Agglomeration and regional growth

Richard E. Baldwin and Philippe Martin

Graduate Institute of International Studies (Geneva) ; University of Paris1 Panthéon-Sorbonne, CERAS-ENPC (Paris) and CEPR

1. INTRODUCTION

Spatial agglomeration of economic activities on the one hand and economic growth on the other hand are processes difficult to separate. Indeed, the emergence and dominance of spatial concentration of economic activities is one of the facts that Kuznets associated with modern economic growth. This strong positive correlation between growth and geographic agglomeration of economic activities has been documented by economic historians (Hohenberg and Lees, 1985 for example), in particular in relation to the industrial revolution in Europe during the nineteenth century. In this case, as the growth rate in Europe as a whole sharply increased, agglomeration materialized itself in an increase of the urbanization rate but also in the formation of industrial clusters in the core of Europe that have been by and large sustained until now. The role of cities in economic growth and technological progress has been emphasized by urban economists (Henderson, 1988, Fujita and Thisse, 1996), development economists (Williamson, 1988) as well as by economists of growth (Lucas, 1988). At the other hand of the spectrum, as emphasized by Baldwin, Martin and Ottaviano (2001), the growth takeoff of Europe took place around the same time (end of eighteenth century) as the sharp divergence between what is now called the North and the South: growth sharply accelerated (for the first time in human economic history) at the same time as a dramatic and sudden process of agglomeration took place at the world level. Hence, as put by Fujita and Thisse (2002b), "agglomeration can be thought as the territorial counterpart of economic growth."

Less dramatically and closer to us, Quah's results (1996) suggest also a positive relation between growth and agglomeration. He finds that among the Cohesion group of countries (Greece, Spain, Portugal and Ireland, though there are no Irish regional data), the two countries that have achieved a high rate of growth and converged in per capita income terms towards the rest of Europe (Spain and Portugal) have also experienced the most marked regional divergence, This is consistent with the results of De la Fuente and Vives (1995), for instance, building on the work of Esteban (1994) who suggest that countries have converged in Europe but that this process of convergence between countries took place at the same time as regions inside countries either failed to converge or even diverged. There are however few direct empirical tests of the relation between agglomeration and growth. Ciccone (2002) analyses the effects of employment density on average labour productivity for 5 European countries at the Nuts 3 regional level. He finds that an increase in agglomeration has a positive effect on the growth of regions. An indirect test of the relationship is performed in the literature on localized technology spillovers. The presence of localized spillovers has been well documented in the empirical literature. Studies by Jacobs (1969) and more recently by Jaffe et al. (1993), Coe and Helpman (1995 and 1997), Ciccone and Hall (1996) provide strong evidence that technology spillovers are neither global nor entirely localized. The diffusion of knowledge across regions and countries does exist but diminishes strongly with physical distance which confirms the role that social interactions between individuals, dependent on spatial proximity, have in such diffusion. A recent study by Keller (2002) shows that even though technology spillovers have become more global with time, " technology is to a substantial degree local, not global, as the benefits from spillovers are declining with distance. » The fact that technology spillovers are localized should in theory lead to a positive link between growth and spatial agglomeration of economic activities as being « close » to innovation clusters has a positive effect on productivity. Hence, these empirical results point to the interest of studying growth and the spatial distribution of economic activities in an integrated framework. From a theoretical point of view, the interest should also be clear.

There is a strong similarity between models of endogenous growth and models of the "new economic geography". They ask questions that are related: one of the objectives of the first field is to analyze how new economic activities emerge through technological innovation; the second field analyzes how these economic activities choose to locate and why they are so spatially concentrated. Hence, the process of creation of new firms/economic activities and the process of location should be thought as joint processes. From a methodological point of view, the two fields are quite close as they both assume (in some versions) similar industrial structures namely, models of monopolistic competition which reflects the role of economies of scale in both fields.

In this chapter, we will attempt to clarify some of the theoretical links between growth and agglomeration. Growth, in the form of innovation, can be at the origin of catastrophic spatial agglomeration in a cumulative process à la Myrdal. One of the surprising features of the Krugman (1991) model, was that the introduction of partial labour mobility in a standard "new trade model" with trade costs could lead to catastrophic agglomeration. The growth analog to this result is that the introduction of endogenous growth in the same type of "new trade model" can lead to the same result. A difference with the labour mobility version is that all the results are derived analytically in the endogenous growth version. Growth also alters the process of location even without catastrophe. In particular, and contrary to the fundamentally static models of the "new economic geography", spatial concentration of economic activities may be consistent with a process of delocation of firms towards poor regions.

The relation between growth and agglomeration depends crucially on capital mobility. Without capital mobility between regions, the incentive for capital accumulation and therefore growth itself is at the heart of the possibility of spatial agglomeration with catastrophe. In the absence of capital mobility, some results are in fact familiar to the New Economic Geography (Fujita, Krugman and Venables, 1999): a gradual lowering of transaction costs between two identical regions first has no effect on economic geography but at some critical level induce catastrophic agglomeration. In the model presented in this chapter, in the absence of migration, "catastrophic" agglomeration means that agents in the south have no more private incentive to accumulate capital and innovate. The circular causality which gives rise to the possibility of a Core-Periphery structure is depicted below and as usual in economic geography models is characterized by both production and demand shifting which reinforce each other. The production shifting takes the form of capital accumulation in one region (and de-accumulation in the other) and the demand shifting takes the form of increased permanent income due to investment in one region (and a decrease in permanent income in the other region).
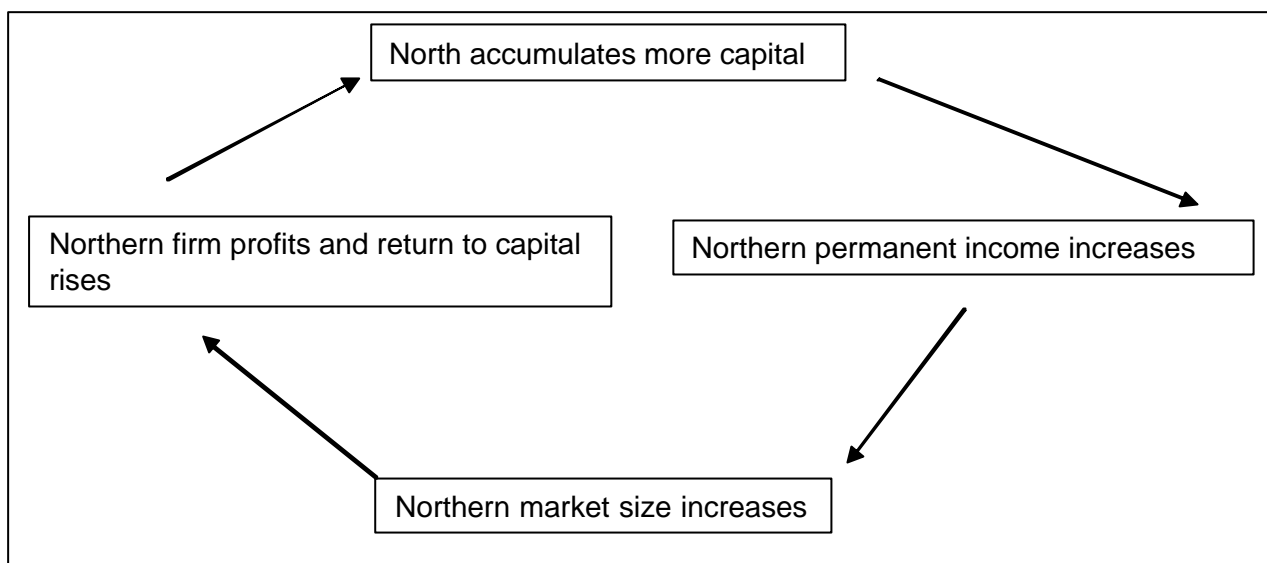
Figure 1: Demand-linked circular causality (a.k.a. backward linkages)

Capital mobility eliminates the possibility of catastrophic agglomeration because in this case production shifting does not induce demand shifting as profits are repatriated. It is therefore stabilizing in this sense. This is in sharp contrast with labour mobility which we know to be destabilizing. However, capital mobility also makes the initial distribution of capital between the two regions a permanent phenomenon so that both the symmetric and the Core-Periphery equilibria are always stable.

In a second section of this chapter, we will concentrate on the opposite causality running from spatial concentration to growth. For this, we will introduce localized technology spillovers which will imply that the spatial distribution of firms will have an impact on the cost of innovation and therefore the growth rate.

This chapter uses modified versions of Baldwin (1999), Baldwin, Martin and Ottaviano (2000) and Martin and Ottaviano (1999). The first two papers analyze models of growth and agglomeration without capital mobility. In contrast to the first paper which uses an exogenous growth model, this chapter analyses endogenous growth. In contrast to the second paper, we restrict our attention to the case of global technology spillovers. The last paper presents a model of growth and agglomeration with perfect capital mobility. Baldwin et al. (2003) also treat some common themes in chapters 6 and 7.

2. THE BASIC FRAMEWORK OF GROWTH AND AGGLOMERATION

Many of the most popular economic geography models focus on labour, examples being Krugman (1991), Krugman and Venables (1995), Ottaviano, Tabuchi and Thisse (2002) and Puga (1999). These are unsuited to the study of growth. The key to all sustained growth is the accumulation of human capital, physical capital and/or knowledge capital – with the accumulation of knowledge capital, i.e. technological progress having a privileged position. We thus need a model in which capital exists and its stock is endogenous.

To present the basic elements of this literature, we organise the discussion with the help of a workhorse model. As Baldwin et al (2003) show, introducing capital into a geography model is relatively simple. The simplest way is accomplished by the 'footloose capital' model (FC model) due to Martin and Rogers (1995). The FC model, however, takes the capital stock as given. Getting to a growth model requires us to add in a capital-producing sector.

Specifically we denote capital as K and labour of L. The capital-producing sector is referred to as the sector I (for innovation and investment, see below) and this comes on top of the two usual sectors, manufactures M and traditional-goods T. The regions (two of them) are symmetric in terms of preferences, technology and trade costs. The usual Dixit-Stiglitz M-sector (manufactures) consists of differentiated goods. Another difference is that the fixed cost is in terms of K. Each variety requires one unit of capital which can be interpreted as an idea, a new technology, a patent, machinery etc.. Production also entails a variable cost (a_M units of labour per unit of output). Its cost function, therefore, is $\boldsymbol{p} + w\, a_M x_i$, where $\boldsymbol{p}$ is K's rental rate, w is the wage rate, and x_i is total output of a typical firm. Traditional goods, which are assumed to be homogenous, are produced by the T-sector under conditions of perfect competition and constant returns. By choice of units, one unit of T is made with one unit of L. The structure of the basic growth and agglomeration model is in figure 2.

Regional labour stocks are fixed and immobile, so that we eliminate one possible source of agglomeration. Each region's K is produced by its I-sector. I is a mnemonic for innovation when interpreting K as knowledge capital, for instruction when interpreting K as human capital, and for investment-goods when interpreting K as physical capital. One possible interpretation of the difference between the situation of capital mobility and one of capital immobility is that in the first case K is physical capital (mobility then means the delocation of plants) or as knowledge capital that is marketable and tradable through patents. The second case, capital immobility, would be more consistent with the interpretation of human capital. In this case, labour immobility implies capital immobility. The I-sector produces one unit of K with a_I units of L, so that the marginal cost of the I sector, F, is $w\, a_I$. Note that this unit of capital in equilibrium is also the fixed cost of the manufacturing sector. As one unit of capital is required to start a new variety, the number of varieties and of firms at the world level is simply the capital stock at the world level: $K^W = K + K^*$. We note n and n^* the number of firms located in north and south respectively. As one unit of capital is required per firm we also know that: $K^W = n + n^*$. However, depending on the assumption we make on capital mobility, the stock of capital produced and owned by one region may or may not be equal to the number of firms producing in that region. In the case of capital mobility, the capital may be produced in one region but the firm that uses this capital unit may be operating in another region. Hence, the number of firms located in one region is, in the case of capital mobility, different from the stock of capital owned by this region.
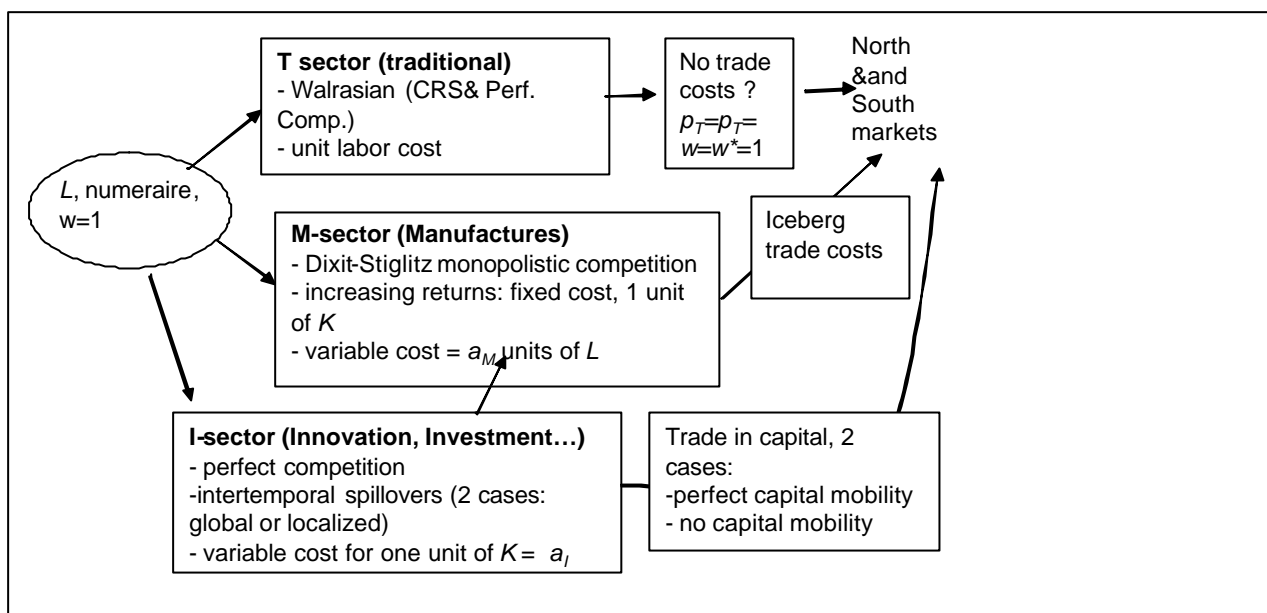
Figure 2: The basic structure of the growth and agglomeration model

To individual I-firms, the innovation cost a_I is a parameter, however following Romer (1990) and Grossman and Helpman (1991), a sector-wide learning curve is assumed. That is, the marginal cost of producing new capital declines (i.e., a_I falls) as the sector's cumulative output rises. Many justifications of this intertemporal externality, classic in the endogenous growth literature, are possible. Romer (1990), for instance, rationalizes it by referring to the non-rival nature of knowledge. We can summarize these standard assumptions on this literature by the following:

$$\dot{K} = \frac{L_I}{a_I} \quad ; \quad F = wa_I \quad ; \quad a_I = 1/K^W \quad ; \quad K^W = K + K^*$$

(1)

where K and K^* are the northern and southern cumulative I-sector production levels. Note that spillovers are global: the North learns as much from an innovation made in the South than in the North. Below, we introduce localized technological spillovers. Following Romer (1990) and Grossman and Helpman (1991), depreciation of knowledge capital is ignored[1]. Finally, the regional K's represent both region-specific capital stocks and region-specific cumulative I-sector production. Because the number of firms, varieties and capital units is equal the growth rate of the number of varieties, on which we will focus, is therefore: $\dot{K}^W / K^W = g$.

