CORE DISCUSSION PAPER

2002/44

Urbanization and Growth

Luisito Bertinelli[1] and Duncan Black[2]

August 2002

Abstract

In a simple urban economics framework, we aim at highlighting how the trade-off between optimal and equilibrium city size behaves when introducing dynamic human capital externalities beside the classical congestion externalities. Our purpose is to show that there are dynamic gains from oversized cities. To this end, we assume that productivity depends on human capital, which is solely accumulated in cities, such that urbanization is the engine of growth. In an empirical illustration, we highlight the link between urbanization and human capital accumulation, by focusing on cross-country panel data.

Keywords: urbanization, human capital accumulation, economic development

JEL Classification: O18, R11

[1]FNRS fellow. CORE, Université catholique de Louvain, 34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium.

[0]UCI Department of Economics, 3151 Social Science Plaza, Irvine CA 92697.

Comments by Masahisa Fujita are gratefully acknowledged. We also benefitted from suggestions by participants at the RSAI meeting in Charleston, IEA conference in Lisbon and seminars at CORE (University of Louvain), University of California Irvine, University of Aix-en-Provence, University of Liège. All remaining shortcomings are ours.

1 Introduction

Throughout history, urbanization has been a key element in the process of development (Bairoch, 1988). Arguably, these two processes are inextricably linked - development does not occur without urbanization - although the causal link between these two processes is not clearcut (see for instance Jacobs (1969)). Today, developed countries are far more urbanized than developing ones – urban population represents about 75% of the total population for the former countries and 37% for the latter (UN, *World Urbanization Prospects, 2001*); urban areas generate 85% of GNP in high-income countries compared to only 55% in developing countries (*World Development Report* 2000).

Many authors have attempted to quantify the economic costs and benefits of urbanization. For the developing world in particular, much attention has been paid to the costs of urbanization and perceived inefficiencies in the process. Increased infant mortality rates, substantial pollution, and pervasive traffic congestion are all features of urban life. Rural-urban migration also may entail a search inefficiency, as in the seminal work by Harris and Todaro (1970). The existence of these various congestion externalities and potential inefficiencies in the migration process have led many authors to conclude that the levels of urbanization in many developing countries are too high. Migration may be privately optimal but socially inefficient.

However, some benefits of urban scale must exist in order to motivate rural-urban migration. The urban-rural wage gap in both developed and developing countries is high and persistent. In addition, the nature of cities appears to provide incentives for investments in education by their residents. Returns to education are higher in urban areas, as are literacy rates and average educational attainment levels. Rural-urban migrants tend to be more educated than those who do not migrate (see, for example, Funkhouser (1998)). Children in urban areas are more likely to be enrolled in school.[1] Contributions in endogenous growth theory emphasize the importance of human capital and knowledge accumulation in economic development (Lucas (1988) and Romer (1986)). To the extent that urbanization encourages human capital accumulation, cities become the engines of economic growth. It is therefore of interest to examine the costs and benefits of urbanization within a dynamic framework.

There are two related issues - the overall degree of urbanization and

[1] For a surveys on the nature and implications of urban migration see Yap (1977) and Williamson (1988).

the degree of urban concentration. Many developing countries experience a high degree of urban primacy, with a large share of the urban population concentrated in a single large city, generally a major port, capital or both.[2] Such uneven distribution of urban population my reflect inefficiencies in the spatial structure of the economy. Severe primacy may arise from the concentration of political power and patronage (Ades and Glaeser, 1995), the degree of tariffs and trade (Hanson, 1995), and the spatial allocation of public infrastructure investment (Henderson, 2000). In addition, the absence of efficient land development markets will tend to lead to cities which are larger than their optimal scale (Henderson (1988) and Duranton (2000)). Henderson (2000) shows that the degree of urban concentration has an impact on economic growth, with too much or too little urban concentration having a relatively negative impact on growth. This relationship changes with desired levels of concentration first rising and then falling with the level of development.

Some nations have enacted deliberate policies in attempts to reduce both the degrees of urbanization and urban concentration (Simmons, 1979). Roughly speaking, there have been two sets of strategies for reducing urbanization and urban concentration. The first have involved attempts to directly influence urban populations, either through forced resettlement of urban residents to rural areas,[3] or strict immigration controls on rural-urban migration.[4] The second have involved the establishment of new industrial satellite cities, as a way of reducing the degree of urban concentration rather than decreasing urbanization per se. Examples of countries pursuing these policies are South Korea and Brazil.

These twin issues of urbanization and urban concentration are related but not identical. In this paper we focus on the former and examine a very simple dynamic model of rural-urban migration with human capital investment in an economy with a single urban location. Technological progress in the model is driven by human capital investment, a standard feature in the growth literature (c.f. Galor and Tsiddon, 1997), and is an increasing function of the economy's average level of human capital. Urban migrants invest in human capital, driving economic growth. At low initial levels of technology a development trap can occur, with the levels of urbanization

[2] Taking OECD and subsaharan countries in 1995, we have average primacy rates of 0.25% respectively 0.41%.

[3] Examples include the extreme resettlement programs of Cambodia or the somewhat more benign administrative controls in China during the early 1970s.

[4] Examples of which include attempts in Indonesia and the Philippines to slow migration to Jakarta and Manila, respectively.

and human capital investment insufficiently high to promote growth. In the absence of government intervention, equilibrium levels of urbanization tend to be statically inefficient, a standard result due to the presence of congestion effects. However, policies designed to remove static inefficiencies due to urbanization will tend to slow the process of economic growth and may even leave the economy stuck in a development trap.

