional utility gap for entrepreneurs:

$$V_i - V_i^* = (\pi_i - \mu \ln P) - (\pi_i^* - \mu \ln P^*) = \frac{\mu}{\sigma}(1-\phi)\,(B - B^*)\, a_i^{1-\sigma} - \frac{\mu}{1-\sigma}(\ln \Delta - \ln \Delta^*), \quad (10)$$

where $B \equiv \frac{L+H}{\Delta}$ and $B^* \equiv \frac{L^*+H^*}{\Delta^*}$. The term $\frac{\mu}{1-\sigma}(\ln \Delta - \ln \Delta^*)$, which is called the supply-link in the trade and geography literature, picks up the effect of differences in price indices, whereas the term $\frac{\mu}{\sigma}(1-\phi)\,(B - B^*)\, a_i^{1-\sigma}$ incorporates the competition effect and the demand link (the shift in demand).[4]

The long-run equilibrium is defined as a situation where the difference in indirect utility between locations is less or equal to the cost of migrating χ:

$$|V_i - V_i^*| \leq \chi. \quad (11)$$

In the following, we will analyze location and spatial sorting of entrepreneurs in two settings. First, we analyze a case where transportation costs are high, and we dub this early agglomeration or the development case. Second, we consider the location of entrepreneurs when transportation costs are low. We call this late agglomeration or the modern case.

3.3 Relocation tendencies

We assume that the large region has more labor and human capital and that it is rich in human capital or entrepreneurs in a relative sense. More precisely, we make the following assumption:

Assumption 1 *$H > H^*$, $L > L^*$, and $\frac{H}{L} > \frac{H^*}{L^*}$.*

The difference in real returns between locations substituting (10) is given by

$$V_i - V_i^* = \frac{\mu\,(1-\underline{a}^\rho)}{\sigma\lambda\,(1-\underline{a}^\alpha)}(1-\phi)\left(\frac{L+H}{H+\phi H^*} - \frac{L^*+H^*}{H^*+\phi H}\right) a_i^{1-\sigma} + \frac{\mu}{\sigma-1}\ln\frac{\frac{H}{H^*}+\phi}{\phi\frac{H}{H^*}+1}, \quad (12)$$

where, $\lambda \equiv \frac{\rho}{(1-\sigma+\rho)} > 0$, $\alpha \equiv 1-\sigma+\rho$, and $a_i^{1-\sigma}$ is an index of firm productivity. We start by asking the question of which firm that has the strongest incentive to migrate to the larger core region, if we start out from the initial situation before any migration has occurred. As is clear from (12), the answer depends on the level of transportation costs, ϕ.

[3] This is essentially a Pareto distribution that has been truncated.

[4] See Baldwin et al. (2003) for a description of these effects in a standard trade and geography model.

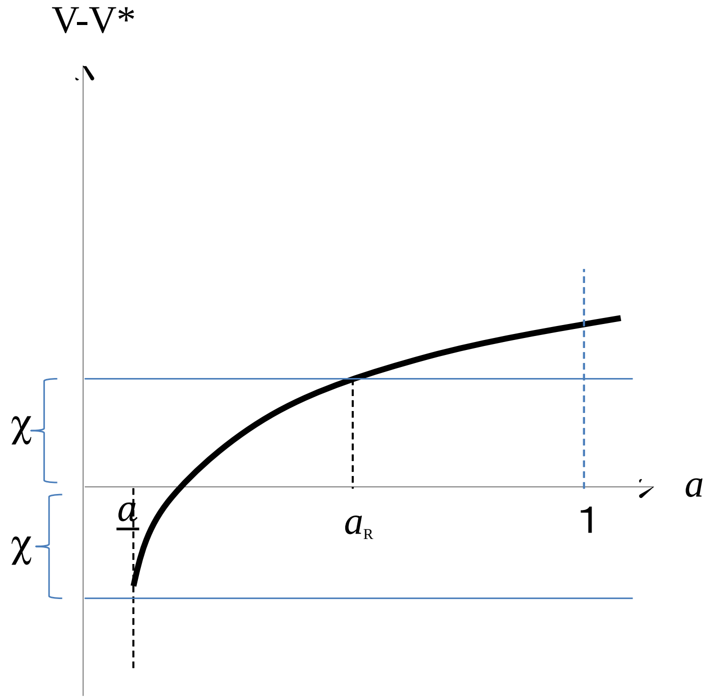

Figure 6: Relocation tendency in the early development case

3.4 Relocation tendencies in the development case (high transportation costs)

We first consider a development case where transportation costs are high. Starting from the case of autarchy where trade costs are prohibitive ($\phi = 0$), we get

$$V - V^* = \frac{\mu\left(1-\underline{a}^{\rho}\right)}{\sigma\lambda\left(1-\underline{a}^{\alpha}\right)}\left(\frac{L+H}{H} - \frac{L^*+H^*}{H^*}\right)a^{1-\sigma} + \frac{\mu}{\sigma-1}\ln\frac{H}{H^*}. \tag{13}$$

Clearly, in this case, the least productive firm will have the strongest incentive to move from the periphery to the larger core region since $\frac{L^*}{H^*} > \frac{L}{H}$. From the point of view of the core, we do instead have $V^* - V = \frac{\mu(1-\underline{a}^{\rho})}{\sigma\lambda(1-\underline{a}^{\alpha})}\left(\frac{L^*}{H^*} - \frac{L}{H}\right)a^{1-\sigma} - \frac{\mu}{\sigma-1}\ln\frac{H}{H^*}$, which implies that the most productive core firms have the strongest incentives to move to the periphery (since $\left(\frac{L^*}{H^*} - \frac{L}{H}\right) > 0$). For simplicity, we will make assumptions that rule out migration from the core to the periphery, so that we only have migration from the periphery to the core (see Appendix 8.1). The situation is described by Figure 6, where firms with marginal costs within the range $[a_R, 1]$ will have incentives to move to the core, whereas the incentives for reverse migration at the other end of the productivity distribution (where $a = \underline{a}$) are not strong enough to overcome the relocation cost χ. The resulting distribution of firms is illustrated in Figure 7.

3.5 Long-run equilibrium in the development case (high transportation costs)

Entrepreneurs are geographically mobile in the long run and migrate in response to differences in the real return. We follow Ottaviano et al. (2002) and assume that entrepreneurs are rational enough to calculate the long-run equilibrium, and that all entrepreneurs that would find it

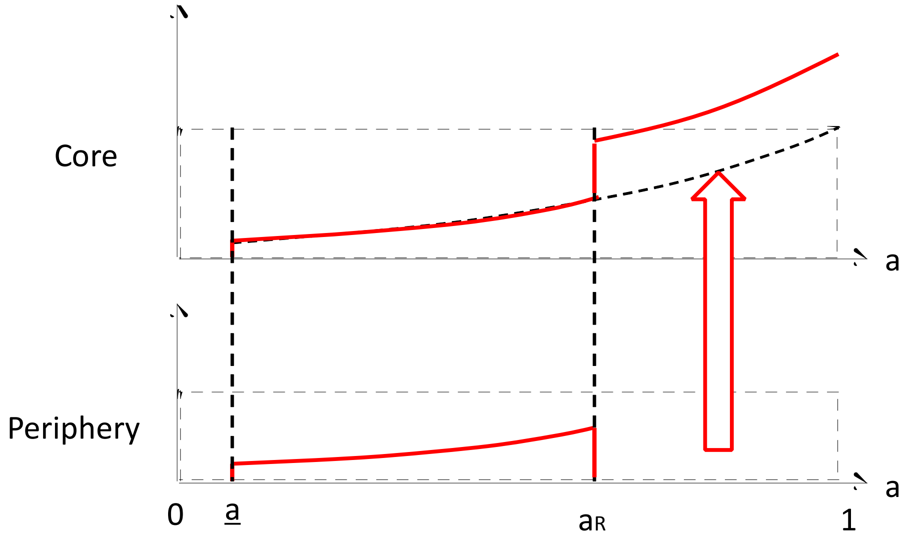

Figure 7: Early development equilibrium

advantageous to migrate in the long run will do so.[5] The long-run equilibrium is characterized by

$$V - V^* = \frac{\mu}{\sigma}(1-\phi)\left(\frac{L+H}{\Delta} - \frac{L^*+H^*}{\Delta^*}\right) a_R^{1-\sigma} - \frac{\mu}{1-\sigma}(\ln\Delta - \ln\Delta^*) = \chi, \tag{14}$$