We assume an infinitely-lived representative consumer (in each country) with preferences:

$$U = \int_{t=0}^{\infty} e^{-rt} \ln Q \, dt \quad ; \quad Q = C_T^{1-a} C_M^a \quad ; \quad C_M = \left(\int_{i=0}^{K+K^*} c_i^{1-1/s} \, di \right)^{\frac{1}{1-1/s}}$$

(2)

where r is the rate of time preference, σ is the constant elasticity of substitution among varieties,

---

1 See Baldwin et al. (2003) for a similar analysis with depreciation.

and the other parameters have the usual meaning. Utility optimization implies that a constant fraction α of total northern consumption expenditure E falls on M-varieties with the rest spent on T. Optimisation by agents in the North also yields unitary elastic demand for T and the CES demand functions for M varieties. The optimal northern consumption path also satisfies the standard Euler equation with log utility which requires[2] $\dot{E}/E = r - \mathbf{r}$ (r is the north's rate of return on investment) and a transversality condition. Southern optimization conditions are isomorphic.

On the supply side, free trade in T equalizes nominal wage rates as long as both regions produce some T (i.e. if $\mathbf{a}$ is not too large). Taking northern labour as numeraire then $w=w^*=1$. As for the M-sector, units are chosen such that $a_M = 1-1/\mathbf{s}$ so that producer prices of varieties are also normalized to 1. With monopolistic competition, equilibrium operating profit is the value of sales divided by $\mathbf{s}$. Using the goods market equilibrium and the optimal pricing rules, the operating profits are given by:

$$\mathbf{p} = bB\frac{E^w}{K^w} \;; \qquad B \equiv \frac{s_E}{s_n + \mathbf{f}(1-s_n)} + \frac{\mathbf{f}(1-s_E)}{\mathbf{f}s_n + 1 - s_n}; \qquad b \equiv \frac{\mathbf{a}}{\mathbf{s}}, \qquad \mathbf{f} \equiv t^{1-\mathbf{s}}$$

$$\mathbf{p}^* = bB^*\frac{E^w}{K^w} \;; \quad B^* \equiv \frac{\mathbf{f}s_E}{s_n + \mathbf{f}(1-s_n)} + \frac{1-s_E}{\mathbf{f}s_n + 1 - s_n}$$

(3)

where $s_E \: ^o \: E/E^w$ is north's share of world expenditure E^w; $s_n = n/(n+n^*)$ is the share of firms which are located in the north, and $0=\phi=1$ is the usual transformation of transaction costs such that ϕ measures the "free-ness" (phi-ness of trade), with $\phi=0$ implying zero free-ness and $\mathbf{f}=1$ implying perfect free-ness (zero trade costs). When capital is immobile, this share is the share of capital owned by the Northern region: s_K. Also, B is a mnemonic for the 'bias' in northern M-sector sales since B measures the extent to which the value of sales of a northern variety exceeds average operating profit per variety worldwide (namely, bE^w/K^w).

3. THE CASE WITHOUT LOCALIZED SPILLOVERS: GROWTH MATTERS FOR GEOGRAPHY

As we shall see, the localisation of the learning spillovers drive growth is a major concern. We start with the simple extreme case considered by Grossman and Helpman (1991) where spillovers are perfectly global. This assumption is already embedded in equation (1).

---

[2] See Barro and Sala-I-Martin 1995 for a derivation using the Hamiltonian approach. Intuitively, the marginal cost of postponing consumption is ρ plus the rate of decline of marginal utility which, given the log preferences is just,

$\dot{E}/E$. The marginal benefit is r, the rate of return on investment. The optimal consumption path must be such that the two are equalized so that agents are indifferent to a small intertemporal reallocation of consumption.

3.1. *The growth equilibrium*

Since the location of innovation and production are irrelevant to the innovation process (knowledge spillovers are global and depend only on past I-sector production), the worldwide equilibrium growth rate can be determined without pinning down the spatial distribution of industry (the location equilibrium). The easiest and most intuitive way of solving for growth equilibria is to use Tobin's q (Baldwin and Forslid 2000). The essence of Tobin's approach is to assert that the equilibrium level of investment is characterized by the equality of the stock market value of a unit of capital – which we denote with the symbol v – and the replacement cost of capital, F. Tobin takes the ratio of these, so what micro economists would naturally call the M-sector free-entry condition (namely $v=F$) becomes Tobin's famous condition $q = v/F = 1$.

Calculating the numerator of Tobin's q (the present value of introducing a new variety) requires a discount rate. In steady state, $\dot{E}/E = 0$ in both regions[3], so the Euler equations imply that $r=r^*=\rho$. Moreover, the present value of a new variety also depends upon the rate at which new varieties are created. In steady state, the growth rate of the capital stock (or of the number of varieties) will be constant and will either be the common $g=g^*$ (in the interior case), or north's g (in the core-periphery case). In either case, the steady-state values of investing in new units of K are:

$$v = \frac{p}{r+g} \quad ; \quad v^* = \frac{p^*}{r+g} \tag{4}$$

It can be checked that the equality, $v=F$, is equivalent to the arbitrage condition present in endogenous growth models such as Grossman and Helpman (1991). The free entry condition in the innovation sector ensures that the growth rate of the value v of capital is equal to growth rate of the marginal cost of an innovation, F, which due to intertemporal spillovers is $-g$. With $r=\mathbf{r}$, and using the definition of F we get the regional q's:

$$q = \frac{p\, K^w}{(\mathbf{r}+g)} \quad ; \quad q^* = \frac{p^*\, K^w}{(\mathbf{r}+g)} \tag{5}$$

In the case of global spillovers, the common growth rate is easy to find because it does not depend on geography. The reason is simply that the cost of innovation and the total size of the market do not depend on the location of firms. Hence, we can just use the special case of the symmetric equilibrium where $s_E = s_n = 1/2$ to find the growth rate.

3.1.1 Endogenous growth and the optimal savings/investment relation

Using equation (3) in that case and imposing that Tobin's q is 1 in equation (5), we get the following relation between growth and world expenditure E^w: $bE^w = g + \mathbf{r}$ where b≡α/σ as is standard in the growth literature. It just says that higher expenditure by increasing profits induces more entry in the manufacturing sector, which implies a higher growth rate. The other equilibrium relation between growth and world expenditure is given by the world labour market

---

[3] To see this, use the world labour market equilibrium: $2L = aE^w(1-1/\mathbf{s}) + (1-a)E^w + g$ which says that world labour supply can be used either in the manufacturing sector, the traditional sector or the innovation sector. It implies that a steady state with constant growth only exists if E^w itself is constant.

equilibrium: $2L = aE^w(1 - 1/s) + (1-a)E^w + g$, which states that labour can be used either in the manufacturing sector (recall the unit labour requirement in this sector is normalized to $1-1/\sigma$), in the T sector or in the innovation sector ($\dot{K}^w$ is the production of the sector per unit of time and $F=1/K^w$ is the labour requirement in the innovation sector). Here the relation is negative as higher expenditure implies that labour resources are diverted from the innovation sector to the manufacturing and traditional sector.

Combining the two we find that the world level of expenditure is simply given by: $E^w = 2L + r$. Using these equations, the growth rate of the number of varieties and of the world capital stock is given by:

$$g = 2Lb - (1-b)r; \qquad b \equiv \frac{a}{s}$$

(6)

This shows that when knowledge spillovers are global in scope, the equilibrium growth depends positively on the size of the world economy (as measured by the endowment of the primary factor) and negatively on the discount rate. Importantly, the equilibrium g does not depend on geography.

Finally, a simple equilibrium relation exists between s_E and s_K, the northern share of expenditures and the northern share of capital ownership. It can be shown that optimizing consumers set expenditure at the permanent income hypothesis level in steady state. That is, they consume labour income plus r times their steady-state wealth, $FK= s_K$, and, $FK^*= (1- s_K)$ in the north and in the south respectively. Hence, $E = L+r s_K$, and $E^* = L+r(1-s_K)$. Note that this is another way to check the level of world expenditure as: $E + E^* = E^w = 2L + r$. Thus, we get:

$$s_E \equiv \frac{E}{E^w} = \frac{L + r\, s_K}{2L + r} = \frac{1}{2} + (\frac{r}{2L + r})(s_K - \frac{1}{2})$$

(7)

This relation between s_E and s_K, can be thought as the optimal savings/expenditure function since it is derived from intertemporal utility maximization. The intuition is simply that an increase in the northern share of capital increases the permanent income in the north and leads therefore to an increase in the northern share of expenditures.

3.1.2 The role of capital mobility

Having worked out the equilibrium growth rate, and thus implicitly defined the amount of resources devoted to consumption, we can turn to working out the spatial division of industry, i.e., the location equilibrium. From now on two roads are open:

1) we can let capital owners decide where to locate production. Capital is mobile even though capital owners are not, so that profits are repatriated in the region where capital is owned. In this case, s_n, the share of firms located in the north and s_K, the share of capital owned by the north, may be different. s_n is then endogenous and determined by an arbitrage condition that says that location of firms is in equilibrium when profits are equalized in the two regions. Because of capital mobility, the decision to accumulate capital will be identical in both regions so that the initial share of capital owned by the north, s_K, is permanent and entirely determined the initial distribution of capital ownership between the two regions.

2) a second solution is to assume that capital is immobile. Presumably, this would be the case if we focus on the interpretation of capital being human (coupled with immobile agents). In this case, the location of production, s_n, is pinned down by capital ownership: $s_n = s_K$.

As we shall see in detail below, the capital mobility assumption is pivotal. Why is this? In standard terminology, allowing capital mobility eliminates demand-linked circular causality (a.k.a. backward linkages); capital moves without its owners, a shift in production leads to no expenditure shifting because profits are repatriated. When capital is immobile, any shock which favours production in one region is satisfied by the creation of new capital in that region. Since the income of the new capital is spent locally, the 'production shifting' leads to 'expenditure shifting'. Of course, expenditure shifting fosters further production shifting (via the famous home market effect), so without capital mobility, the model features demand-linked circular causality. As is well known, this form of linkage is de-stabilising, so – as we shall see in detail below – capital mobility in a growth model is a <u>stabilizing</u> force (Baldwin 1999).

Because the case of capital mobility is simpler, we start with it.

3.2. Perfect capital mobility: the location equilibrium

With capital mobility, an obvious question arises: where does capital locate? Capital owned in one region can be located elsewhere. Again, the arbitrage condition, which implies that profits across regions need to be equal for firms to be indifferent between the two locations, pins down the equilibrium location of firms. Using equation (3), and imposing the equality of profits, we get that there is no more incentive for relocation when the following relation between s_n and s_E is satisfied:

$$s_n = \frac{1}{2} + \left(\frac{1+f}{1-f}\right)(s_E - \frac{1}{2}), \qquad 0 \le s_n \le 1$$

(8)

where the equilibrium s_n equals unity or zero when the s_n implied by (8) is outside the zero-unity range. This is an example of the "home market" effect. Since $(1+\phi)/(1-\phi)$ is greater than one, this relationship tells us that a change in market size leads to a more than proportional change in the spatial allocation of industry.

Combining equations (7) and (8), we get the equilibrium relation between the share of firms located in the north, s_n, and the share of capital owned by the north, s_K:

$$s_n = \frac{1}{2} + \left(\frac{r}{2L+r}\right)\left(\frac{1+f}{1-f}\right)(s_K - \frac{1}{2}) \qquad 0 \le s_n \le 1$$

(9)

Note also that if the initial distribution of capital in the north is such that $s_K > \frac{1}{2}$, then more firms will be located in the north than in the south: $s_n > \frac{1}{2}$. An increase in the share of capital in the north, s_K, induces relocation to the north as it increases expenditure and market size there. Note also that lower transaction costs (higher f) will reinforce the home market effect, implying that an unequal distribution of capital ownership will translate in an even more unequal distribution of firms.

3.2.1   Stability of the location equilibrium

It is easy to see that the division of industry described above will not change over time. With perfect capital mobility, operating profits have to be the same in both regions which also

implies that the value of capital has to be the same in both regions. Hence, $p = p*$ and $q=q*=1$. This, together with the assumption of constant returns to scale, and the assumption of global spillovers (implying that the cost of innovation is the same in both regions) means that the two regions will accumulate capital at the same constant rate so that any initial distribution of capital is stable. Moreover, since neither backward nor forward linkages operating in this model with capital mobility, no "catastrophic" agglomeration scenario can unfold (see Martin and Ottaviano 1999). Hence, the equilibrium described by (9) is always stable. In particular, the symmetric equilibrium where $s_n = s_K = 1/2$, is always stable for any level of transaction costs on trade in goods.

To see this point in more detail, one can analyze the effect of an exogenous increase in s_n, by a small amount and check the impact of this perturbation on the ratio of profits in the north to profits in the south. That is, ask the question whether an increase in geographic concentration in the north decreases or increases the incentive to relocate in the north. The symmetric equilibrium is stable, if and only if $\partial(p/p*)/\partial s_n$ is negative. Indeed this is the case for all positive levels of transaction costs since, evaluated at the equilibrium geography:

$$\frac{\partial\left(p/p*\right)}{\partial s_n} = -\frac{(1-f)^2}{(1+f)^2}\frac{1}{s_E(1-s_E)} < 0$$

Evaluated at the equilibrium given by (9), an exogenous increase in the share of firms located in the north always decreases relative profits there, so that it leads firms to go back to the south. The location equilibrium determined in (9) is always stable. The reason is that when more firms locate in the north, this increases competition there (and decreases it in the south).