where

$$\Delta = H\int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + H^*\int_{a_R}^{1} a_k^{1-\sigma} dG(a) + \phi H^*\int_{\underline{a}}^{a_R} a_k^{1-\sigma} dG(a),$$
$$\Delta^* = \phi H\int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + \phi H^*\int_{a_R}^{1} a_k^{1-\sigma} dG(a) + H^*\int_{\underline{a}}^{a_R} a_k^{1-\sigma} dG(a), \tag{15}$$

and where $a_R^{1-\sigma}$ is the productivity of the marginal firm that is just indifferent between migrating and not. Using (14) gives the implicit function that determines a_R:

$$\begin{aligned}&\frac{\mu}{\sigma\lambda}(1-\phi)(1-\underline{a}^{\rho})\left(\begin{array}{c}\frac{(L+H)+(1-a_R^{\rho})H^*}{(1-\underline{a}^{\alpha})(H+\phi H^*)+(1-\phi)(1-a_R^{\alpha})H^*}\\ -\frac{(L^*+H^*)-(1-a_R^{\rho})H^*}{(1-\underline{a}^{\alpha})(\phi H+H^*)-(1-\phi)(1-a_R^{\alpha})H^*}\end{array}\right)a_R^{1-\sigma}\\ &+\frac{\mu}{(\sigma-1)}\ln\frac{(1-\underline{a}^{\alpha})(H+\phi H^*)+(1-\phi)(1-a_R^{\alpha})H^*}{(1-\underline{a}^{\alpha})(\phi H+H^*)-(1-\phi)(1-a_R^{\alpha})H^*}-\chi=0,\end{aligned} \tag{16}$$

[5] Our assumption essentially implies that the relevant segment of entrepreneurs relocates simultaneously. An alternative specification would be to assume that migration costs vary with the migration pressure so that firms migrate in order as in Baldwin and Okubo (2006). However, this assumption would lead to several instances of two-way migration in our setting, which would make the analysis less transparent.

where the term $(1-a_R^\rho)H^*$ represents entrepreneurs that have migrated to the larger region.

Using the implicit function theorem on (16), we can calculate $\frac{da_R}{d\phi}$ and $\frac{da_R}{dH}$. We here show the sign of these derivatives at $\phi = 0$, but we illustrate in our simulations below that the results hold for a range of sufficiently high trade costs.

Proposition 1 $\frac{da_R}{d\phi} < 0$, *lower transportation costs imply that a larger range of firms migrate to the larger region at* $\phi = 0$, *under the sufficient condition that* $\sigma > \frac{1+a_R^\rho}{2a_R^\rho}$.

Proof: See Appendix 8.2.

Proposition 2 $\frac{da_R}{dH} < 0$, *the larger the core region, the larger the range of firms that migrate to the core at* $\phi = 0$.

Proof: See Appendix 8.3.

3.6 Relocation tendencies in a modern economy (low transportation costs)

We will now redo our experiment in the case when transportation costs have fallen below a certain threshold value, $\phi > \widetilde{\phi}$, which is defined below. Thus, we once more consider the same starting point but now assume that we are in a more modern economy where transportation costs are low. The migration tendencies are again given by

$$V_i - V_i^* = \frac{\mu}{\sigma\lambda}(1-\phi)\frac{(1-\underline{a}^\rho)}{(1-\underline{a}^\alpha)}\left(\frac{\frac{L}{H}+1}{1+\phi\frac{H^*}{H}} - \frac{\frac{L^*}{H^*}+1}{1+\phi\frac{H}{H^*}}\right)a_i^{1-\sigma} + \frac{\mu}{\sigma-1}\ln\frac{\frac{H}{H^*}+\phi}{\phi\frac{H}{H^*}+1}. \tag{17}$$

The relationship between firm productivity and migration tendencies is determined by the term $\left(\frac{\frac{L}{H}+1}{1+\phi\frac{H^*}{H}} - \frac{\frac{L^*}{H^*}+1}{1+\phi\frac{H}{H^*}}\right)$. This term switches from negative to positive for

$$\widetilde{\phi} = \frac{HL^* - H^*L}{H^2 + HL - H^{*2} - H^*L^*}. \tag{18}$$

That is, the agglomeration process changes nature at $\phi = \widetilde{\phi}$. We now analyze what we call late agglomeration or the modern economy case where $\phi > \widetilde{\phi}$, which implies that the term $\left(\frac{\frac{L}{H}+1}{1+\phi\frac{H^*}{H}} - \frac{\frac{L^*}{H^*}+1}{1+\phi\frac{H}{H^*}}\right) > 0$. Consequently, it is the most productive firms that have the largest incentives to migrate to the larger core region. The migration pattern in this case, which is similar to the case in Baldwin and Okubo (2006), is shown in Figure 8. Firms with marginal costs within the range $[\underline{a}, a_R]$ will move to the core region, since their gains from doing so are larger than χ. The resulting firm distributions in the two regions are illustrated in Figure 9.

In this case, there is no tendency for two-way migration since with $\phi > \widetilde{\phi}$, we have that $V - V^*|_{a=1} = \frac{\mu}{\sigma\lambda}(1-\phi)\left(\frac{\frac{L}{H}+1}{1+\phi\frac{H^*}{H}} - \frac{\frac{L^*}{H^*}+1}{1+\phi\frac{H}{H^*}}\right)a_i^{1-\sigma} + \frac{\mu}{\sigma-1}\ln\frac{\frac{H}{H^*}+\phi}{\phi\frac{H}{H^*}+1} > 0$ for all permissible $a's$. That is, even the least productive firm (with $a = 1$) will prefer the core absent moving costs.

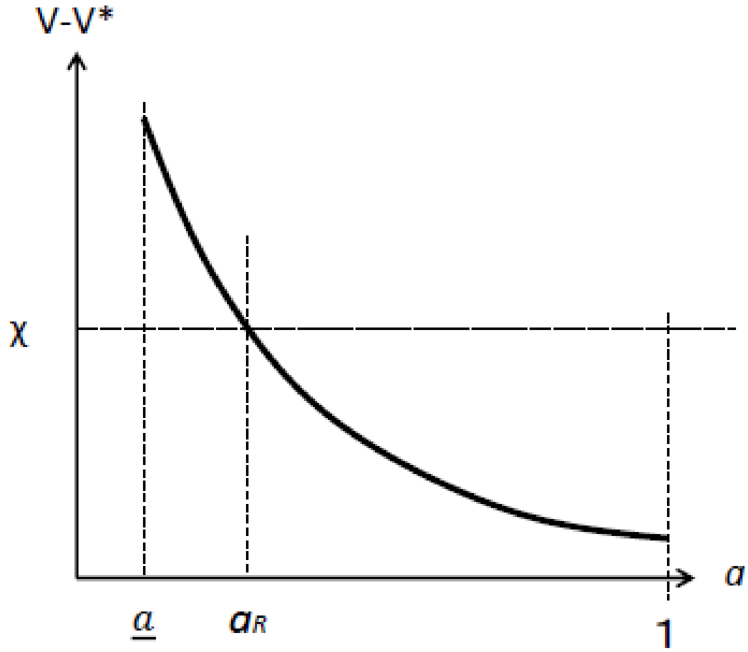

Figure 8: Relocation tendency in the modern economy case.

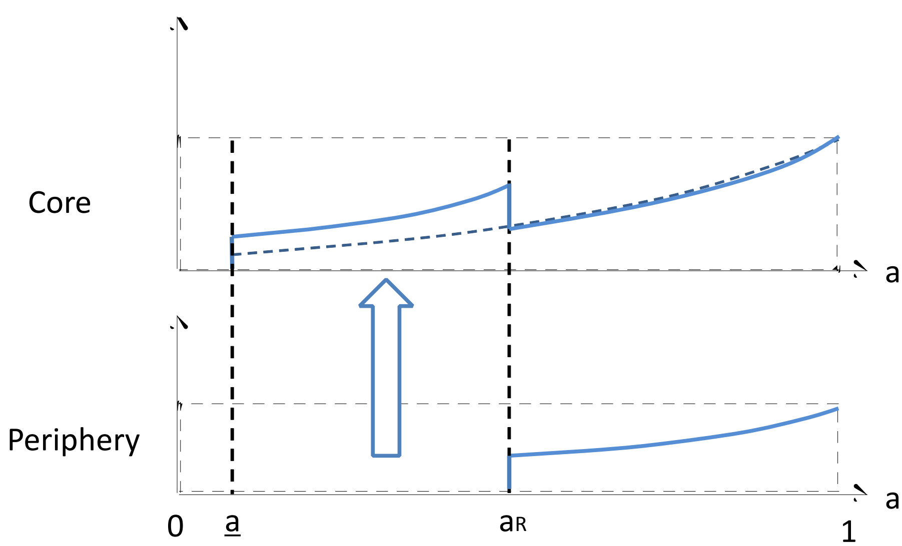

Figure 9: Modern equilibrium

3.7 Long-run equilibrium in the modern case (low transportation costs)

The long-run equilibrium is determined by (14), with

$$\Delta = H \int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + H^* \int_{\underline{a}}^{a_R} a_k^{1-\sigma} dG(a) + \phi H^* \int_{a_R}^{1} a_k^{1-\sigma} dG(a),$$

$$\Delta^* = \phi H \int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + \phi H^* \int_{\underline{a}}^{a_R} a_k^{1-\sigma} dG(a) + H^* \int_{a_R}^{1} a_k^{1-\sigma} dG(a) \tag{19}$$

and the implicit function determining a_R is given by

$$\frac{\mu}{\sigma\lambda}(1-\phi)(1-\underline{a}^{\rho}) \left(\frac{L + H + a_R^{\rho} H^*}{(1-\underline{a}^{\alpha})(H + \phi H^*) + (1-\phi) a_R^{\alpha} H^*} - \frac{L^* + H^* - a_R^{\rho} H^*}{(1-\underline{a}^{\alpha})(\phi H + H^*) - (1-\phi) a_R^{\alpha} H^*} \right) a_R^{1-\sigma} \tag{20}$$

$$+ \frac{\mu}{(\sigma-1)} \ln \frac{H + \phi H^* + (1-\phi) a_R^{\alpha} H^*}{\phi H + H^* - (1-\phi) a_R^{\alpha} H^*} - \chi = 0.$$

Implicit differentiation of (20) leads to the following proposition:

PROPOSITION 3 $\frac{da_R}{dH} > 0$. *The larger the core, the larger the range of firms that migrate to the core.*

Proof: See Appendix 8.4.

A larger size difference will once more lead to more agglomeration to the large region, whereas lower transportation costs have a hump-shaped effect on agglomeration. An important difference between this case and the development case is that while the most productive firms were left in the periphery in the previous development equilibrium, it is now the smallest and least efficient firms that stay in the small region. Moreover, a_R is hump-shaped in the level of transportation costs (ϕ), as illustrated in the simulations below.

So far, we have analyzed two types of agglomeration processes that are separated by the level of transportation costs. A remaining question is what happens when transportation costs gradually fall. Will the early development agglomeration process have reached full agglomeration before $\phi = \widetilde{\phi}$, when the late (modern) agglomeration path becomes feasible and, if so, is it possible that some entrepreneur has an incentive to move back to the periphery? We will start by answering the former question.

4 Sustain point analysis

The sustain point is here defined as the trade cost at which all firms are agglomerated in the larger core economy, that is, the point at which the last firm is indifferent between the two locations. The sustain point will differ for an agglomeration in a development economy and for an agglomeration in a modern economy (with low transportation costs). The last firm to migrate in the development economy case is the most productive one ($a = \underline{a}$). Using $V_i - V_i^* = \chi$ at the sustain point, and noting that we have $\Delta = \lambda \frac{1-\underline{a}^\alpha}{1-\underline{a}^\rho}\left(H + H^*\right), \Delta^* = \phi\lambda\frac{1-\underline{a}^\alpha}{1-\underline{a}^\rho}(H + H^*), B = \frac{1-\underline{a}^\rho}{1-\underline{a}^\alpha}\frac{H+H^*+L}{\lambda(H+H^*)}$, and $B^* = \frac{1-\underline{a}^\rho}{1-\underline{a}^\alpha}\frac{L^*}{\phi\lambda(H+H^*)}$ in this case, gives the following function that implicitly determines the sustain point:[6]

$$\frac{\mu}{\sigma\lambda}\frac{1-\underline{a}^\rho}{1-\underline{a}^\alpha}(1-\phi^{SD})\frac{1}{H+H^*}\left(L + H + H^* - \frac{L^*}{\phi^{SD}}\right)\underline{a}^{1-\sigma} - \frac{\mu}{(\sigma-1)}\ln\phi^{SD} = \chi. \tag{21}$$

Is it then the case that the early agglomeration process (agglomeration in a development economy) has led to full agglomeration when the late agglomeration process becomes feasible? The answer is that it depends:

Proposition 4 *Full agglomeration in the early agglomeration case occurs before the late agglomeration process starts to exist $\phi^{SD} < \widetilde{\phi}$ iff $\frac{HL^*-H^*L}{H^2+HL-H^{*2}-H^*L^*} < e^{\frac{1-\sigma}{\mu}\chi}$.*

Proof: See Appendix 8.5.

5 Equilibrium switches

Once trade freeness is higher than $\widetilde{\phi}$, which separates the early and late agglomeration cases, the model has two equilibria. A question is then if it is possible that the economy switches from one to the other or, to be more precise, if it is possible that the early agglomeration configuration of the industry switches to the late agglomeration (modern) configuration. We discuss here if it is possible that early development full agglomeration, when $\phi > \widetilde{\phi}$, switches to the modern equilibrium path. The deviation incentive for an entrepreneur in the agglomerated core is

$$V^*(a) - V(a) = -\frac{\mu}{\sigma\lambda}\frac{1-\underline{a}^\rho}{1-\underline{a}^\alpha}(1-\phi)\left(\frac{L+H+H^*}{H+H^*} - \frac{L^*}{\phi(H+H^*)}\right)a^{1-\sigma} - \frac{\mu}{(\sigma-1)}\ln\frac{1}{\phi}. \tag{22}$$

Clearly, if the expression in the second parenthesis is positive, $\frac{L+H+H^*}{H+H^*} - \frac{L^*}{\phi(H+H^*)} > 0$, no firm (entrepreneur) would deviate to the periphery. A sufficient condition for stability is therefore that

$$\phi > \frac{L^*}{L+H+H^*}. \tag{23}$$

For instance, if we maintain the assumption that $L = L^*$, we have from (18) that

[6] In a similar way, the last firm to move in the late agglomeration (modern economy) case is the least efficient firm ($a = 1$) and the sustain point is therefore in this case determined by:

$$\frac{\mu}{\sigma\lambda}\frac{1-\underline{a}^\rho}{1-\underline{a}^\alpha}(1-\phi^{SM})\frac{1}{H+H^*}\left(L + H + H^* - \frac{L^*}{\phi^{SM}}\right) - \frac{\mu}{(\sigma-1)}\ln\phi^{SM} = \chi.$$

$$\phi > \frac{L}{L+H+H^*} = \widetilde{\phi}. \tag{24}$$

Thus, there is a large range of parameter values for which switches do not occur. This is an example of hysteresis or of path dependence where history matters. The early agglomeration will remain as the economy turns to a modern phase, and this economy will have a higher degree of agglomeration **as** compared to an economy that start agglomerating in the modern phase all the way until free trade.

6 A relocation policy experiment

Several developing countries do, at times, have limited internal migration, which will tend to postpone agglomeration.[7] We will now use the model to analyze the effects of this policy. That is, we will compare the effects of early and late agglomeration. The two cases occur for different levels of transportation costs. Countries that postpone agglomeration will have a different agglomeration process because transportation costs are lower at a later stage of development. Our two policies are: i) early migration (agglomeration in a development economy) and then development (lower trade costs), ii) development (lower transportation costs) and then migration and agglomeration (the modern economy case). The model is solved by numerical simulation where we use $L = 3, L^* = 2, H = 6, H^* = 2.5, \sigma = 2, \rho = 4, \underline{a} = 0.1, \mu = 0.8$ and $\chi = 0.6$. Note that these parameter values respect the assumptions made in the theory section that $L > L^*, H > H^*$, and $\frac{H}{L} > \frac{H^*}{L^*}$. The simulations show how the two policies may imply very different development paths. The agglomeration in a development case and the agglomeration in the modern agglomeration case are here separated by $\widetilde{\phi} = 0.174$. Figure 6 shows the relocation path (the path of a_R) in the two cases, where the development case is depicted by the dashed line. A range of the least productive firms (firms with a in the range $[0.59, 1]$) agglomerate to the larger region already at autarky. Lower trade costs lead to further agglomeration until all human capital has agglomerated to the larger core region at $\phi = \widetilde{\phi}$. The industry remains fully agglomerated to the core as ϕ increases towards one. On the other hand, if human capital (firms) is allowed to migrate only at a later stage (when $\phi > \widetilde{\phi}$), the economy instead follows the path of the modern economy depicted by a solid line in Figure 6. Agglomeration is here slower, and it is the most productive firms that are the first to migrate to the core. Full agglomeration may never materialize in this case, as in Figure 6. Agglomeration rents are hump-shaped as usual, and the point where agglomeration rents are maximal is also the point where the maximal range of firms relocate to the core.[8]

The two development paths have different implications for the agglomeration premium at

[7] One important example is China. See e.g. Bosker et al. (2012) and Walley and Zang (2007).