3.2.2   Does capital flow from rich to poor?

An interesting question that can be analyzed in this framework is: "Do firms relocate towards the north or towards the south?" In economic geography models without growth, industrial concentration implies that firms are destroyed in the south and relocated in the north. Here, the relocation story is richer because of the constant creation of new firms. To see what is the direction of relocation we need to look at the difference between the share of capital owned by the north and the share of firms located in north. The expression is easier taking ratios, so:

$$s_K - s_n = 2\frac{L(1-f)-rf}{(1-f)(2L+r)}(s_K - \frac{1}{2})$$

(10)

In the symmetric equilibrium, where both regions are endowed originally with the same amount of capital there is no relocation of course. If the initial distribution of capital is such that $s_K > 1/2$, so that the north is richer than the south, then the direction of the capital flows is ambiguous; it depends on the sign of $L(1-f)-rf$. If this expression is positive, then $s_K > s_n$ so that some of the capital owned by the north relocates to the south.

The ambiguity of the direction of capital flows stems from the fact that it is governed by two opposite effects, namely the market crowding effect (which is a dispersion force that makes the poor capital region attractive because firms installed there face less competition), and the market access effect (which is an agglomeration force that makes the rich region attractive because of its high level of income and expenditure). The first effect dominates when trade is quite closed (f is low). Note that when the rate of time preference is high or more generally when

the return to capital is high, the capital rich region becomes more attractive because the market access effect is reinforced. There is a threshold level of transaction costs that determines the direction of capital flows. It is given by:

$$\boldsymbol{f}^{CP} = \frac{L}{L + \boldsymbol{r}}$$

(11)

When transaction costs are below this level, relocation takes place towards the south and vice-versa. The reason why we attach CP (for Core-Periphery) to this threshold will become clear later when we analyze the case of capital immobility, as we will see that this threshold value is the one for which the symmetric equilibrium becomes unstable.

An interesting feature here is that concentration of wealth and of economic activities in the north (s_K and $s_n > \frac{1}{2}$), is compatible with relocation of firms from north to south ($s_K < s_n$) when $\boldsymbol{f} < \boldsymbol{f}^{CP}$. This comes from the introduction of growth and the fact that a larger number of newly created firms are created and owned by the north than by the south.

3.3. *No capital mobility: "new growth" and "new geography"*

The previous section described a growth and geography equilibrium where agglomeration forces were present,[4] but where no "catastrophe" could take place since all circular causality had been ruled out. As discussed above, eliminating capital mobility in a growth model is actually de-stabilising since anything that changes the spatial allocation of industry and thus capital will simultaneously change that spatial allocation of expenditure. And, as is well known, the home market effect means that any change in expenditure's spatial allocation induces a knock-on change in the location of industry.

More formally, restricting capital mobility (together with the assumptions of labour immobility) has two implications. First, the number of firms and the number of units of capital owned in a region are identical: $s_n = s_K$. Second, because the arbitrage condition of the previous section does not hold, profits may be different in the two regions. This in turn implies that, contrary to the previous section, the two regions may not have the same incentive to accumulate capital so that the initial ownership of capital does not need to be permanent. This means that the analysis will be quite different from the previous section. We will ask the following questions which are the usual ones in the "new economic geography" models. Starting from an equal distribution of capital, the symmetric equilibrium, we will determine under which circumstances it remains a stable equilibrium. Then we will look at the Core-Periphery equilibrium and again ask when this equilibrium is stable.

3.3.1 Stability of the symmetric equilibrium

We first consider interior steady states where both regions are investing, so $q = 1$ and $q^* = 1$. Using (3) and (5) in (6), $q = q^* = 1$ and imposing $s_n = s_K$ we get:

---

[4] We define agglomeration as the phenomenon where the concentration of economic activity creates forces that foster the concentration of economic activity. The home market effect, which did operate in the pervious section, shows agglomeration forces are present since a division where $s_L = s_K > \frac{1}{2}$ would not be an equilibrium. Due to the home market effect, such a division would encourage further concentration of economic activity in the north.

$$s_K = \frac{1}{2} + (\frac{1+f}{1-f})(s_E - \frac{1}{2})$$

(12)

which of course is just (8) with s_n replaced by s_K. In other words, it now determines the location of capital ownership as well as the location of production. Together with equation (7) which implied that production shifting led to expenditure shifting, this defines a second positive relation between s_E and s_K, i.e. expenditure shifting leads to production shifting.

The intuition is that a relative increase in northern demand increases profits in the north and therefore the marginal value of an extra unit of capital. In other words, the numerator of Tobin's *q* increases in the north. Hence, this raises the incentive to innovate there and the north indeed increases its share of capital s_K. The intuition is therefore very close to the "home market effect" except that it influences here the location of capital accumulation. Together with the optimal saving relation of (7), it is easy to check that the symmetric solution $s_E = s_K = \frac{1}{2}$ is always an equilibrium, in particular it is an equilibrium for all levels of transaction costs. The symmetric equilibrium is the unique equilibrium for which both regions accumulate capital (*q* = *q* =1*). However, the fact that there are two positive equilibrium relations between s_E and s_K, the share of expenditures and the share of capital in the north, should warn us that the symmetric equilibrium may not be stable. Indeed, in this model a 'circular causality' specific to the presence of growth and capital immobility tends to de-stabilize the symmetric equilibrium because of the demand-linked cycle in which production shifting leads to expenditure shifting and *vice versa*. The particular variant present here is based on the mechanism first introduced by Baldwin (1999) in a neo-classical growth model.

There are several ways to study the symmetric equilibrium's stability. We can first graph the two equilibrium relations between s_E and s_K, the "Permanent Income" relation (call it PI) given by equation (7) and the "Optimal Investment" relation (call it OI) given by equation (12). In the case where the slope of the PI relation is less than the OI relation we get the left-panel of Figure 3. At the right of the permanent income relation, s_E, the share of expenditures in the north, is too low given the high share of capital owned by the north (agents do not consume enough). The opposite is true at the left of the PI relation. At the right of the optimal investment relation, s_K, the share of capital in the north, is too high given the low level of s_E, the share of expenditures in the north (agents invest too much). The opposite is true is at the left of the OI relation. This graphical analysis suggests that in this case the symmetric equilibrium is stable.

In the case where the slope of the PI relation is steeper than the OI then the same reasoning leads to the right-panel of the diagram. This suggests that in this case, the symmetric equilibrium is unstable. According to this graphical analysis, the transaction cost below which the symmetric equilibrium becomes unstable is exactly the one for which the slope of the PI curve equals the slope of OI curve. The slope of the PI curve is: *?/(2L+?)* which is the share of capital income in total income. The slope of the OI curve is: *(1-f)/(1+f)*. The two slopes are equal for a level of transaction costs which we saw above: it turns out to be the threshold level, which we define as f^{CP}, is given by equation (11), namely $f^{CP}=L/(L+?)$. When the "free-ness" of trade is higher than this level, our graphical analysis suggests that the stable equilibrium is not stable.
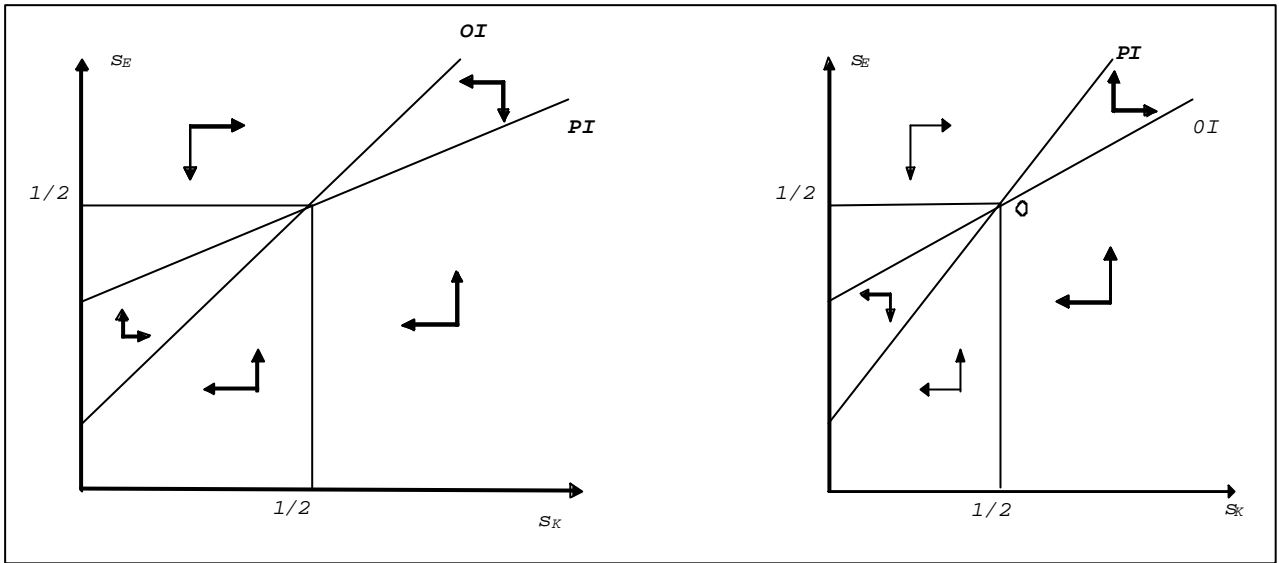
Figure 3: The northern shares of expenditure and capital, stable and unstable cases

To gain more intuition on this result, we can also study the symmetric equilibrium's stability in a different and more rigorous way. We can analyze the effect of an exogenous increase s_K, by a small amount and check the impact of this perturbation on Tobin's q, allowing expenditure shares to adjust according to (7). The symmetric equilibrium is stable, if and only if $\partial q/\partial s_K$ is negative: in this case, an increase in the share of northern capital lowers Tobin's q in the north (and therefore the incentive to innovate) and raises it in the south (by symmetry $\partial q/\partial s_K$ and $\partial q^*/\partial s_K$ have opposite signs). Thus when $\partial q/\partial s_K < 0$, the perturbation generates self-correcting forces in the sense that the incentive to accumulate more capital in the north falls and increases in the south. If the derivative is positive, the increase in the share of capital in the north reinforces the incentive to accumulate more capital in the north: the symmetric equilibrium is unstable in this case. Differentiating the definition of q with respect to s_K, we have:

$$\left(\frac{\partial q/q}{\partial s_K}\right)_{s_K=1/2} = 2\left(\frac{1-f}{1+f}\right)\left(\frac{\partial s_E}{\partial s_K}\right)_{s_K=1/2} - 2\left(\frac{1-f}{1+f}\right)^2$$

(14)

This expression illustrates the two forces affecting stability. The first term is positive, so it represents the destabilizing force, namely the ***demand-linked effect***. This effect was absent of the stability analysis in the case of capital mobility because an increase in profits in one region led to delocation of capital but not to more local capital accumulation. In the case of capital immobility, the only adjustment mechanism when profits increase in one region is that agents in that region accumulate more capital up to point where the profits of accumulating capital are driven back to zero. This "local" accumulation process generates a higher permanent income and a higher level of expenditure in one region only which is the reason for the circular causality.

The negative second term reflects the stabilizing ***market crowding effect,*** which was the only present in the capital mobility case. Clearly, reducing trade costs (an increase in f) erodes the stabilizing force more quickly than it erodes the destabilizing demand-linkage.

15

Using (7) to find $\P s_E/\P s_K = r/[2L+r\ ]$, the critical level of f at which the symmetric equilibrium becomes unstable is defined by the point where (13) switches sign. It is easy to check that again this critical level is given by f^{CP} of equation (11).

When trade costs are high the symmetric equilibrium is stable and gradually reducing trade costs produces standard, static effects – more trade, lower prices for imported goods, and higher welfare. There is, however, no impact on industrial location, so during an initial phase, the global distribution of industry appears unaffected. As trade free-ness moves beyond f^{CP}, however, the equilibrium enters a qualitatively distinct phase. The symmetric distribution of industry becomes unstable, and northern and southern industrial structures begin to diverge; to be concrete, assume industry agglomerates in the north. Since s_K cannot jump, crossing f^{CP} triggers transitional dynamics in which northern industrial output and investment rise and southern industrial output and investment fall. Moreover, in a very well defined sense, the south would appear to be in the midst of a 'vicious' cycle. The demand linkages would have southern firms lowering employment and abstaining from investment, because southern wealth is falling, and southern wealth is falling since southern firms are failing to invest. By the same logic, the north would appear to be in the midst of a 'virtuous' cycle.

3.3.2   The Core-Periphery equilibrium

In addition to the symmetric equilibrium, a core-periphery outcome (s_K =0 or 1, but we will focus only on the second one where the north gets the core) can also exist. For s_K =1 to be an equilibrium, it must be that $q = v/F = 1$ and $q^* = v^*/F^* <1$ for this distribution of capital ownership: continuous accumulation is profitable in the north since $v=F$, but $v^*<F^*$ so no southern agent would choose to setup a new firm. Defining the Core-Periphery equilibrium this way, it implies that it is stable whenever it exists. Using (3), (5) and (6), (7), q^* with s_K =1 simplifies to:

$$q^* = \frac{(1+f^2)L + f^2 r}{(2L+r)f}$$

(14)

If q^* is less than 1 when s_K =1, then the Core-Periphery equilibrium exists and is stable as there is no incentive for the south to innovate in this case. The threshold f that solves q*=1 defines the starting point of the core-periphery set. Again, this threshold is f^{CP} of equation (11). This implies that at the level of the transaction costs for which the symmetric equilibrium becomes unstable, the Core-Periphery becomes a stable equilibrium.

When transaction costs are high enough, the Core-Periphery equilibrium is not a stable equilibrium: in this case the south would have an incentive to innovate because the profits in the south are high enough. This is because even though the southern market is small in this case (it has no capital income in the Core-Periphery equilibrium), it is protected from northern competition thanks to high transaction costs. When transaction costs are low enough, this protection diminishes and the fact that the market in the south is small becomes more important: in this case, above the threshold f^{CP}, it becomes non profitable to operate a firm in the south.