[8] Some firms may want to relocate back from the core, when we have passed the hump and trade freeness is sufficiently higher than $\widetilde{\phi}$. Assuming that human capital (firms) dies after a couple of time periods eliminates this possibility.

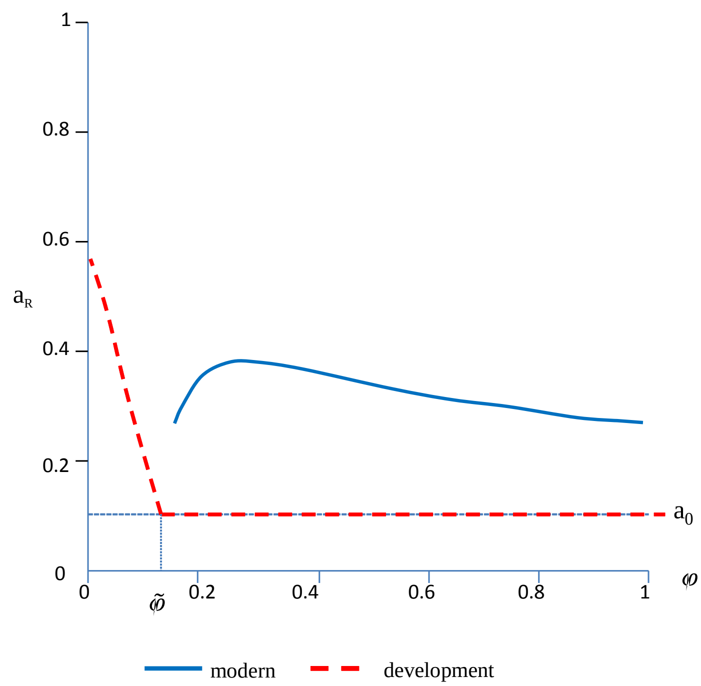

Figure 10: Cut-offs

the core since in the development case, it is the least productive firms that move to the core whereas in the modern case, it is the most productive firms that move to the core. We use a weighted productivity index as a measure of average productivity (a harmonic mean weighted by output shares) as in Melitz (2003), which absent migration is given by $\Phi = \left(\int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) \right)^{\frac{1}{\sigma-1}}$. With migration, we have the following indices for the core

$$\Phi_{development}^{core} = \left(\frac{H \int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + H^* \int_{a_R}^{1} a_k^{1-\sigma} dG(a)}{H + H^* \int_{a_R}^{1} dG(a)} \right)^{\frac{1}{\sigma-1}}, \quad \Phi_{Modern}^{core} = \left(\frac{H \int_{\underline{a}}^{1} a_k^{1-\sigma} dG(a) + H^* \int_{\underline{a}}^{a_R} a_k^{1-\sigma} dG(a)}{H + H^* \int_{\underline{a}}^{a_R} dG(a)} \right)^{\frac{1}{\sigma-1}} \tag{25}$$

and for the periphery

$$\Phi_{development}^{periphery} = \frac{H^* \left(\int_{\underline{a}}^{a_R} a_k^{1-\sigma} dG(a) \right)^{\frac{1}{\sigma-1}}}{H^* \int_{\underline{a}}^{a_R} dG(a)}, \quad \Phi_{Modern}^{periphery} = \frac{H^* \left(\int_{a_R}^{1} a_k^{1-\sigma} dG(a) \right)^{\frac{1}{\sigma-1}}}{H^* \int_{a_R}^{1} dG(a)}. \tag{26}$$

Figures 7 and 8 plot the productivity premia in the core in the development case and the modern case, i.e. $\frac{\Phi_{development}^{core}}{\Phi_{development}^{periphery}}$ and $\frac{\Phi_{Modern}^{core}}{\Phi_{Modern}^{periphery}}$. The agglomeration premium is actually negative and rapidly falling in the developing case. This is due to the fact that it is the least productive firms that agglomerate in the core, and close to $\widetilde{\phi}$ only very productive firms remain in the periphery. In the modern case, we do instead have a positive and less variable productivity premium. Both these properties are consistent with stylized evidence.[9]

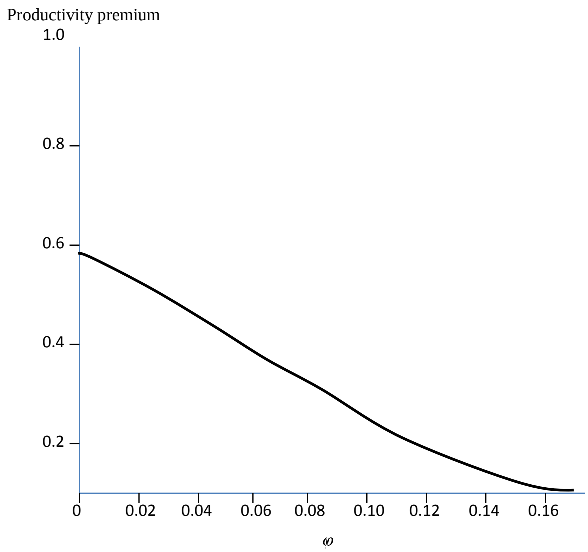

Figure 11a: Productivity premium in the development case

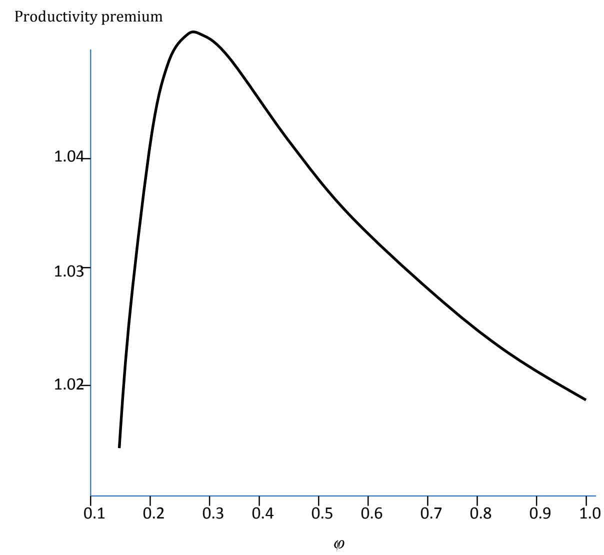

Figure 11b: Productivity premium in the modern case

Next, we compare the two development paths in terms of social welfare. Indirect utility is given by $V = w - \mu \ln P$, and we use the sum of the indirect utility for all individuals as the measure of social welfare. Thus, we must take the location of the entrepreneurs into account. Social welfare in the modern agglomeration case is given by

$$H\left(\int_{\underline{a}}^{1} \pi dG(a) - \mu \ln P\right) + H^* \int_{\underline{a}}^{a_R} \pi dG(a) - \frac{a_R^\rho - \underline{a}^\rho}{1 - \underline{a}^\rho} H^* \mu \ln P + H^* \int_{a_R}^{1} \pi^* dG(a) - \frac{1 - a_R^\rho}{1 - \underline{a}^\rho} H^* \mu \ln P^*$$
$$+ L(w - \mu \ln P) + L^*(w - \mu \ln P^*), \tag{27}$$

where $\frac{a_R^\rho - \underline{a}^\rho}{1 - \underline{a}^\rho}$ is the share of the peripheral entrepreneurs that have migrated to the core and $\frac{1 - a_R^\rho}{1 - \underline{a}^\rho}$ the share of these entrepreneurs still living in the periphery. Social welfare in the development (early agglomeration) case is in the same way given by

$$H\left(\int_{\underline{a}}^{1} \pi dG(a) - \mu \ln P\right) + H^* \int_{a_R}^{1} \pi dG(a) - \frac{1 - a_R^\rho}{1 - \underline{a}^\rho} H^* \mu \ln P + H^* \int_{\underline{a}}^{a_R} \pi^* dG(a) - \frac{a_R^\rho - \underline{a}^\rho}{1 - \underline{a}^\rho} H^* \mu \ln P^*$$
$$+ L(w - \mu \ln P) + L^*(w - \mu \ln P^*). \tag{28}$$

[9] The very bad performance of the core in the simulations of the development case **does,** of course, **depend** on the fact that there are **no** other externalities between firms such as e.g. local productivity spillovers.