Using s_K =1, the remaining aspects of the core-periphery steady state are simple to calculate. In particular, since s_K =1, q=1, and q*<1, we have that no labour is used in the innovation or manufacturing sectors in the south and all innovation is made in the north.

Note that the core-periphery outcome ($s_K=1$) is reached only asymptotically. This is because the stock of capital in the south does not depreciate and once the level of f^{CP} is crossed, stays constant, whereas the stock of capital in the north keeps growing at rate g. Figure 4 summarises the model's stability properties in a diagram with f and s_K on the axes:
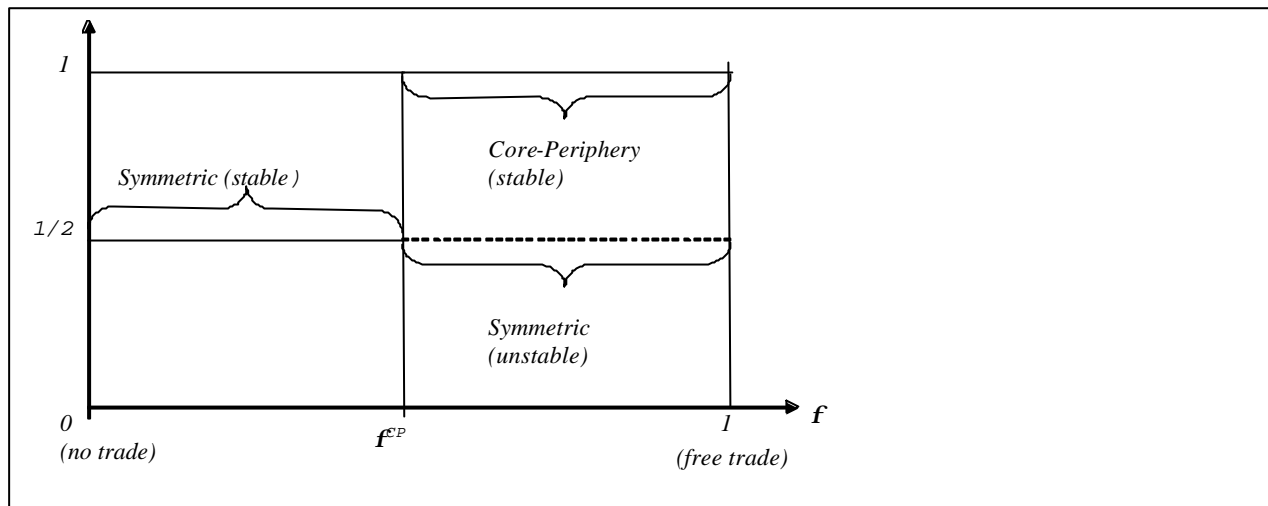

Figure 4: Stability properties of the CP equilibrium

Following the tradition of the "new economic geography" we have analyzed here the existence and stability conditions of the symmetric and Core-Periphery equilibria. In this simple model we can go further and analyze what would happen if we started from a situation in which the north had more capital than the south ($1/2 <s_K <1$). It can be checked, using equations (3), (5) and (6) that in this case $q <1$ (and $q^*>1$) if:

$$(1-s_K)(s_K - \frac{1}{2})(1-f)(rf - L(1-f)) < 0$$

that is if $f < f^{CP}$. Hence, in this case, the north would not innovate (the large stock of capital implies a high degree of market crowding) and the south would innovate. Hence, if we start from such an interior asymmetric equilibrium then one would converge back to the symmetric equilibrium as long as transaction costs are high enough. If $f > f^{CP}$, then the economy converges to the core-periphery equilibrium.

3.4. Concluding remark

Comparing perfect capital mobility to no capital mobility, we conclude that:

- when trade costs are high , the absence of capital mobility leads to convergence between the two regions: if one region starts with more capital than the other then, the two regions converge to the symmetric equilibrium. On the contrary, with capital mobility, any initial distribution of capital ownership becomes permanent. However, some of the firms owned by the north will relocate and produce in the south. This will produce some sort of convergence in terms of GDP but not in terms of GNP.

- when trade costs are low, the absence of capital mobility leads to divergence between the two regions[5]: asymptotically, whatever the initial distribution of capital, all the capital is accumulated and owned by one region. With capital mobility, as long as all the capital is not entirely owned by the north, some firms will still produce in the south. However, some of the southern capital will delocate to the north.

Hence, in the case of mobile capital (physical or tradable innovations such as patents), the key parameter for regional income distribution is the "exogenous" initial distribution of capital. In the case of immobile (human) capital, the key parameter is the level of transaction costs. The regional distribution of capital affects the long term regional income distribution "only" to the extent that it determines which region becomes the core, through a small initial advantage in capital endowments for example. To simplify matters we have used a model where only one type of capital exists. To make it more realistic, in particular for the European case, it would be interesting to extend it and take into account the different natures of capital so that part of the capital is mobile and part is not.

Can we derive some policy implications from this analysis? One striking result is that when regions are not well integrated (high transport/trade costs) capital immobility is conducive to regional convergence[6]. However, when regions are well integrated, the opposite result is true. To the extent that public policies can alter capital mobility, the policy implication is clear: capital mobility, both physical and human, should be facilitated between countries which are well integrated on the trade side. In the European context, this implies that the "single market" was right to foster free movement of goods and capital at the same time.

4. THE CASE WITH LOCALIZED SPILLOVERS: GEOGRAPHY MATTERS FOR GROWTH (AND VICE VERSA)

In the previous section, we showed that growth could dramatically alter economic geography in the sense that the process of accumulation of capital teamed with capital immobility could lead to catastrophic agglomeration. However, geography had no impact on growth. This was due to the fact that we assumed global spillovers: the learning curve, which as in any endogenous growth model, was at the origin of sustained growth, was global in the sense that the north and the south would learn equally from an innovation made in any region. In this section, we analyze how localized spillovers give a role in growth to the geography of production and innovation activities.

Localized spillovers are not the only way that geography can affect growth. Martin and Ottaviano (2001) generate a feedback between growth and agglomeration by assuming vertical

---

[5] This result however is not general. Urban (2002) integrates a neo-classical growth model into a static geography model without physical capital mobility. Contrary to the models presented here, he shows that lower transaction costs lead to convergence between the poor and the rich country. The reason is that the classic local decreasing returns effect implies that there is more incentive to accumulate capital in the poor country and in his model this effect does not depend on transaction costs. On the contrary, the home market effect, the divergence force, decreases as transaction costs diminish.

6 Basevi and Ottaviano (2002) modify this type of model to investigate the intermediate situation in which capital mobility is neither absent nor perfectly free.

linkages rather than local spillovers in innovation. Because the innovation sector uses manufacturing goods as an input, the location of manufacturing affects the cost of innovation through trade costs. Yamamoto (2002) presents a similar model with circular causation between growth and agglomeration coming from the vertical linkages between the intermediate goods sector and the innovation sector. A different type of geography and growth model where trade is absent is proposed by Quah (2002). The knowledge spillovers are imperfect both across space and time so that quite intuitively spatial clusters can appear. The reasoning is not very different from Grossman and Helpman (1991) who show that when knowledge spillovers are localized the increasing returns activity concentrates in one location.

4.1. *Necessary extensions of the basic model*

Introducing localized technological spillovers requires a minor modification to one of the assumptions made in the previous section[7]. Equation (1) that described the innovation sector assumed global spillovers in the sense that the marginal cost of an innovation, identical in both regions, was: $F = wa_I = 1/K^W$, so that it was decreasing in the total stock of existing capital; in the Grossman and Helpman (1991), spillovers were global. Grossman and Helpman (1991) also consider the other polar of extreme where knowledge spillovers are purely local. Since the "geography of knowledge" is an important topic for policy makers and a subject that has attracted a great deal of empirical work, it is more convenient to allow non-polar assumptions concerning knowledge spillovers as introduced by Baldwin and Forslid (2000). Specifically, suppose that these spillovers are localized in the sense that the cost of R&D in one region also depends on the location of firms (stock of knowledge capital). Hence, the northern cost of innovation depends more on the number of firms located in the north than in the south so that equation (1) becomes (taking into account that the wage rate is equal to 1):

$$F = a_I \quad ; \quad a_I \equiv \frac{1}{K^W A}; \qquad A \equiv s_n + \lambda\,(1 - s_n), \qquad 0 \leq \lambda \leq 1$$

(15)

where λ (a mnemonic for learning spillovers) measures the degree of to which learning from knowledge creation in one region facilitates knowledge creation in the other region. The fully global spillovers case is where $\lambda=1$; the fully local case is $\lambda=0$. To put it differently, the higher the transaction costs on the mobility of ideas, technologies, and innovations. The cost function of the innovation sector in the south is isomorphic.

Again the case of perfect capital mobility is easier than the case without capital mobility. Hence, we will start with the former following some of the analysis of Martin and Ottaviano (1999)[8] and then describe the model without capital mobility following Baldwin, Martin and Ottaviano (2001).

---

7 Here, localized technology spillovers are assumed. Duranton and Puga (2001) provide micro-foundations for the link between local diversity and innovation in a model with localized spillovers. Firms that innovate locate in diversified cities and then relocate in specialized cities to commence mass production.

8 The set of basic results is enriched by other contributions Ottaviano (1996) as well as Manzocchi and Ottaviano (2001) extend the model of Martin and Ottaviano (1999) to a three-region economy.

4.2. *The case of perfect knowledge capital mobility*

There are three endogenous variables that we are interested in: the growth rate g which is common to both regions in steady state; s_n, the share of firms that are producing in the north; and s_E, the share of expenditure in the north which also can be thought as a measure of income inequality between north and south. Remember that with perfect capital mobility, s_K the share of capital in the north is given by the initial distribution of capital as the stocks of capital in both regions are growing at the same rate. We want to find the different equilibrium relations between these three endogenous variables.

Due to localized spillovers, it is less costly to innovate in the region with the highest number of firms (which represent also capital or innovations). This implies that, because of perfect capital mobility, all the innovation will take place in the region with a higher number of firms[9]. Remember that due to perfect competition the value of an innovation is equal to its marginal cost. The shares of firms are perfectly tradable across regions (perfect capital mobility) so the value of capital (or firms) cannot differ from one region to another and no innovation will take place in the south. But the south will be able to simply buy (without transaction costs) innovations or capital produced in the north. Hence, in the case when $s_K > 1/2$, that is when the initial stock of capital is higher in the north than in the south, we know from the previous section that this will imply that more firms will be located in the north ($s_n > 1/2$) so that all innovation will take place in the north. In this case the world labour market equilibrium will be given by:

$$2L = \frac{g}{s_n + \boldsymbol{l}\,(1 - s_n)} + \boldsymbol{a}\,\frac{\boldsymbol{s} - 1}{\boldsymbol{s}} E^W + (1 - \boldsymbol{a})E^W \tag{16}$$

Remember also that world expenditure is given by: $E^W = 2L + \boldsymbol{r}FK^W$. The value and marginal cost of capital is given by F in (15). Using this and equation (16), we get the growth rate of capital g as a function of s_n, our first equilibrium relation:

$$g = 2bL[s_n + \boldsymbol{l}\,(1 - s_n)] - \boldsymbol{r}(1 - b) \qquad 1/2 < s_n \leq 1 \tag{17}$$

Compared to the growth rate derived in the previous section, this one differs because of the presence of localized spillovers: spatial concentration of firms (a higher s_n) implies a lower cost of innovation and therefore a higher growth rate[10]. Note also that for a given geography of production (a given s_n), less localized spillovers (a higher λ) also implies a lower cost of innovation in the north (as the innovation sector in the north benefits more from spillovers of firms producing in the south) and a higher growth rate.

The arbitrage condition consistent with the assumption of perfect capital mobility requires profits to be equalized in the two locations so that $\boldsymbol{p} = \boldsymbol{p}* = bE^W/K^W$. This gives the same equilibrium relation between s_n and s_E as in the previous section (equation 8).

---

[9] Fujita and Thisse (2002a and 2002b, chapter 11) show in a version where mobile skilled workers are at the origin of innovation, that this sector is always agglomerated in one region. This is the same type of mechanism that applies here: if capital (in the form of shares, patents etc…) is costlessly tradable and there are localized spillovers that lower the cost of innovation in the region with the highest number of innovators, then the symmetric spatial configuration is never stable.

[10] The result that spatial concentration is good for growth is also present in the model of growth with mobility of skilled workers by Fujita and Thisse (2002a and 2002b, chapter 11).

To find the third equilibrium relation, one between s_E and g, remember that due to intertemporal optimization, $E = L + rvK$ where v is the value of capital which itself is equal to the discounted value of future profits. Using these relations, it is easy to get the last equilibrium relation:

$$s_E = \frac{1}{2} + b\frac{r}{g+r}(s_K - \frac{1}{2})$$

(18)

Note that income inequality between the two regions is decreasing in the growth rate as long as the north is initially richer than in the south in capital stocks ($s_K > \frac{1}{2}$). This is because the value of capital decreases with growth due to faster entry of new firms.

The equilibrium characterized by these three relations is stable for the same reasons as in the case of perfect capital mobility of the previous section. Capital mobility allows southerners to save and invest buying capital accumulated in the north (in the form of patents or shares). Hence, the lack of an innovation sector does not prevent the south from accumulating capital: the initial inequality in wealth does not lead to self-sustaining divergence. No "circular causality" mechanism which would lead to a core-periphery pattern, as in the "new geography" models of the type of Krugman (1991), will occur.

Using equations (8), (17) and (18), the equilibrium is the solution to a quadratic equation given in appendix I. One can find the transaction cost such that relocation goes from north to south in the case where $s_K > 1/2$ (which implies also that $s_n > \frac{1}{2}$). $s_K > s_n$ if:

$$f < \frac{lL(1-s_K) + Ls_K}{lL(1-s_K) + Ls_K + r}$$

Note that when all the capital is owned by the north ($s_K = 1$), then the threshold level of transaction cost is again f^{CP} given in the previous section. Note also that in the less extreme case where $s_K < 1$, less localized spillovers imply, everything else constant, that relocation will take place towards the south. The reason is that less localized spillovers imply a lower cost of innovation in the north, and therefore a lower value of capital of which the north is better endowed with. Hence, less localized spillovers generate, for a given distribution of capital, a more equal distribution of incomes and expenditures and therefore attract firms in the south.

One could analyze the properties of this equilibrium by analyzing the equilibrium location s_n given in appendix. However, it is more revealing to use a graphical analysis.

Equation (8) provides a positive relation between s_n and s_E, the well known "demand-linked" effect. In figure 5, this relation is given by the curve s_n (s_E) in the NE quadrant. Equation (17) provides a positive relation between g and s_n. This is the localized spillovers effect: when industrial agglomeration increases in the region where the innovation sector is located, the cost of innovation decreases and the growth rate increases. This relation is given by the line $g(s_n)$ in the NW quadrant. Finally, equation (19) provides a negative relation between s_E and g. This is a "competition" effect: the monopoly profits of existing firms decrease as more firms are created; as the north is more dependent on this capital income, the northern share of income and expenditures decreases. This relation is described by the curve s_E (g) in the SE quadrant.

This graph can be used to analyze the relation between the geography of income, the geography of production and growth.

4.2.1   Spatial equity and efficiency

An increase in regional inequality in capital endowments s_K shifts to the right the $s_E(g)$ in the SE quadrant. The impact is therefore an increase in income inequality and an increase in spatial inequality in the sense that s_n increases. However, because the economic geography becomes less dispersed and therefore more efficient from the point of view of localized technology spillovers, the growth rate g is higher. Hence the introduction of growth and localized spillovers in a geography model is at the origin of a trade-off between spatial equity and efficiency (see Martin 1999 for an analysis along these lines) which may have important implications for public policies.
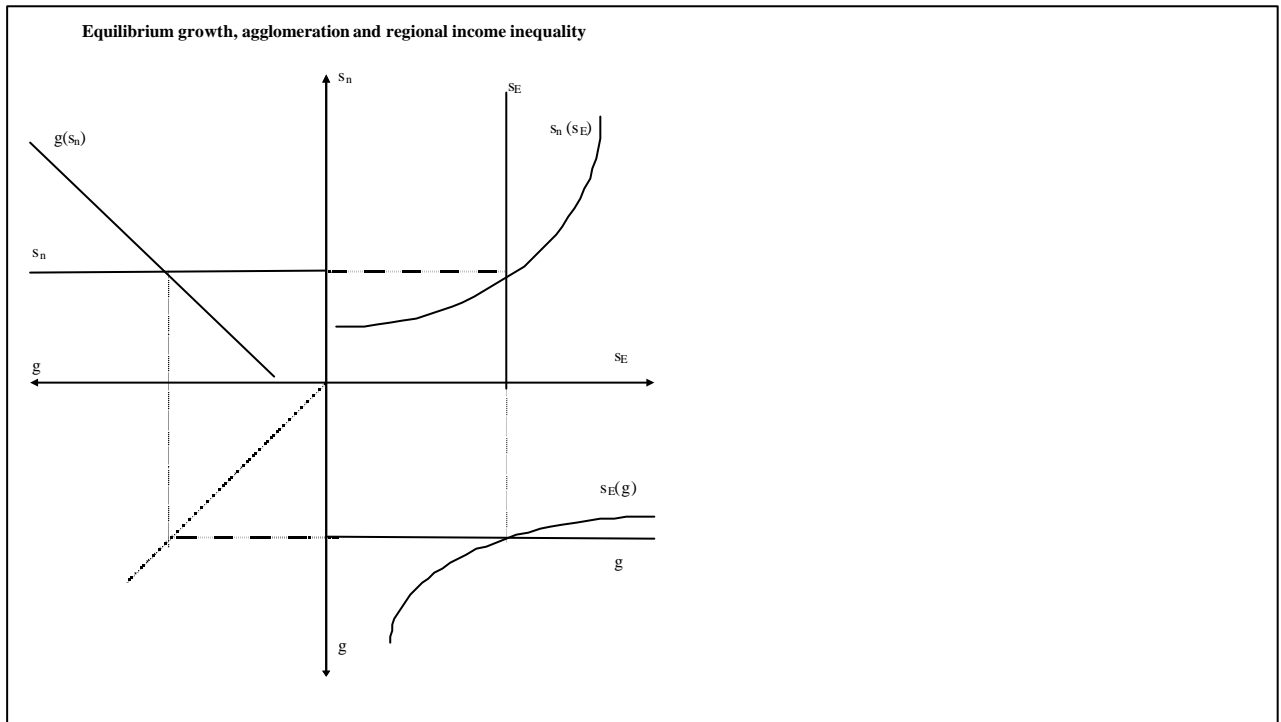

Equilibrium growth, agglomeration and regional income inequality

Figure 5: Spatial equity and efficiency

It is also easy to analyze the impact of lower trade costs on goods (higher f). For a given income disparity, it increases spatial inequality so that the schedule $s_n(s_E)$ shifts up in the NE quadrant. This in turn increases the growth rate which leads to lower income inequality, an effect that mitigates the initial impact on spatial inequality. Overall even though spatial inequality has increased, the growth rate has increased and nominal income disparities have decreased[11]. Martin (1999) also shows that lowering transport costs inside the poor region will have exactly the opposite effect as it leads firms to relocate into that region.

It is also interesting to analyze the effects of an increase in l that is less localized technology spillovers. This can be interpreted as lowering transaction costs between regions on ideas and information. Public policies that improve infrastructure on telecommunication, the

---

[11] These results of course depend on the assumption that agglomeration of economic activities decreases the cost of innovation. If congestion costs exist when the agglomeration becomes too large, lowering trade costs between regions may have very different effects. Baldwin et al. (2003) show in this case that lower trade costs may lead to an equilibrium with high spatial inequality, high income inequality and low growth.

internet or education may be interpreted as affecting l. This shifts the $g(s_n)$ to the left in the NW quadrant so that growth increases for a given geography of production. This lowers income disparities between the two regions as monopolistic profits are eroded by the entry of new firms. This in turn brings a decrease in spatial inequality on the geography of production as s_n decreases. More generally, an exogenous increase in growth will lead to less spatial agglomeration and less regional income inequality. This is important because it implies that, even in the presence of localized technology spillovers, the sign of the correlation between growth and agglomeration depends on the nature of the causal relationship.

4.2.2   Welfare implications

The structure of the model is simple enough so that it is fairly easy, at least compared to the other models, to present some welfare implications. One question we can ask is whether the concentration of economic activities, generated by market forces, is too small or too important from a welfare point of view (see Baldwin et al. 2003, chapters 10 and 11 for a more detailed analysis). Two distortions, which are directly linked to economic geography, exist here. First, when investors choose their location they do not take into account the impact of their decision on the cost of innovation in the north where the innovation sector is located. Localized positive spillovers are not internalized in the location decision and from that point of view the "market" economic geography will display too little spatial concentration. Second the location decision also has an impact on the welfare of immobile consumers which is not internalized by investors. This happens for two reasons. On the one hand an increase in spatial concentration affects negatively the cost and therefore the value of existing capital so that the wealth of capital owners in both regions decreases. This affects more the north than the south. On the other hand, when spatial concentration in the north increases, consumers in the north gain because of the lower transport costs they incur. Symmetrically, consumers in the south loose. V and V^*, the indirect individual utilities of north and south respectively, as a function of the spatial concentration s_n and of the growth rate g are given by:

$$V = C + \frac{1}{r}\ln\left(1 + \frac{? \, s_K}{L s_n}\right) + \frac{a}{r(s-1)}\ln\left[s_n + f(1-s_n)\right] + \frac{a \, g}{r^2(s-1)} \qquad (19)$$

$$V^* = C + \frac{1}{r}\ln\left[1 + \frac{? \, (1-s_K)}{L s_n}\right] + \frac{a}{r(s-1)}\ln\left[1 - s_n + f \, s_n\right] + \frac{a \, g}{r^2(s-1)}$$

where C is a constant. We can analyze how a change in the spatial concentration s_n affects welfare in both regions:

$$\frac{\partial V}{\partial s_n} = \frac{2L a^2(1-l)}{r^2 s(s-1)} - \frac{s_K}{L s_n^{\,2} + r s_K s_n} + \frac{a}{r(s-1)}\frac{1-f}{s_n + f(1-s_n)} \qquad (20)$$

$$\frac{\partial V^*}{\partial s_n} = \frac{2L a^2(1-l)}{r^2 s(s-1)} - \frac{1-s_K}{L s_n^{\,2} + r(1-s_K)s_n} - \frac{a}{r(s-1)}\frac{1-f}{1-s_n + f s_n}$$

There are three welfare effects of a change in spatial concentration. The first term is identically positive in both regions: an increase in spatial concentration increases growth because,

through localized spillovers, it decreases the cost of innovation. The second term is negative in both regions: the decrease in the cost of innovation also diminishes the value of existing firms and therefore diminishes the wealth of capital owners. Because the north owns more capital than the south, this negative effect is larger in the north than in the south. Finally, the last term represents the welfare impact of higher concentration on transaction costs. This welfare effect is positive in the north and negative in the south.

To analyze whether the market geography displays too much or too little concentration in the north implies to evaluate these two equations at the market equilibrium. It can be checked that as long as l is sufficiently small (technological spillovers are sufficiently localized), the effect of an increase in spatial concentration is always positive on the north. It is interesting that the north will gain less by an increase in geographical concentration if it owns a larger share of the capital. Another way to say this is that capital owners may loose from geographical concentration in the north. Geographical concentration in the north may improve welfare in the south. This is in stark contrast with standard economic geography models without growth where the southerners always loose following an increase in concentration in the north. Here the positive effect on growth may more than compensate the negative impact of concentration on transaction costs and on wealth. This will be so if l is sufficiently small (technological spillovers are sufficiently localized), and if transaction costs are low enough.

4.3. The case without capital mobility: the possibility of a growth take-off and agglomeration

As in the case of globalised spillovers, allowing perfect capital mobility stabilises the localised spillovers model by eliminating demand-linked circular causality. We turn now to the opposite assumption – capital immobility. As we shall see, this opens the door to some spectacular interactions between growth and geography.

Here we follow the analysis of Baldwin, Martin and Ottaviano (2001) and simply sketch the nature of the solution. The model is identical to the one described in the previous section except for the introduction of localized spillovers as described in the previous section. This has several consequences: the geography of production has now an impact on the cost of innovation so that as in the previous section, the global growth rate is affected by geography. The value of capital, which can differ in the two regions as capital mobility is absent, is itself affected by geography because the innovation sector is perfectly competitive. Hence, the marginal cost of capital and innovation is equal to its value. In turn, this affects wealth and expenditures in the two regions so that profits will depend on geography in this way too. This implies that the two relations between the share of capital in the north (s_K) and the share of expenditures in the north (s_E) are going to be much more complex than in the case without localized spillovers.

4.3.1 The long-run equilibria and their stability

The optimal savings/expenditure function derived from intertemporal utility maximisation, which we interpreted as a permanent income relation in the previous section (equation 8) becomes:

$$s_E = 1/2 + \frac{rl(2s_K - 1)}{2\{2LAA^* + r[A(1-s_K) + A^* s_K]\}} \tag{21}$$

where A is given in (15) and A* is the symmetric. The permanent income relation is such that s_E is always increasing in s_K: an increase in the northern share of capital increases the northern share of expenditures. When we consider interior steady states where both regions are investing (innovating), so that $q = 1$ and $q^* = 1$, the second relation between s_E and s_K, which we called the optimal investment one, becomes, in the presence of localized spillovers:

$$s_E = 1/2 + \frac{(2s_K - 1)(\boldsymbol{l} + \boldsymbol{l}\boldsymbol{f}^2 - 2\boldsymbol{f})}{2(1 - \boldsymbol{f}^2)\left[A(1 - s_K) + A^* s_K\right]}$$

(22)

Note of course that $s_E = s_K = \frac{1}{2}$, the symmetric equilibrium is a solution to the two equilibrium relations (23) and (24). Two other solutions to this system may exist which are given by:

$$s_K = 1/2 \pm \frac{1}{2}\sqrt{\left(\frac{1 + \boldsymbol{l}}{1 - \boldsymbol{l}}\right)\left(\frac{1 + \boldsymbol{l}\Lambda}{1 - \boldsymbol{l}\Lambda}\right)} \quad ; \quad \Lambda \equiv \left[1 - \frac{2\boldsymbol{r}\boldsymbol{f}(1 - \boldsymbol{l}\boldsymbol{f})}{L(\boldsymbol{l} + \boldsymbol{l}\boldsymbol{f}^2 - 2\boldsymbol{f})}\right]^{-1}$$

(23)

Both s_E and s_K converge to $\frac{1}{2}$ either as $\boldsymbol{l}$ approaches 1 or as $\boldsymbol{f}$ approaches the value:

$$\boldsymbol{f}^{cat} = \frac{L(1 + \boldsymbol{l}) + \boldsymbol{r} - \sqrt{(1 - \boldsymbol{l}^2)\left[L(1 + \boldsymbol{l}) + \boldsymbol{r}\right]^2 + \boldsymbol{l}^2 \boldsymbol{r}^2}}{\boldsymbol{l}\left[L(1 + \boldsymbol{l}) + 2\boldsymbol{r}\right]}$$

(24)

from above. For levels of $\boldsymbol{f}$ below $\boldsymbol{f}^{cat}$, these two solutions are imaginary. In addition, for levels of $\boldsymbol{f}$ above another critical value:

$$\boldsymbol{f}^{CP'} = \frac{2L + \boldsymbol{r} - \sqrt{(2L + \boldsymbol{r})^2 - 4\boldsymbol{l}^2 L(L + \boldsymbol{r})}}{2\boldsymbol{l}(L + \boldsymbol{r})}$$

(25)

one of the solutions is negative and the other one is above unity. Since both violate boundary conditions for s_K, the corresponding steady state outcomes are the corner solutions $s_K = 0$ and $s_K = 1$. Note that for $\boldsymbol{l} = 1$, $\boldsymbol{f}^{cat} = \boldsymbol{f}^{CP'} = \boldsymbol{f}^{CP}$ as defined in the previous section.

As in the case without localized spillovers, we can study the stability of the Core-Periphery equilibrium by analyzing the value of q^* at $s_K = 1$:

$$q^*\Big|_{s_K=1} = \frac{\boldsymbol{l}\left[L(1 + \boldsymbol{f}^2) + \boldsymbol{r}\boldsymbol{f}^2\right]}{\boldsymbol{f}(2L + \boldsymbol{r})}$$

(26)

When $q^* < 1$, we know that then the Core-Periphery equilibrium is stable as the south has no incentive to innovate any more. It is easy to check that $q^* < 1$, when $\boldsymbol{f} > \boldsymbol{f}^{CP'}$.