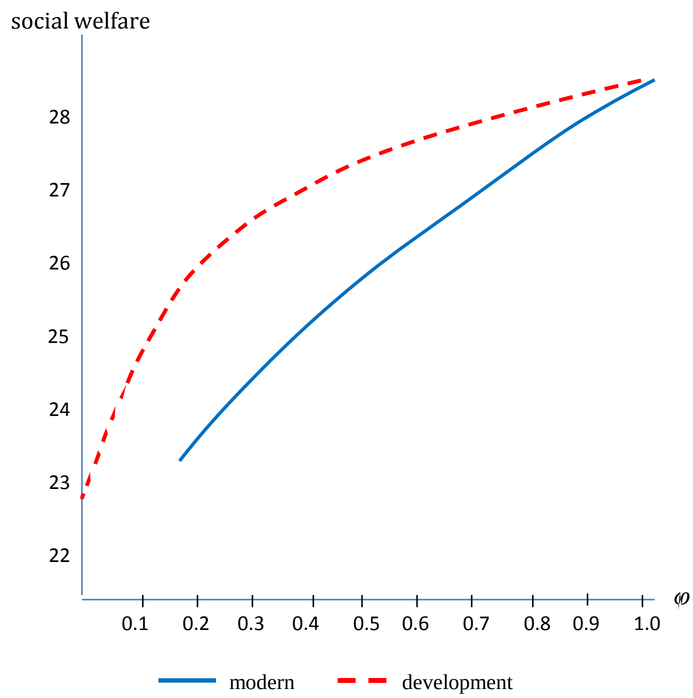

Figure 12: Social welfare

Figure 12 shows the development of social welfare in the two cases. The early agglomeration path is clearly superior in terms of total social welfare and the two curves only converge at zero transportation costs ($\phi = 1$) where location ceases to have any importance.

However, even if total social welfare is always higher here in the early agglomeration case, it does create considerably higher tensions between workers in the two regions. The real return to entrepreneurs in the two regions cannot differ by more than the moving cost (χ) in equilibrium, but the real return to unskilled labor can differ sharply between regions. This difference depends solely on the price index, since nominal wages are equal. Figure 13 shows a measure of the inverse of the price index (Δ) in the two regions in the two agglomeration cases. It is clear that unskilled workers in the core always have a higher real wage due to the lower price index in the core. However, the difference in real returns for unskilled labor between regions is much larger in the early agglomeration (development) case, as seen by comparing the distance between the two dashed curves and the two solid ones. It may seem surprising that welfare is always higher for unskilled labor in the core under the early agglomeration regime, since low-productivity firms are the first to sort to the core in this case. However, full agglomeration in the core occurs almost at the same time as the modern agglomeration process becomes possible. So the core in the development case contains all firms as compared to the not fully agglomerated core in the modern case. Thus, in sum, the development path allowing for early agglomeration is superior in terms of total welfare, but may lead to higher social tensions due to wage differences between the core and the periphery. It is also noteworthy that whereas the early agglomerator ends up in full agglomeration relatively early on, this may never materialize for the later agglomerator

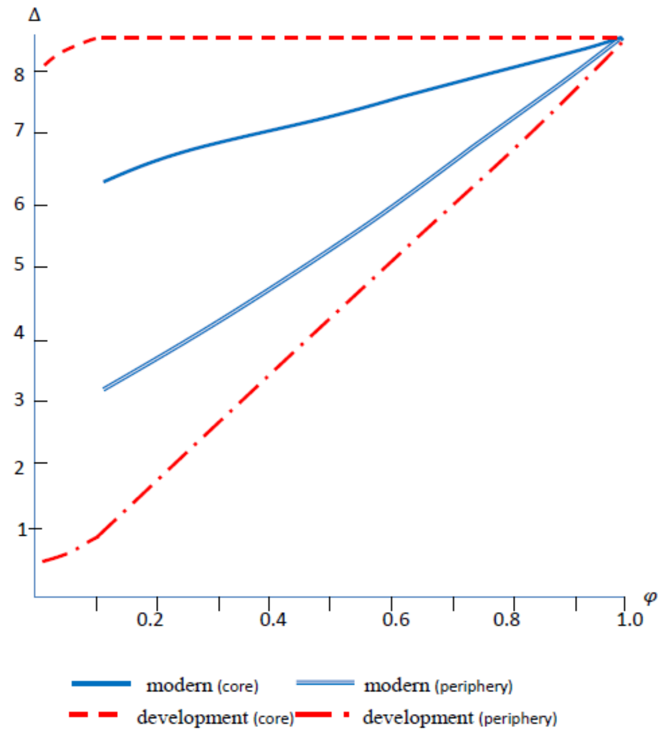

Figure 13: Welfare measure for low-skilled workers.

(modern case) unless the transportation costs become very low.

7 Conclusions

This paper starts out from the observation that agglomeration may have very different implications in developing countries and advanced economies, respectively. We develop a model that generates different agglomeration paths in advanced economies with low transportation costs, as compared to developing economies with high transportation costs. We show that countries that limit the geographical movement of human capital (and firms) at early stages of development may end up in a different equilibrium than countries that allow for geographical mobility early on. At early stages of development (when transportation costs are high), the model has an equilibrium where low-productivity firms concentrate in the large and human capital rich region, whereas the most productive firms agglomerate to the smaller region with a relatively high endowment of labor. We relate this type of equilibrium to countries at an early stage of development, where a separation of the administrative center (with high-educated officials) and the industrial center is often observed, and where industrial productivity is far higher in the industrial center. As transportation costs fall, the model switches to an equilibrium where the most productive firms concentrate to the larger and human capital rich region. This corresponds to a modern equilibrium where highly productive firms concentrate in the largest and most human capital rich regions. Whereas our numerical simulations illustrate that free mobility of human capital from the outset is superior from the perspective of total social welfare, it does come at the cost of considerably higher differences in wages between workers in the core and in the periphery.

References

[1] Baldwin R., R. Forslid, P. Martin, G.I.P. Ottaviano and F. Robert-Nicoud, (2003). *Economic Geography and Public Policy*, Princeton University Press.

[2] Baldwin R. and T. Okubo, (2006). "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting", *Journal of Econiomic Geography* 6(3), 323-346.

[3] Baldwin, R., and T. Okubo (2014). "International trade, offshoring and heterogeneous firms." *Review of International Economics* 22.1 : 59-72.

[4] Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3), 507-553.

[5] Borck, R., and Pfluger, M. (2006). Agglomeration and tax competition. *European Economic Review*, 50(3), 647-668.

[6] Bosker M., Brakman S., Garretsen H. and M.Schramm (2012). "Relaxing Hukou: Increased labor mobility and China's economic geography", *Journal of Urban Economics* 72, 252–266.

[7] Desmet, K, E. Ghani, S. O'Connell, and E.Rossi-Hansberg. 2012. "The Spatial Development of India." *Policy Research Working Paper* 6060. World Bank, Washington, DC.

[8] Eeckhout, J., R. Pinheiro and K. Schmidheiny, (2013). "Spatial Sorting.", *Journal of Political Economy*, forthcoming.

[9] Forslid R. (1999). "Agglomeration with Human and Physical Capital: an Analytically Solvable Case," CEPR Discussion Papers, no. 2102.

[10] Forslid R. and G.Ottaviano (2003)."An Analytically Solvable Core-Periphery Model", *Journal of Economic Geography* 3, pp.229-240.

[11] Fujita, M., P. Krugman and A.J.Venables, (1999). The Spatial Economy: Cities, Regions and International Trade. MIT Press, Cambridge, Massachusetts.

[12] Fujita, M., Thisse, J.-F. (2002) *Economics of Agglomeration*. Cambridge University Press.

[13] Forslid, R., & Okubo, T. (2012). On the development strategy of countries of intermediate size—An analysis of heterogeneous firms in a multi-region framework. *European Economic Review*, 56(4), 747-756.

[14] Ghani, E, A.G. Goswami, and W.Kerr. (2012). "Is India's Manufacturing Moving Away From Cities? " Working Paper 12-090. Harvard Business School, Cambridge, MA

[15] Gollin D., Jedwab R. and D.Vollrath (2015)."Urbanization with and without Industrialization", *Journal of Economic Growth*, forthcoming.

[16] Mendels, F. F. (1972). Proto-industrialization: the first phase of the industrialization process. *The journal of economic history*, 32(01), 241-261.

[17] Melitz, M. J. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, 71(6), 1695–1725.

[18] Melo, P., D. Graham and R. Noland, (2009). "A Meta-analysis of Estimates of Urban Agglomeration Economies," *Regional Science and Urban Economics* 39, 332-342.

[19] Nocke, V. (2006). A gap for me: Entrepreneurs and entry. Journal of the European Economic Association, 4(5), 929-956.

[20] Ogilvie, S., and Cerman, M. (1996). *European proto-industrialization: an introductory handbook*. Cambridge University Press.

[21] Pfluger, M. (2004). A simple, analytically solvable, Chamberlinian agglomeration model. *Regional Science and Urban Economics*, 34(5), 565-573.

[22] Ottaviano G.I.P., T. Tabuchi and J-F. Thisse, (2002). "Agglomeration and Trade Revisited", *International Economic Review* 43, 409-436.

[23] Rosenthal, S. and W. Strange, (2004). "Evidence on the Nature and Sources of Agglomeration Economies", Chapter 49, 2119-2171, in J. Henderson and J. Thisse (Eds) *Handbook of Urban and Regional Economics vol.4*, Elsevier, Amsterdam.

[24] Saito, O (1985) Puroto kougyou-ka no jidai-seiou to nihon no hikaku-shi, (The age of proto-industrialization: Western Europe and Japan in historical and comparative perspective), written in Japanese, *Nihon Hyoron sha.*

[25] Settsu T. (2015). "Industrial Structure, Prefectural Inequality, and Convergence in Pre-war Japan (1874-1940)", *RCESR Discussion Paper Series*, No. DP15-1.

[26] Whalley J. and S. Zhang (2007). "A numerical simulation analysis of (Hukou) labor mobility restrictions in China", *Journal of Development Economics* 83, 392–410.

8 Appendix

8.1 Conditions ruling out two-way migration

In order to rule out two-way migration, we need two conditions:

The first is that

$$V - V^*|_{a=1} > -(V - V^*)|_{a=\underline{a}}\,. \tag{29}$$

This condition implies that the incentive for the least productive firm to move from the periphery to the core is stronger that the incentive for the most productive firm to move from the core to the periphery. Using (13), this condition becomes:

$$\frac{\mu\left(1-\underline{a}^{\rho}\right)}{\sigma\lambda\left(1-\underline{a}^{\alpha}\right)}\left(\frac{L}{H}-\frac{L^{*}}{H^{*}}\right)\left(1+a^{1-\sigma}\right)+\frac{2\mu}{\sigma-1}\ln\frac{H}{H^{*}}>0. \tag{30}$$

By inspection, the condition will always hold for H sufficiently large.

Second, we need to assume that χ is sufficiently large to preclude migration from the core to the periphery but not larger than to allow migration from the periphery to the core. Thus,

$$\left.(V-V^{*})\right|_{a=1}>\chi>-\left.(V-V^{*})\right|_{a=\underline{a}}, \tag{31}$$

which using (13) becomes

$$\frac{\mu\left(1-\underline{a}^{\rho}\right)}{\sigma\lambda\left(1-\underline{a}^{\alpha}\right)}\left(\frac{L}{H}-\frac{L^{*}}{H^{*}}\right)+\frac{\mu}{\sigma-1}\ln\frac{H}{H^{*}}>\chi>\frac{\mu\left(1-\underline{a}^{\rho}\right)}{\sigma\lambda\left(1-\underline{a}^{\alpha}\right)}\left(\frac{L^{*}}{H^{*}}-\frac{L}{H}\right)\underline{a}^{1-\sigma}-\frac{\mu}{\sigma-1}\ln\frac{H}{H^{*}}. \tag{32}$$

In sum, for high trade costs, two-way migration is ruled out by having a sufficiently large H and a χ of an intermediate level as defined by (32).

8.2 Proposition 1: Conditions for $\frac{da_R}{d\phi}<0$

We want to determine the sign of $\frac{da_R}{d\phi}$, and we start by establishing the conditions for $B<B^*$ as a corollary.

8.2.1 Conditions for $B<B^*$ in the development case

$$B-B^{*}=\frac{L+H+(1-a_R^{\rho})H^{*}}{(1-a^{\alpha})(H+\phi H^{*})+(1-\phi)(1-a^{\alpha})H^{*}}-\frac{L^{*}+H^{*}-(1-a_R^{\rho})H^{*}}{(1-a^{\alpha})(\phi H+H^{*})-(1-\phi)(1-a^{\alpha})H^{*}} \tag{33}$$

$$=\left[\begin{array}{c}\frac{\left(L+H+(1-a_R^{\rho})H^{*}\right)((1-a^{\alpha})(\phi H+H^{*})-(1-\phi)(1-a^{\alpha})H^{*})}{\{(1-a^{\alpha})(H+\phi H^{*})+(1-\phi)(1-a^{\alpha})H^{*}\}\{(1-a^{\alpha})(\phi H+H^{*})-(1-\phi)(1-a^{\alpha})H^{*}\}}\\ -\frac{\left(L^{*}+H^{*}-(1-a_R^{\rho})H^{*}\right)((1-a^{\alpha})(H+\phi H^{*})+(1-\phi)(1-a^{\alpha})H^{*})}{\{(1-a^{\alpha})(H+\phi H^{*})+(1-\phi)(1-a^{\alpha})H^{*}\}\{(1-a^{\alpha})(\phi H+H^{*})-(1-\phi)(1-a^{\alpha})H^{*}\}}\end{array}\right]$$

The sign of this expression depends on the numerators. The numerator of the first term can be simplified as:

$$\begin{aligned}&(L+H+(1-a_R^{\rho})H^{*})(1-a^{\alpha})(\phi H+H^{*})-(L+H+(1-a_R^{\rho})H^{*})(1-\phi)(1-a^{\alpha})H^{*}\\&=(L+H)(1-a^{\alpha})(\phi H+H^{*})+(1-a_R^{\rho})H^{*}(1-a^{\alpha})(\phi H+H^{*})-(L+H+(1-a_R^{\rho})H^{*})(1-\phi)(1-a^{\alpha})H^{*}\\&=((1-a^{\alpha})(H+H^{*}))\phi((L+H)+(1-a_R^{\rho})H^{*}).\end{aligned}$$

The numerator in the second term can be simplified in a similar way:

$$(L^*+H^*-(1-a_R^\rho)H^*)(1-a^\alpha)(H+\phi H^*)+(L^*+H^*-(1-a_R^\rho)H^*)(1-\phi)(1-a^\alpha)H^*$$
$$=(L^*+H^*)(1-a^\alpha)(H+\phi H^*)-(1-a_R^\rho)H^*(1-a)(H+\phi H^*)+(L^*+H^*)(1-\phi)(1-a^\alpha)H^*$$
$$-(1-aR)H^*(1-\phi)(1-a^\alpha)H^*$$
$$=(1-a^\alpha)(H+H^*)((L^*+H^*)-(1-a_R^\rho)H^*)$$

The difference of the numerators is now given by

$$(1-a^\alpha)(H+H^*)\{\phi((L+H)+(1-a_R^\rho)H^*)-((L^*+H^*)-(1-a_R^\rho)H^*)\}.$$

Thus,

$$B<B^* \qquad if \qquad \phi<\frac{((L^*+H^*)-(1-a_R^\rho)H^*)}{((L+H)+(1-a_R^\rho)H^*)}\leq\frac{(L^*+H^*)}{(L+H)}. \tag{34}$$