The stability of the symmetric equilibrium can be studied following the same method as in the case without localized spillovers. We turn to signing $\partial q / \partial s_K$ evaluated at the symmetric equilibrium. Differentiating the definition of q with respect to s_K, we have:

$$\left(\frac{\partial \bar{q}/\bar{q}}{\partial s_K}\right)_{\big|_{s_K=1/2}} = 2\left(\frac{1 - \boldsymbol{f}}{1 + \boldsymbol{f}}\right)\left(\frac{ds_E}{ds_K}\right)_{\big|_{s_K=1/2}} + \frac{4}{1 + \boldsymbol{l}}\frac{1 + \boldsymbol{f}^2}{(1 + \boldsymbol{f})^2}\left[1 - \boldsymbol{l} - \frac{(1 - \boldsymbol{f})^2}{1 + \boldsymbol{f}^2}\right]$$

(27)

Using (21) to find $\partial s_E / \partial s_K = 2\boldsymbol{l}\boldsymbol{r}/\left[(1 + \boldsymbol{l})(L(1 + \boldsymbol{l}) + \boldsymbol{r})\right]$, when evaluated at $s_K = \frac{1}{2}$, we see that the system is unstable (the expression in (27) is positive) for sufficiently low trade costs

(i.e. $f \approx 1$). The two effects discussed in the previous section in the case without localized spillovers are still present. The first positive term is the **demand-linked effect**: an increase in s_K increases north's capital income, expenditure share and local profits so that the value of an innovation (the numerator of Tobin's q) increases. The last negative term is the stabilizing **market crowding effect**. The second (positive) term is new and can be thought of as the **localized spillovers effect**: an increase in s_K implies a lower cost of innovation in the north (the denominator of Tobin's q) and therefore increase the incentive to innovate in the north.

4.3.2 Possibility of catastrophic agglomeration

It is possible to show that $f^{cat} < f^{CP'} < f^{CP}$. Hence, localized spillovers make the *catastrophic* agglomeration possible for higher transaction costs. The critical level at which the expression in (27) becomes positive is f^{cat}. The appendix uses standard stability tests involving eigenvalues and derives the same result. Figure 6~~Figure 6~~ summarises the model's stability properties in a diagram with f and s_K on the axes. It shows that up to f^{cat}, only the symmetric equilibrium exists and is stable. Between f^{cat} and $f^{CP'}$, the symmetric steady state looses its stability to the two neighbouring interior steady states, which are thus saddle points by continuity. This is called a "supercritical pitchfork bifurcation". After f^{CP}, only the Core-Periphery equilibria are stable. Note that these can be attained only asymptotically because, due to the absence of capital depreciation, the south share of capital never goes to zero even after it stops investing (i.e. after f^{CP}).
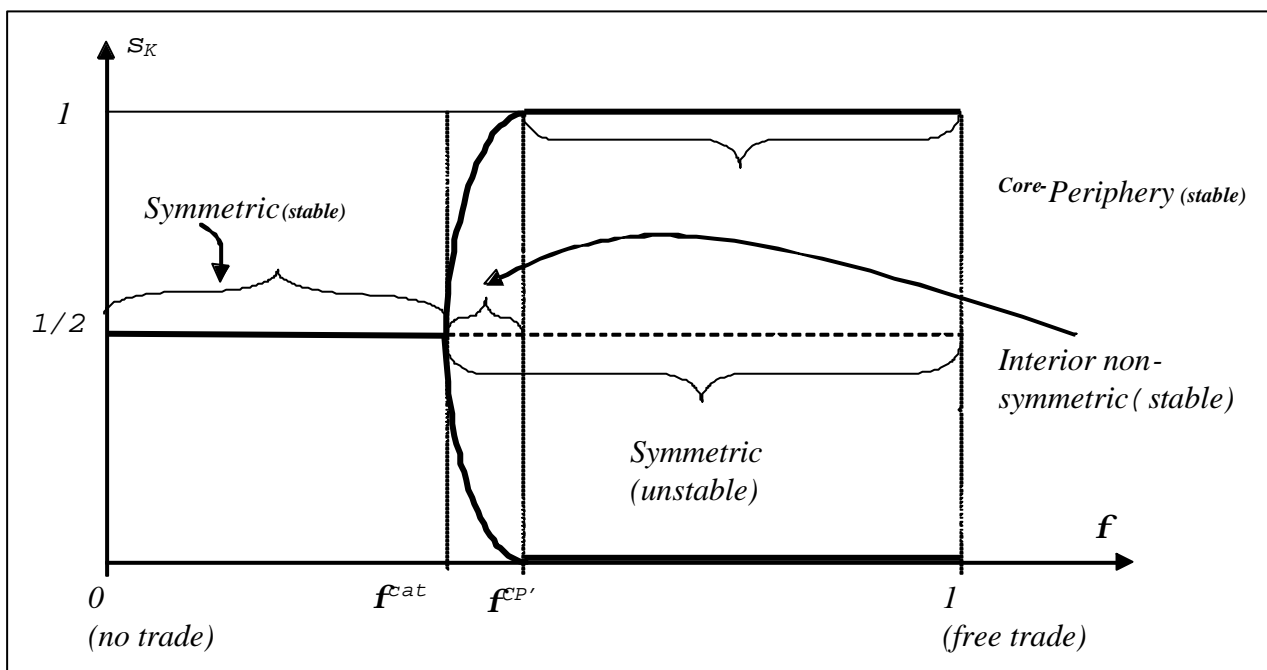

Figure 6: Stability properties of equilibria in the presence of localized spillovers

4.3.3 Geography affects growth

Introducing localized technology spillovers implies that economic geography affects the global growth rate and the model generates endogenous stages of growth. In the version with capital mobility, the result that geography affects growth was already present. However, because

26

of the absence of possible catastrophe, the relation between geography and growth was linear. This is not the case here. There are different stages of growth in the sense that if we think that transaction costs are lowered with time (in line with the "new economic geography" literature), then as economic geography is altered in a non linear way, the growth rate itself changes in a non linear manner. When transaction costs are high so that $f < f^{cat}$, the equilibrium economic geography is such that industry is dispersed between the two regions. This implies that spillovers are minimized and the cost of innovation is maximum. Using the optimal investment condition $q = q^* = 1$, and the fact that $s_K = \frac{1}{2}$, it is easy to find the growth rate (see also equation (18) using $s_K = s_n = \frac{1}{2}$) in that first stage:

$$g = bL(1 + l) - r(1 - b) \tag{28}$$

The growth rate of course increases with l. Asymptotically, when $s_K = 1$, spillovers are maximized so that the cost of innovation is minimized. Again using equation (18) with $s_K = s_n = 1$, the growth rate is in that stage:

$$g = 2bL - r(1 - b) \tag{29}$$

which is, of course, identical to the solution when spillovers are global since in the core-periphery outcome, all innovators are located in the same region and so no learning is diminished by λ.

The growth rate in that final stage is higher than the growth rate in the first stage when transactions costs are high. In the former stage, innovation has stopped in the south which then is entirely specialized in the traditional good. In the intermediate stage, which we call the take-off stage, i.e. when transaction costs are such that $f^{cat} < f < f^{CP'}$, the growth rate cannot be analytically found. However, it can be characterized as a take-off stage as the pitchfork bifurcation properties of the system entail that the economy leaves a neighbourhood of the symmetric steady state equilibrium to reach a neighbourhood of the asymmetric steady state equilibrium in finite time.

We have seen that a gradual lowering of transaction costs on goods (an increase in f) leads, once the transaction cost passes a certain threshold, to a catastrophic agglomeration characterized by a sudden acceleration of innovation in one region (take-off) mirrored by the sudden halt of innovation in the other region. The north (the take-off region) enters a virtuous circle in which the increase in its share of capital expands its relative market size and reduces its relative cost of innovation which in turn induces further innovation and investment. In contrast, the south enters a vicious circle in which lower wealth leads to lower market size and lower profits for local firms. It also leads to an increase in the cost of innovation so that the incentive to innovate diminishes. Hence, growth affects geography which itself affects growth and agglomeration is driven by the appearance of growth poles and sinks.

4.3.4   Can the Periphery Gain from Agglomeration?

In most geography models, agglomeration is a win-lose bargain. Residents of the region that gains the industry typically enjoys an increase in welfare while those left in the periphery see their real incomes fall. Allowing for endogenous growth opens the door to an important caveat to this pessimistic scenario.

Growth Compensation for the Periphery

The continual lowering of trade costs produces uneven spatial development – real per-capita income rises in the core region (since it saves the trade costs on all M-varieties) and falls in the peripheral one (since it pays the trade costs on all M-varieties). However, the emergence of regional imbalances is accompanied by faster growth in all regions (growth take-off). Of course, this is good also for the periphery and creates a tension between the static loss due to relocation and the dynamic gain due to faster growth. Thus, while the core is unambiguously better off, the take-off has ambiguous effects of peripheral welfare.

Intuition is served by , which plots the long run levels of welfare in the two regions as functions of trade freeness. In particular, it depicts a scenario in which lower trade costs drive all industry towards the north.

When trade is sufficiently closed, freer trade raises welfare in both regions because it lowers the price of imported manufactured goods. As trade freeness rises above the break point, north and south welfare levels diverge. The north benefits from agglomeration and faster growth. The south benefits only from the latter, while it is harmed by the former. This explains why the south's post-take-off welfare is always below the north's. Once full agglomeration has been reached (i.e., freeness has risen above the sustain point), the north's welfare is constant.

The behaviour of south's welfare is more complex. If the expenditures share of manufacturing goods μ is low enough, the increase in the growth rate has only a mild impact on welfare and the static loss dominates. In this case, the South loses from the take-off. This case is shown by the solid line (the lowest one in the diagram). On the contrary, if the share μ is sufficiently large, the dynamic gain dominates and the take-off benefits both regions, as shown by the dotted line. Finally, for intermediate values of μ, the south initially loses but eventually attains a welfare level that exceeds its pre-take-off situation. This is illustrated by the dashed curve.
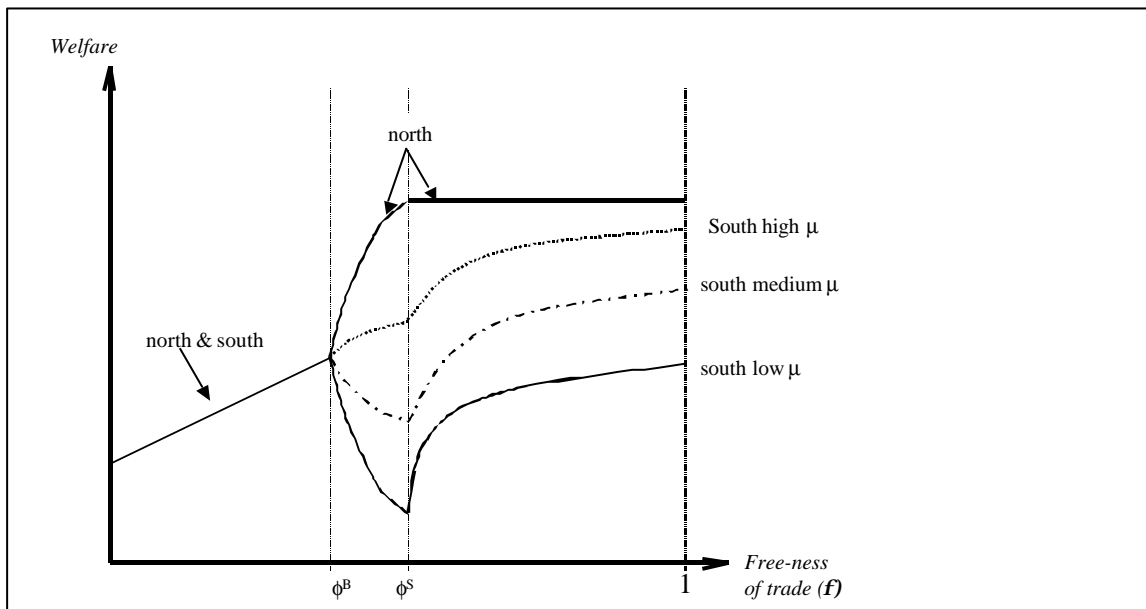

Figure 7: Agglomeration, growth and welfare

Importantly, after the take-off lowering transaction costs always improves welfare in the south because it decreases the price of goods imported from the north. Thus, even though the south may have been made worse off by agglomeration in the north, resisting further reductions in transaction costs is not welfare improving.

4.4. *The geography of goods and ideas: Stabilising and destabilising integration*

The main focus in the "new economic geography" literature has been on the consequence of falling transaction costs on trade in goods. We have shown that in a dynamic model with endogenous growth and localized spillovers, lower transaction costs on goods have an effect on industry location but also on the growth rate. These effects can be "catastrophic" or not, depending on the mobility of capital.

Economic integration, however, is a multi-faceted phenomenon. Up to this point, we have look at two types of closer integration – lowering the cost of trade in goods, and making capital more mobility. There is another aspect of integration, however, which can also be important, namely integration that results in lower barriers to the spatial diffusion of leaning knowledge spillovers. What might be called the cost of trading ideas. In the model introduced above (localised spillovers and immobile capital) we can study the impact of making trade in ideas freer by changing the 'learning spillover' parameter λ[12].