8.2.2 $\frac{da_R}{d\phi}$

We now turn to the sign of $\frac{da_R}{d\phi}$. a_R is from (16) defined by the implicit function $F=0$, where

$$F\equiv(1-\phi)\frac{(1-\underline{a}^\rho)}{\sigma\lambda}\begin{pmatrix}\frac{L+H+(1-a_R^\rho)H^*}{(1-\underline{a}^\alpha)(H+\phi H^*)+(1-\phi)(1-a_R^\alpha)H^*}\\-\frac{L^*+H^*-(1-a_R^\rho)H^*}{(1-\underline{a}^\alpha)(\phi H+H^*)-(1-\phi)(1-a_R^\alpha)H^*}\end{pmatrix}a_R^{1-\sigma} \tag{35}$$
$$+\ln\frac{(1-\underline{a}^\alpha)\,(H+\phi H^*)+(1-\phi)(1-a_R^\alpha)H^*}{(1-\underline{a}^\alpha)\,(\phi H+H^*)-(1-\phi)(1-a_R^\alpha)H^*}-\frac{\chi}{\mu},$$

and where $\alpha\equiv 1-\sigma+\rho$. Implicit differentiation gives $\frac{da_R}{d\phi}=-\frac{\frac{dF}{d\phi}}{\frac{dF}{da_R}}$, and we start by calculating $\frac{dF}{da_R}$:

$$\frac{dF}{da_R}=(1-\phi)\frac{((1-\sigma)(L+H)a_R^{-\sigma}-\alpha a_R^{\alpha-1}H^*)\Delta+\lambda(L+H+(1-a_R^\rho)H^*)(\alpha(1-\phi)a_R^{\alpha-1}H^*)}{\Delta^2}$$
$$-(1-\phi)\frac{((1-\sigma)(L^*+H^*)a_R^{-\sigma}+\alpha a_R^{\alpha-1}H^*)\Delta-\lambda(L^*+H^*-(1-a_R^\rho)H^*)(\alpha(1-\phi)a_R^{\alpha-1}H^*)}{\Delta^{*2}}$$
$$-\frac{\sigma}{(\sigma-1)}(1-\phi)\alpha a_R^{\alpha-1}H^*\left(\frac{1}{\Delta}+\frac{1}{\Delta^*}\right)$$
$$=(1-\phi)\begin{pmatrix}\frac{((1-\sigma)(L+H)a_R^{-\sigma}-\alpha a_R^{\alpha-1}H^*)\Delta+\lambda(L+H+(1-a_R^\rho)H^*)(\alpha(1-\phi)a_R^{\alpha-1}H^*)}{\Delta^2}\\-\frac{((1-\sigma)(L^*+H^*)a_R^{-\sigma}+\alpha a_R^{\alpha-1}H^*)\Delta-\lambda(L^*+H^*-(1-a_R^\rho)H^*)(\alpha(1-\phi)a_R^{\alpha-1}H^*)}{\Delta^{*2}}-\frac{\sigma}{(\sigma-1)}\alpha a_R^{\alpha-1}H^*\left(\frac{1}{\Delta}+\frac{1}{\Delta^*}\right)\end{pmatrix}$$
$$=(1-\phi)\begin{pmatrix}\frac{((1-\sigma)(L+H)a_R^{-\sigma})}{\Delta}-\frac{((1-\sigma)(L^*+H^*)a_R^{-\sigma})}{\Delta^*}\\+\left(B\frac{(\alpha(1-\phi)a_R^{\alpha-1}H^*)}{\Delta}+B^*\frac{(\alpha(1-\phi)a_R^{\alpha-1}H^*)}{\Delta^*}\right)-\frac{2\sigma-1}{(\sigma-1)}\alpha a_R^{\alpha-1}H^*\left(\frac{1}{\Delta}+\frac{1}{\Delta^*}\right)\end{pmatrix}$$
$$=(1-\phi)\left((\sigma-1)a_R^{-\sigma}\left(\frac{L^*+H^*}{\Delta^*}-\frac{L+H}{\Delta}\right)+\left((1-\phi)(B+B^*)-\frac{2\sigma-1}{(\sigma-1)}\right)\alpha a_R^{\alpha-1}H^*\left(\frac{1}{\Delta}+\frac{1}{\Delta^*}\right)\right)$$

where $B \equiv \frac{L+H+(1-a_R^{\rho})H^*}{\Delta}$ and $B^* \equiv \frac{L^*+H^*-(1-a_R^{\rho})H^*}{\Delta^*}$. We have from (34) that $B < B^*$ *if* $\phi < \frac{((L^*+H^*)-(1-a_R^{\rho})H^*)}{((L+H)+(1-a_R^{\rho})H^*)}$ and we assume this condition to hold. Then $\frac{L+H}{\Delta} < \frac{L^*+H^*}{\Delta^*}$. Finally, consider the term $(1-\phi)(B+B^*) - \frac{2\sigma-1}{(\sigma-1)}$. When $\phi = 0$ there is no relocation, which means that $a_R = 1$.

To proceed we make the following sufficient assumption on σ :

Assumption 2 (sigma ass) $\sigma > \frac{1+a_R^{\rho}}{2a_R^{\rho}}$

Given this assumption, we have that $(1-\phi)(B+B^*) - \frac{2\sigma-1}{(\sigma-1)} = \frac{1}{(1-\underline{a}^{\rho})}(\frac{L+H}{H} + \frac{L^*+H^*}{H^*}) - \frac{2\sigma-1}{(\sigma-1)} > 0$. We therefore have $\frac{dF}{da_R} > 0$, at $\phi = 0$.

Next we prove that $\frac{dF}{d\phi} > 0$

$$\begin{aligned}
\frac{dF}{d\phi} &= \left[\begin{array}{c} -\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{L+H+(1-a_R^{\rho})H^*}{\Delta} - \frac{L^*+H^*-(1-a_R^{\rho})H^*}{\Delta^*}\right)a_R^{1-\sigma} \\ +(1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\left(-\frac{(L+H+(1-a_R^{\rho})H^*)a^{1-\sigma}}{\Delta^2}\frac{d\Delta}{d\phi} + \frac{(L^*+H^*-(1-a_R^{\rho})H^*)a^{1-\sigma}}{\Delta^{*2}}\frac{d\Delta^*}{d\phi}\right)\right) \\ +\frac{1}{\sigma-1}(\frac{1}{\Delta}\frac{d\Delta}{d\phi} - \frac{1}{\Delta^*}\frac{d\Delta^*}{d\phi}) \end{array} \right] \\
&= -\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(B - B^*\right)a_R^{1-\sigma} + \frac{(1-\underline{a}^{\rho})}{\sigma\lambda}(1-\phi)\left(\left(-B\frac{a_R^{1-\sigma}}{\Delta}\frac{d\Delta}{d\phi} + B^*\frac{a_R^{1-\sigma}}{\Delta^*}\frac{d\Delta^*}{d\phi}\right)\right) \\
&+ \frac{1}{\sigma-1}(\frac{1}{\Delta}\frac{d\Delta}{d\phi} - \frac{1}{\Delta^*}\frac{d\Delta^*}{d\phi}) \\
&= \frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(B^* - B\right)a_R^{1-\sigma} - \left(\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}(1-\phi)Ba_R^{1-\sigma} - \frac{1}{\sigma-1}\right)\frac{1}{\Delta}\frac{d\Delta}{d\phi} \\
&+ \left(\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}(1-\phi)B^*a_R^{1-\sigma} - \frac{1}{\sigma-1}\right)\frac{1}{\Delta^*}\frac{d\Delta^*}{d\phi})
\end{aligned}$$

The sign of $\frac{dF}{d\phi}$ therefore depends on the sign of

$$\begin{aligned}
&\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(B^* - B\right)a_R^{1-\sigma} - \left(\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}(1-\phi)Ba_R^{1-\sigma} - \frac{1}{\sigma-1}\right)\frac{1}{\Delta}\frac{d\Delta}{d\phi} \\
&+ \left(\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}(1-\phi)B^*a_R^{1-\sigma} - \frac{1}{\sigma-1}\right)\frac{1}{\Delta^*}\frac{d\Delta^*}{d\phi}).
\end{aligned} \tag{36}$$

Now

$$\begin{aligned}
\frac{d\Delta}{d\phi} &= \left(1-\underline{a}^{\alpha}\right)H^* - (1-a_R^{\alpha})H^* > 0, \\
\frac{d\Delta^*}{d\phi} &= \left(1-\underline{a}^{\alpha}\right)H + (1-a_R^{\alpha})H^* > 0,
\end{aligned}$$

and since $\Delta > \Delta^*$ and $H > H^*$, we have that

$$\frac{1}{\Delta}\frac{d\Delta}{d\phi} < \frac{1}{\Delta^*}\frac{d\Delta^*}{d\phi}.$$

Moreover from (34) $B^* > B$, which implies that

$$\frac{dF}{d\phi} > 0.$$

Using $\frac{dF}{da_R} > 0$ and $\frac{dF}{d\phi} > 0$, we get

$$\frac{da_R}{d\phi} = -\frac{\frac{dF}{d\phi}}{\frac{dF}{da_R}} < 0$$

for $\phi = 0$.