---

[12] Keller (2002) shows that technology spillovers have indeed become considerably more global between 1970 and 1995.
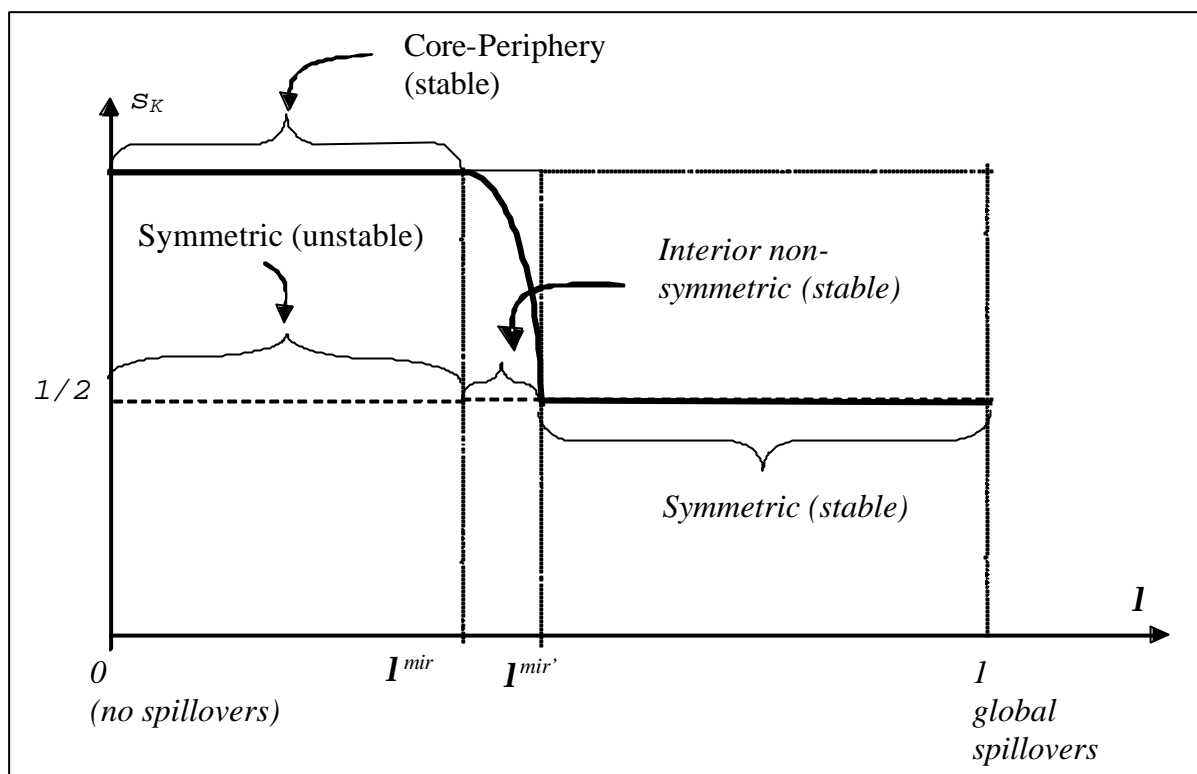
Figure 8:

Falling transaction costs on ideas: Stability properties of equilibria in the presence of localized spillovers

4.4.1   Globalisation and the newly industrialised countries

One exercise that illustrates the interactions focuses on the fact that both f^{cat} and $f^{CP'}$ are increasing in l. The intuition is that as spillovers are becoming more global, an increase in the northern share of capital does not decrease much the relative cost of innovation in the north (a destabilizing effect), so that the capital income effect (the stabilizing effect based on lower transaction costs on goods) must be stronger. One important implication is that from a situation with full agglomeration in the north ($s_K = 1$) and fixed transaction costs on goods, a gradual increase in l (more globalized spillovers due for example to falling telecommunication costs) initially has no impact on southern industry. However, because the cost of innovation in the south decreases with l, Tobin's q in the south increases with l. At some point, when l is high enough, q^* becomes more than 1, and the south begins to innovate. The value of this threshold level which we call l^{mir} (for "miracle") is:

$$?^{mir} = \frac{f(2L + ?)}{L(1 + f^2) + ?f^2}$$

(30)

As in the case of falling transaction costs on goods, there is a second critical value where the symmetric equilibrium becomes stable. This value, denoted as $l^{mir'}$ is the level of l such that $\P q / \P s_K$ evaluated at the symmetric equilibrium becomes negative. As with the north take-off, the

"miracle" in the south would appear as a virtuous circle: as it starts investing, its wealth and permanent income rise so that market size in the south and profits made by local firms increase. In turn, as the number of innovations made in the south increases, the cost of future innovations decreases. This "miracle" implies a jump in the investment rate, as Tobin's q in the south is more than 1, and rapid industrialization. Also incomes between the south and the north converge. 8 describes the effect of an increase in l on the model's stability properties in a diagram with l and s_K on the axes.
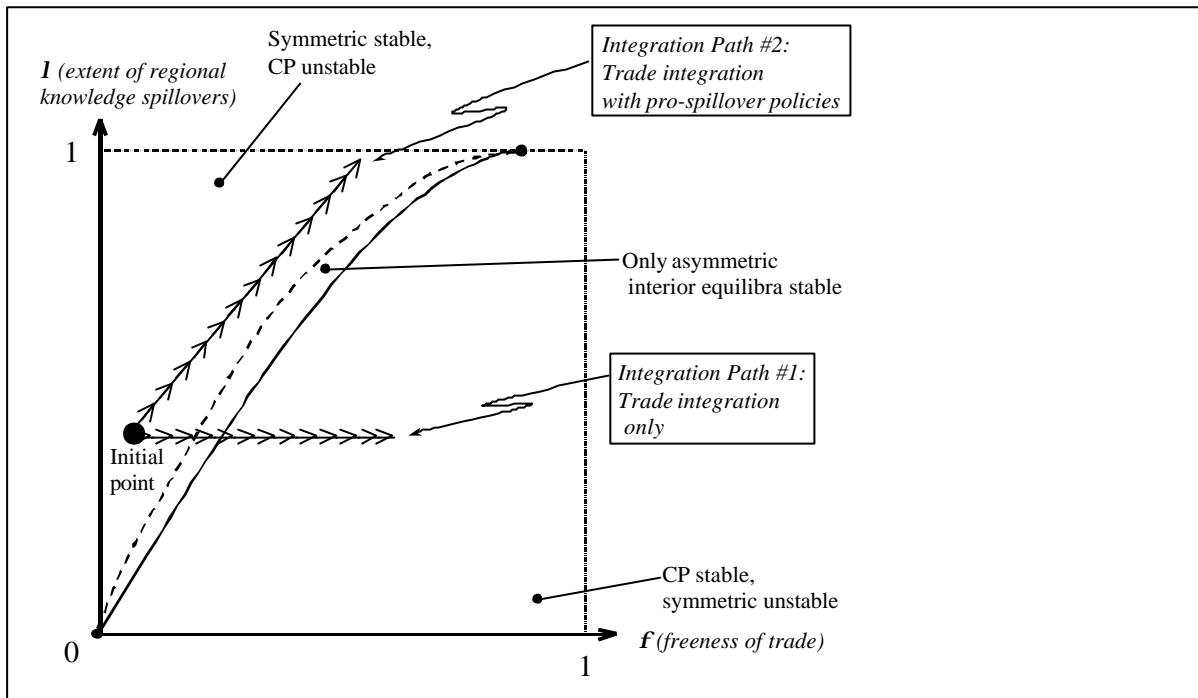

Figure 9: Stability Map for LS Model: Stabilizing and Destabilizing Integration

4.4.2   The learning-linked circular causality

Another way to characterise the essential interplay between the cost of trading goods and ideas is to focus on the symmetric outcome rather than the fully-agglomerated outcome.

It is useful to point out that the localisation of learning in the I-sector creates its own distinct agglomeration force. This new force, which is very much akin to a cost-linkage that operates in the I-sector rather than the I-sector, can be called learning-link circular causality. That is, if a region gets a slightly large amount of knowledge, it becomes a more attractive (cheaper) place to produce more knowledge, all else equal. Since a faster rate of knowledge creation sustains and deepens the region's advantage, an initial bit of 'knowledge shifting' leads to 'knowledge-creation shifting' which in turns leads to knowledge shifting.

Given this logic, it should be clear that making it easier to trade ideas (i.e. raising λ) tends to stabilise the symmetric equilibrium. We see therefore that there can be a tension between the de-stabilising tendency that arises when goods become cheaper to trade and the stabilising tendency that arises when ideas become easier to trade.

To investigate a scenario in which the cost of sharing ideas λ changes together with the cost of trading goods ϕ, 9depicts what Baldwin and Forslid (2000) call a "stability map". This shows how the model's stability properties vary with λ and ϕ. The diagram plots the break and sustain points against the various possible values of λ and ϕ. The dashed curve is the break point and the solid curve is the sustain point. The curves partition the map into three regions. When trade is not very free, and/or knowledge spillovers are very free, then the symmetric outcome is stable and the CP outcomes are unstable. This is the northwest region in the diagram. When trade is quite free and/or knowledge flows are very restricted, only the CP outcomes are stable. This is the southeast region of the map. For a narrow range of ϕ's, two asymmetric interior equilibria are the only stable equilibria and this is shown as the area between the two curves.

The results also point out to a stark difference between lowering transaction costs on goods and lowering transaction costs on ideas. Lower transaction costs on goods may foster divergence in incomes if it triggers an agglomeration process. However, lowering transaction costs on ideas has the opposite effect as it can make the core-periphery equilibrium unstable and trigger a sudden industrialization in the south which leads to convergence. In our model, the distinction between transaction costs on goods and transaction costs on ideas is an easy one. However, in reality trading goods often implies exchanging ideas in the process so that the processes that govern the evolution of the two types of transaction costs are certainly intertwined.

5. OTHER CONTRIBUTIONS

An early attempt to link growth and geography models was Walz (1996) who introduces endogenously expanding product variety in a model with vertical linkages and migration. His assumption of costless migration leads to a bang-bang migration behaviour. Walz (1997) extends the model to a three-region setting.

Black and Henderson (1999) model the relation between urbanization and growth: there are localized knowledge spillovers so that urbanization affects the endogenous growth of the economy. Growth itself affects the size of cities. However, the assumption of a migration process that is determined by a city developer seems rather restrictive.

The models of growth and geography find their antecedents in models of endogenous growth and trade in particular Grossman and Helpman (1991). Some of the results on geography are already present as these authors show that with free trade and when knowledge spillovers are localised, the increasing returns activity concentrates in one location. Models of endogenous growth and trade do not however all imply that free trade leads to divergence if trade also involves capital goods as shown by Goh and Olivier (2002).

Fujita and Thisse (2002a, 2002b chapter 11) propose a combination of a Krugman type core-periphery model and a Grossman-Helpman growth type model with horizontally differentiated products. As in the previous section and the earlier literature on growth and agglomeration, they use a set-up where the fixed cost of firms is a patent. They analyse the two cases of tradability and non tradability of the patents. The skilled workers who produce these patents are themselves mobile and they show that this is destabilising factor. In the case of tradable patents, when transaction costs are sufficiently low, a Core-Periphery pattern emerges with all the R&D sector as well as most of the manufacturing sector concentrated in one region. In this case, agglomeration takes place because of workers mobility and not because of growth

but growth is influenced positively by spatial concentration because of the presence of localized spillovers. With non-tradability of patents, a third destabilising force is added (the first being the mobility of skilled workers and the second being the localized spillovers) so that the Core-Periphery becomes sustainable at higher trade costs.

Baldwin (1999) presents a neoclassical growth model combined with an economic geography model. That paper shows that growth can affect the location industry since changes in regional capital stocks change the relative size of regional markets and this, via the home market effect, alters that spatial allocation of industry. The key to this is the fact that forces that encourage production in one region also tend to encourage capital accumulation in that region. To put it differently, capital accumulation is another way in which expenditure shifting can be tied to production shifting. Moreover, when this demand-linkage is neutralised by assuming that all capital earnings are repatriated, the linkage is broken.

Figure 10: Real Per Capita Income Changes CP vs CC Models
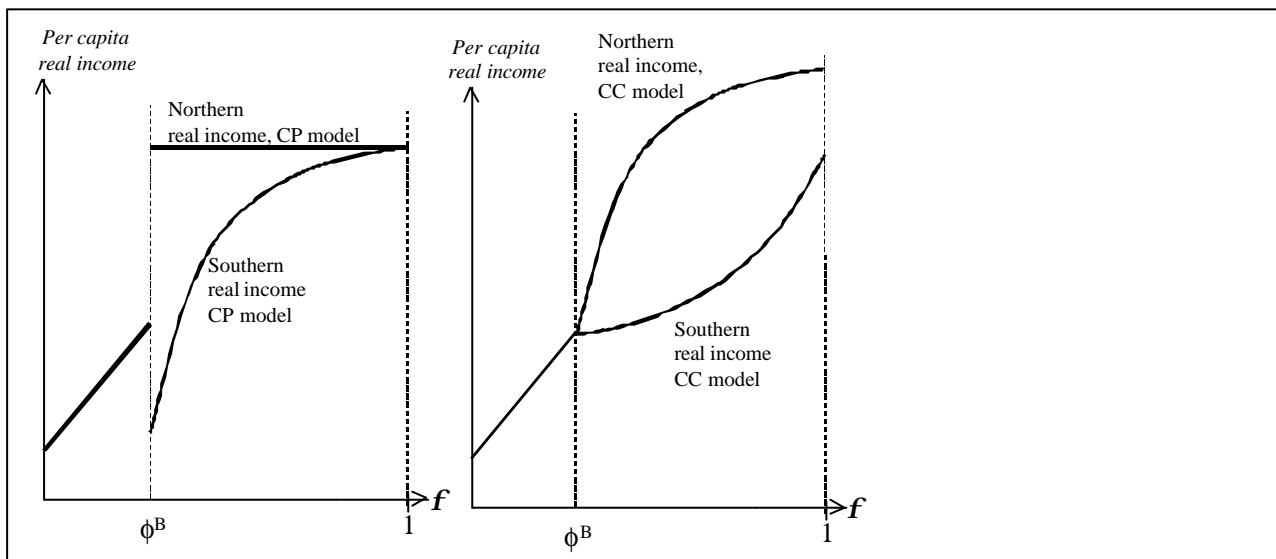
When capital is immobile, the paper illustrates a second novel feature – geography can affect regional growth, at least in the medium-run. In particular, the Perroux (1955) notion of 'growth poles and growth sinks' appears very clearly. Consider, for instance, initially symmetric regions facing trade costs that are high enough to ensure that the symmetric outcome is stable. When trade becomes free enough, symmetry becomes unstable. To be concrete, assume a small shock puts the north a little a bit ahead so the core will eventually end up in the north. The instability arises since the reward to capital rises in the north and falls in the south. This in turn would induce northern residents to raise their investment rate above the rate necessary to sustain the initial capital stock. The consequence might be called agglomeration-induced, investment-led growth. The north's investment rate rises, boosting its capital-labour ratio climbs, and thus its per capita income and output. This expansion of market size further favours investment in the region. In short, the north has become a growth pole. Circular causality has an interesting interpretation in this context. Investment in the growing region is favoured precisely because expenditure in the region is growing. And expenditure is growing due to the high investment rate.

The reverse process operates in the south. The lower rate of return induces southern consumers/savers to stop investing, so depreciation erodes the southern capital stock and southern per capita income and output begin to drop. Given the particular depreciation process assumed, foreign firms shut down one by one. In the simple models we work with here, workers displaced by the downsizing of the south's industrial sector immediately find new jobs in the non-industrial sector. However, if finding a new job or expanding the non-industrial sector took time, the periphery's downward spiral would be associated with above-normal unemployment; the same labour market features would imply 'labour shortages' in the growing region. More colloquially, the declining region would resemble a 'rust belt' and the ascending region would resemble a 'boom belt'.