8.3 Proposition2: Proof $\frac{da_R}{dH} < 0$ in development case

$$F \equiv$$

$$\frac{(1-\phi)\left(1-\underline{a}^{\rho}\right)}{\sigma\lambda}\left(\frac{L+H+(1-a_R^{\rho})H^*}{\left(1-\underline{a}^{\alpha}\right)\left(H+\phi H^*\right)+(1-\phi)(1-a_R^{\alpha})H^*}-\frac{L^*+H^*-(1-a_R^{\rho})H^*}{\left(1-\underline{a}^{\alpha}\right)\left(\phi H+H^*\right)-(1-\phi)(1-a_R^{\alpha})H^*}\right)a_R^{1-\sigma} \tag{37}$$

$$+\frac{1}{(\sigma-1)}\ln\frac{\left(H+\phi H^*\right)\left(1-\underline{a}^{\alpha}\right)+(1-\phi)(1-a_R^{\alpha})H^*}{\left(\phi H+H^*\right)\left(1-\underline{a}^{\alpha}\right)-(1-\phi)(1-a_R^{\alpha})H^*}-\frac{\chi}{\mu}.$$

We here investigate the sign of $\frac{da_R}{dH}$, and

$$\frac{da_R}{dH} = -\frac{\frac{dF}{dH}}{\frac{dF}{da_R}}. \tag{38}$$

From the previous proof we know that $\frac{dF}{da_R} > 0$ if $\phi = 0$.

Next we turn to the sign of $\frac{dF}{dH}$:

$$\frac{dF}{dH} = (1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{a_R^{1-\sigma}\Delta-(L+H+(1-a_R^{\rho})H^*)}{\Delta^2}+\frac{(L^*+H^*-(1-a_R^{\rho})H^*)\phi}{\Delta^{*2}}\right)+\frac{1}{(\sigma-1)}\left(\frac{1}{\Delta}-\frac{\phi}{\Delta^*}\right)$$

$$= (1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{1}{\Delta}-\frac{B}{\Delta}+\frac{B^*}{\Delta^*}\phi\right)a_R^{1-\sigma}+\frac{1}{(\sigma-1)}\left(\frac{1}{\Delta}-\frac{\phi}{\Delta^*}\right)$$

$$> (1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{\Delta-(L+H+(1-a_R^{\rho})H^*)(1-\phi)}{\Delta^2}\right)+\frac{1}{(\sigma-1)}\left(\frac{1}{\Delta}-\frac{\phi}{\Delta^*}\right)$$

Now

$$\Delta-(L+H+(1-a_R^{\rho})H^*)(1-\phi) = \left(\left(1-\underline{a}^{\alpha}\right)\left(H+\phi H^*\right)+(1-\phi)(1-a_R^{\alpha})H^*\right)-(1-\phi)((L+H)+(1-a_R^{\rho})H^*) > 0.$$

Moreover $\frac{1}{\Delta}-\frac{\phi}{\Delta^*} > 0$ since

$$\phi\Delta-\Delta^* = \phi\left(\left(1-\underline{a}^{\alpha}\right)\left(H+\phi H^*\right)+(1-\phi)(1-a_R^{\alpha})H^*\right)-\left(\left(1-\underline{a}^{\alpha}\right)\left(\phi H+H^*\right)-(1-\phi)(1-a_R^{\alpha})H^*\right) < 0.$$

We therefore have that $\frac{dF}{dH} > 0$.

Using (38) we get that

$$\frac{da_R}{dH} < 0$$

if the labor to capital ratio is sufficiently high and if $\phi = 0$.

8.4 Proposition 3: Proof $\frac{da_R}{dH} > 0$ in modern case

$$F \equiv (1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{(L+H)a_R^{1-\sigma}+a_R^{\alpha}H^*}{(1-\underline{a}^{\alpha})(H+\phi H^*)+(1-\phi)a_R^{\alpha}H^*}-\frac{(L^*+H^*)a_R^{1-\sigma}-a_R^{\alpha}H^*}{(1-\underline{a}^{\alpha})(\phi H+H^*)-(1-\phi)a_R^{\alpha}H^*}\right)$$
$$+\frac{1}{(\sigma-1)}\ln\frac{(1-\underline{a}^{\alpha})(H+\phi H^*)+(1-\phi)a_R^{\alpha}H^*}{(1-\underline{a}^{\alpha})(\phi H+H^*)-(1-\phi)a_R^{\alpha}H^*}$$

We here investigate the sign of $\frac{da_R}{dH}$, and

$$\frac{da_R}{dH} = -\frac{\frac{dF}{dH}}{\frac{dF}{da_R}}, \tag{39}$$

where $\frac{dF}{da_R} > 0$ (see previous Proof for Proposition 2).

Next

$$\frac{dF}{dH} = (1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{a_R^{1-\sigma}\Delta-((L+H)a_R^{1-\sigma}+a_R^{\alpha}H^*)}{\Delta^2}+\frac{((L+H)a_R^{1-\sigma}+a_R^{\alpha}H^*)\phi}{\Delta^{*2}}\right)+\frac{1}{(\sigma-1)}\left(\frac{1}{\Delta}-\frac{\phi}{\Delta^*}\right)$$
$$> (1-\phi)\frac{(1-\underline{a}^{\rho})}{\sigma\lambda}\left(\frac{\Delta-((L+H)a_R^{1-\sigma}+a_R^{\alpha}H^*)(1-\phi)}{\Delta^2}\right)+\frac{1}{(\sigma-1)}\left(\frac{1}{\Delta}-\frac{\phi}{\Delta^*}\right)$$

Now $\frac{1}{\Delta}-\frac{\phi}{\Delta^*} > 0$ since

$$\phi\Delta-\Delta^* = \phi((1-\underline{a}^{\alpha})(H+\phi H^*)+(1-\phi)a_R^{\alpha}H^*)-((1-\underline{a}^{\alpha})(\phi H+H^*)-(1-\phi)a_R^{\alpha}H^*) < 0.$$

Moreover

$$\Delta-((L+H)a_R^{1-\sigma}+a_R^{\alpha}H^*)(1-\phi) = \left((H+\phi H^*)(1-\underline{a}^{\alpha})+(1-\phi)a_R^{\alpha}H^*-(1-\phi)((L+H)+a_R^{\rho}H^*)\right) > 0.$$

We therefore have that $\frac{dF}{dH} > 0$, which from (39) gives:

$$\frac{da_R}{dH} > 0.$$

8.5 Proof of separation of equilibria

At ϕ^{SD}, $V_i - V_i^* = 0$

$$V_i - V_i^* = \frac{\mu(1-\underline{a}^{\rho})}{\sigma\lambda(1-\underline{a}^{\alpha})}(1-\phi^{SD})\frac{1}{H+H^*}\left(L+H+H^*-\frac{L^*}{\phi^{SD}}\right)\underline{a}^{1-\sigma}-\frac{\mu}{(\sigma-1)}\ln\phi^{SD}-\chi = 0.$$

Let us discuss a case where the development case of full agglomeration arises before the switch point, i.e. $\phi^{SD} < \widetilde{\phi}$. Since the agglomeration rent is positive above the sustain point, we get $V_i - V_i^*|_{\widetilde{\phi}} > 0$:

$$V_i - V_i^* = \frac{\mu \left(1 - \underline{a}^{\rho}\right)}{\sigma\lambda \left(1 - \underline{a}^{\alpha}\right)}(1 - \widetilde{\phi})\frac{1}{H + H^*}\left(L + H + H^* - \frac{L^*}{\widetilde{\phi}}\right)\underline{a}^{1-\sigma} - \frac{\mu}{(\sigma - 1)}\ln\widetilde{\phi} - \chi > 0.$$

Since $L + H + H^* - \frac{L^*}{\widetilde{\phi}} = 0$ at $\widetilde{\phi}$, we get

$$-\frac{\mu}{(\sigma - 1)}\ln\widetilde{\phi} - \chi > 0.$$

Pugging in $\widetilde{\phi}$, from (18) we get

$$\frac{HL^* - H^*L}{H^2 + HL - H^{*2} - H^*L^*} < e^{\frac{1-\sigma}{\mu}\chi}.$$