Allowing growth leads to another feature not commonly found in economic geography models that exclude considerations of growth. In the standard Core-Periphery economic geography model falling trade cost can produce asymmetries in initially symmetric regions. At intermediate trade costs, the two regions' also experience divergence of their real per capita incomes, but eventually, free trade re-equalises incomes. This is illustrated in the left panel of figure 10 with the heavy solid lines (CP model stands for Core-Periphery model) . At the break point, all industry moves north (for convenience, the diagram assumes all H moves immediately) and this raises northern per capita income.

In Baldwin (1999), which assumes capital is immobile, the core-periphery outcomes comes about as a result in a change in the two region's capital labour ratios, with the north's rising and the south's falling. Thus even at free trade, the per capita incomes of the north are permanently higher than those of the south. This is shown in the right panel of figure 10 with CC model standing for constructed capital model.

As an aside, we should also note that Baldwin (1999) also added a new element to the growth literature, which presumes that such neoclassical growth models predict convergence of regional income levels. In Baldwin (1999), however, progressive trade liberalisation between symmetric nations eventually produces the core-periphery outcome. Thus, contrary to the standard assertion in the growth literature, in this neoclassical growth model, economic integration produces divergence in real per-capita income levels.

6. CONCLUDING REMARKS

Introducing growth into economic geography models increases the degree of complexity of models that are already quite complex. Is it worth it? What insights do we gain from the marriage of growth and geography?

Geography and growth models display a number of features that do not appear in more tradition economic geography models such as Krugman (1991). These features are interesting since they help us organise our thinking about important real-world phenomenon and policies. Specifically:

1.   In these models growth affects geography by creating what could be called growth-linked circular causality; forces that foster the location of industry in a region also foster the investment, i.e. the accumulation of human, physical and/or knowledge capital in that region. Since these new factors earn incomes and spend part of these incomes locally, capital accumulation alters relative market size.

2.   The agglomeration process in these models would look like the appearance of growth poles and sinks – firms want to be in the growing region, people want to invest in that region since it is growing and this investment in turn makes the region grow faster. The opposite spiral would appear to be operating in the 'growth sink'.

3.   The simple geography models predict that everyone should be indifferent to agglomeration once trade became really free. In growth and geography models, region capital-labour ratios 'endowments' are permanently altered by agglomeration. Thus regional, real per capital income differentials do not disappear as trade gets perfectly free;

4.   Economic integration is a multi-faceted phenomenon in the real world, yet the standard models focus almost exclusive in the cost of selling goods at a distance. Economic integration has a much richer meaning in geography and growth models. Geography and growth models show that the cost of moving capital across borders (capital mobility) and the cost of moving ideas across borders (learning spillovers) are also important aspects of economic integration. In particular, these other policies can mitigate or extenuate the de-stabilising aspects of freer trade.

5.   Perhaps the most important new feature of geography and growth models is the way in which they allow us to crystallise our thinking about the interplay between the location of economic activity and the growth rate of economic activity.

6.   One aspect of this interplay is important for policy analysis. Taking the standard economic geography models at face value produces higher protectionist policy implications (see Baldwin et al 2003, chapter 12) since agglomeration of industry is always a win-lose situation. In the geography and growth models, the result is not so stark. The continual lowering of trade costs does produce uneven spatial development – real per-capita income rises in the core region and falls in the peripheral one. However, the emergence of regional imbalances is accompanied by faster growth in all regions. Of course, this is good also for the periphery and creates a tension between the static loss due to relocation and the dynamic gain due to faster growth. Thus, while the core is unambiguously better off, the take-off has ambiguous effects of peripheral welfare. A similar application is to realise that regional policies at the national level that seek to avoid geography concentration of industry may cost the country as a whole in growth terms. The introduction of growth in geography models thus points to the existence of a possible spatial equity-efficiency tradeoff.

7.   Perhaps the most sweeping application of these models – and one that is not yet complete – concerns what might be called the grand unified theory of globalisation and geography. Here are the stylised facts of globalisation since the mid 19th century that the unified theory would have to explain. The world has seen two waves of globalisation – one from roughly 1850 to 1914 and one from the 1960s to the present. At a high level of abstraction the key facts are:

        Industrialisation/De-industrialisation. In the first wave, the 'North' (Western Europe and the US) industrialised while South (especially India and China) de-industrialised. In the second wave, the South (East Asia) industrialises while the North de-industrialises.

        Divergence. The first wave sees North and South incomes diverge massively, while the second wave witnesses a convergence, at least between the North and the industrialising South.

        Trade. International trade in goods and factors (labour migration and long-term capital flows) exploded in the first wave. After being shut down by two world wars, a surge of protectionism and the Great Depression, the second wave was marked by a return of trade and

capital flows to levels that have recently topped those seen in Victorian England. Mass international migration, however, remains small by the standards of the first wave.

Growth Take-off. Sometime before the first globalisation wave kicks in, the "Industrial Revolution" triggers modern growth in the North, but the South continues to stagnant in per capita terms. Modern growth, that is a self-sustain growth process whereby output per hour worked rises steadily year-by-year, begins in the UK but spreads to Western Europe and the US around the middle of the 19th century. Of course, this it not independent of the income divergence since big differences in income levels come from sustained differences in growth rates – not from one-time shifts of the location of industry. Moreover, the limited income convergence in the second wave is linked to spectacular growth in the industrialising South and a moderate slowdown in the North.

Urbanisation. While some of the largest cities in the world were in the South prior to the 19th century, the first globalisation wave is accompanied by a rapid and historically unprecedented urbanisation in the North. Northern urbanisation continued during the second wave but cities grew even more rapidly in the South.

The closest the non-growth geography models have come to this is Krugman and Venables (1995), a paper that was known 'on the streets' as history-of-the-world-part-I. This paper sews together the first three of the five facts as follows. In 1750 or so, the world's economic geography was quite homogeneous, i.e. poor and agrarian. With domestic and international trade costs nearly prohibitive, each village essentially had to make all its own goods; this meant manufactured goods were dear and the available range of varieties limited. As trade costs fell, both inside and between nations, specialisation became feasible and this triggered a process of what Myrdal called cumulative causality. Modelling this circular-causality process is the heart of the new economic geography's contribution, so an aside is in order. Migration of firms or workers de-homogenise the world, turning it into economically big and small regions (markets). When industries are imperfectly competitive and trade is costly, Krugman's 'home market effect' favours the location of industry in large regions, but since industries are marked by increasing returns, getting a disproportionate share of industry means a region's labour is disproportionately productive and this in turn results in higher real wages and/or a higher return to capital. The circle is closed by noting that capital and labour are attracted to the region with higher rewards and their migration makes the big region bigger and the small region smaller.

According to Krugman-Venables, advances in transport technology in the early 19th century triggered this de-homogenisation of the world's economic geography, and, as history would have it, the North won at the South's expense. This single event is the root cause of the first three facts: Northern industrialisation and Southern de-industrialisation, the rapid expansion of international trade (England becomes the world's workshop providing cheap and varied manufactured goods in exchange for raw materials and this specialisation both fosters trade and is fostered by it), and income divergence (due to increasing returns in industry and decreasing returns in other sectors, a high share of industry in GDP means high labour productivity and thus high incomes).

One problem with this story is that the magnitudes just don't fit. One-time concentrations of industry just cannot account for the observed income gaps. Here is the argument. Krugman-Venables ignore endogenous technological progress, assuming that physical technology is identical in the North and South. Thus in the Krugman-Venables story, the difference in incomes

between the UK and India must be due to the difference in industry's share in the UK and Indian output mix and the productivity gap between industry and traditional sectors. If the UK's per capita income was 100 in 1850, India's was 23 according to Maddison (1995, Tables C16 & D1), so the income gap to be explained is 77. Moreover, Crafts (1989, p.417) tells us that in 1840, 47% of the UK workforce was in industry, and Bairoch (1982, Table 9) tells us that India was only 4.7% as industrialised as the UK in 1860, so (ignoring the mismatch in dates) we can conclude that the static allocation of industry can only account for the income difference if industrial workers are 171 times – i.e. 17,100% – more productive than workers in the traditional sector.

This just cannot be right. Plainly, the real story must lie elsewhere and growth is the obvious suspect. Indeed, since the headline story in the 19th century was the spread of modern growth, the Krugman-Venables story is a bit like Hamlet with the Prince. Clearly, one has to add endogenous growth to the Krugman-Venables story to account for the facts on income divergence/convergence as well as on growth take-offs.

As shown in this chapter, allowing for endogenous growth, localised spillovers and some capital immobility, we can get the fourth fact of globalisation into a unified framework. The only facts left un-accounted for concerns urbanisation. To get this into the story, one would have to allow internal geography in the regions considered, but once the technical difficulties were mastered, the economics would be straightforward. In the first wave of globalisation, economic activity characterised by localised spillovers is concentrating in the North. It would not therefore be too surprising that urbanisation proceeded faster in the North than in the South during this era. Likewise, in the second wave of globalisation, the industrialisation of the South (emergence of the Asian tigers, etc.) strengthens the forces that foster within South concentration of economic activity, i.e. urbanisation, while the de-industrialisation of the North does the opposite.

REFERENCES

Baldwin, R. and R. Forslid (1997) Trade liberalization and endogenous growth: A q-theory approach, <u>Journal of International Economics</u> 50, 497-517.

Baldwin, R.E. (1999) Agglomeration and endogenous capital, <u>European Economic Review</u>, 43, 253-280.

Baldwin, R. and R. Forslid (2000) The Core-Periphery Model and Endogenous Growth: Stabilising and De-Stabilising Integration, <u>Economica</u> 67, 307-324.

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano and Robert-Nicoud, 2003, <u>Economic Geography and Public Policy</u>, forthcoming, Princeton University Press.

Baldwin, R., P. Martin and G. Ottaviano (2001) Global Income Divergence, Trade and Industrialization: The Geography of Growth Take-Off, <u>Journal of Economic Growth</u> 6, 5-37.

Barro, R. and X. Sala-I-Martin, 1995, <u>Economic Growth</u>, New York Mc Graw Hill.

Basevi, G. and G. Ottaviano (2002) The district goes global: Export vs FDI, <u>Journal of Regional Science</u> 42, 107-126.

Bottazzi, L. (2001) "Globalization and Local Proximity in Innovation: A Dynamic Process" European Economic Review.

Black, D, and V. Henderson (1999) A theory of urban growth, Journal of Political Economy 107, 252-284.

Ciccone, A. (2002) "Agglomeration-Effects in Europe", European Economic Review, Volume 46, Issue 2, February , Pages 213-227

Ciccone, A. and R. Hall (1996) Productivity and the Density of Economic Activity, American Economic Review 87, 54-70

Coe, D. and E. Helpman (1995) International R&D Spillovers, European Economic Review 39, 859-887.

Coe, D., E. Helpman and A. Hoffmaister (1997) North-South R&D Spillovers, The Economic Journal 107,134-149

Duranton G. and D. Puga (2001), "Nursery Cities: Urban Diversity, Process Innovation and the Life Cycle of Products," American Economic Review, vol 5, 1454-1477.

Fujita, M. and J. Thisse, 2002a, "Does Geographical Agglomeration Foster Economic Growth? And Who Gains and Looses from it?" CEPR DP 3135.

----------------------------, 2002b, "Economics of Agglomeration", Oxford University Press.

Grossman, G. and E. Helpman (1991), Innovation and Growth in the World Economy (Cambridge MA: MIT Press).

Goh, A. and J. Olivier, «Learning by doing, trade in capital goods and growth», 2002, Journal of International Economics, 56, 411-444.

Hohenberg P. and L.H. Lees, 1985, *The Making of Urban Europe (1000-1950)*, Cambridge (Mass.), Harvard University Press.

Jacobs, J. (1969) The Economy of Cities (New York: Vintage).

Jaffe A., M. Trajtenberg and R. Henderson (1993) Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, Quarterly Journal of Economics.

Keller W. (2002) Geographic Localization of International Technology Diffusion, American Economic Review 92, 120-142.

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

Krugman P. R. and A. J.Venables (1995) Globalization and the inequality of nations, Quarterly Journal of Economics 60, 857-880.

Lucas, R.E. (1988) On the mechanics of economic development, Journal of Monetary Economics} 22, 3-42.

Manzocchi, S. and G. Ottaviano (2001) Outsiders in economic integration: The case of a transition economy, Economics of Transition 9, 229-249

Martin, P. (1999) "Public Policies, Regional Inequalities and Growth", Journal of Public Economics, (73)1 pp. 85-105

Martin, P. and G. Ottaviano (1999) Growing Locations: Industry Location in a Model of Endogenous Growth, European Economic Review 43, 281-302.

Martin, P. and G. Ottaviano (2001) Growth and Agglomeration, International Economic Review 42, 947-968.

Martin, P. et C. A. Rogers, 1995, "Industrial Location and Public Infrastructure," *Journal of International Economics*, 39, 335-351.

Ottaviano, G.I.P. (1996) The location effects of isolation, Swiss Journal of Statistics and Economics 132, 427-440.

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

Perroux, F. (1995) Note sur la croissance, Economie Appliquée 1-2, 307-320.

Puga D. (1999), The rise and fall of regional inequalities, European Economic Review, vol. 43, pp. 303-334.

Quah, D. (2002), "Spatial Agglomeration Dynamics", CEPR DP 3208.

Romer, P., (1990) Endogenous Technological Change, Journal of Political Economy 98.5, part II, S71-S102.

Urban, D. (2002) "Neoclassical Growth, Manufacturing Agglomeration and Terms of Trade", mimeo LSE.

Waltz, U. (1996) Transport Costs, Intermediate Goods and Localized Growth, Regional Science and Urban Economics 26, 671-795.

Walz, U. (1997) Growth and deeper regional integration in a three-country model, *Review of International Economics* 5, 492-507.

Williamson, J.G., 1988, «Migration and Urbanization », in H. Chenery and T.N. Srinivasan

(eds.), *Handbook of Development Economic*s, volume 1, Amsterdam: North Holland, 425-465.

Yamamoto, K. (2001), Agglomeration and Growth with Innovation in the Intermediate Goods Sector, mimeo, Graduate School of Economics, Kyoto University, forthcoming *Regional Science and Urban Economics.*