Standard theories about the existence of cities stress the tradeoff between the costs and benefits of agglomeration. Benefits of urbanization arise due to the presence of Marshallian externalities (Henderson, 1974) or due to firm level increasing returns in imperfectly competitive markets (Fujita *et al*, 1999). Dispersion forces offsetting benefits of employment and population concentration arise due to such factors as increased commuting expenditures within cities or transportation costs between the urban area and the locations of immobile factors of production. In either case, urban scale is generally self-limiting, with the costs of agglomeration eventually outweighing the benefits.

In the sequel, a stylized urban economics model is depicted. The aim is to take as granted the main results of previous models. Hence, we abstract from any scale benefits per se, and focus on the self-limiting nature of urban congestion effects. The presence of congestion costs, combined with the migration of workers in response to net earning differentials, will always ensure a kind of Malthusian equilibrium in which urban net incomes are always driven down to the level of the prevailing non-urban wage, despite the presence of productivity benefits in the urban location. In the absence of any restrictions on migration, in equilibrium urbanization does not provide any static boost to the economy's net per capita output. However, urban migration will have dynamic benefits due to the investment in human capital by urban migrants. Though with congestion externalities, from a static perspective the equilibrium level of urbanization is generally too high, congestion costs also have the effect of discouraging human capital investment, which has dynamic implications.

The paper is organized as follows. Section 2 provides some empirical hints about the relation of urbanization and human capital accumulation. In section 3 we outline the basic model and solve for the dynamic equilibrium path of the economy. In section 4 we consider the impact of actions by a myopic planner on the economy's growth path. Section 5 concludes.

2 Empirical Foreword

There have been extensive empirical work linking both urbanization as well as human capital to growth (Eaton and Eckstein (1997), Barro (2001), Temple (2001),...). The factor augmenting or productivity enhancing nature of urbanization respectively human capital accumulation are standard explanations of growth in urban and growth economics literature.

Although urbanization and human capital levels are closely interlinked, there has been no formal attempt to study their direct inetraction.

As pointed above, it is a well documented fact that *levels* as well as *returns* to education are higher in urban than in rural areas. The former fact can be related to the activities in cities compared to remote areas. Since Christaller (1933) and subsequent empirical studies, we know that the number of diversified activities increase with city size. Agglomeration of people do allow for higher degrees of segmentation in production activities, and hence higher diversity which is valued by consumers. The broader the array of production, the higher the degree of specialization of each particular activity and hence, the higher the skill requirements.

But this is not the end of the story. This latter fact, i.e. higher returns to education in urbanized areas can directly be related to external effects of human capital. Knowledge diffusion being intrinsically local, urbanized areas constitute locations minimizing distance between people, favoring hence the internalization of knowledge spillovers.

Drawing upon this, we provide empirically related facts by running macro level regressions. In the foregoing, we discuss the results of these estimations. Details on data, estimation techniques and results can be found in the Appendix.

Starting from a panel of approximately 100 countries over 30 years (by 5-year periods observations), we regress human capital accumulation on urbanization, taking into account for endogeneity issues.[5] Estimating a growth-convergence type specification for human capital across countries, and accounting for urbanization and time specific dummies, it results that urbanization is positively related to human capital accumulation even after controlling for the general level of development.

Due to the boundedness of the rate of urbanization between 0 and 1, we may expect non-linear effects of this latter variable on human capital accumulation. Thus, departing from the linear specification, we add a square

[5] Human capital accumulation is proxied as the 5-year increase in the average years of education of the population aged 25 or more.

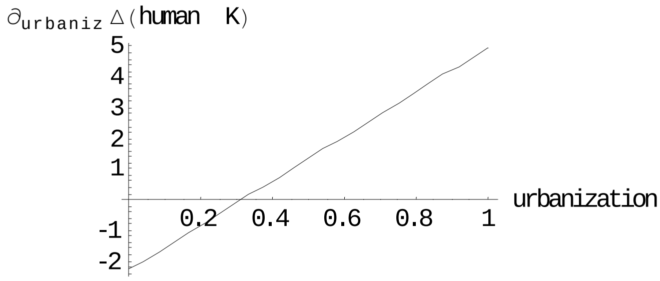

Figure 1: Marginal impact of urbanization on human capital accumulation

urbanization term to our specification, keeping our control variables too. The result points to an increasing *marginal* impact of urbanization on human capital accumulation (see Figure 1). The zero crossing point, bellow which urbanization deters human capital accumulation, is given for levels of urbanization approximately 30%. In 1960, 42 countries were bellow this level of urbanization (39% of our sample), and still 25 countries were concerned in 1990 (23% of our sample).[6]

This result is consistent with urban history. In first stages of development, cities entail industries requiring low skilled worker (think of steel cities or textile cities for instance). Hence, no human capital externalities may arise as people have no or low levels of education, and congestion costs largely dominate (Bairoch, 1985). However, as Duranton and Puga (2001) have shown formally, more service oriented, skill intensive activities in cities appear with development going on, favoring consequently education.

Finally, in order to support the idea that the marginal impact of urbanization on human capital accumulation varies with the absolute level of development and educational attainment we use interaction terms:

$$urbaniz_{it}[\alpha + \lambda_1 hsyr_{it}] \tag{1}$$

$$urbaniz_{it}[\alpha + \lambda_1 hsyr_{it} + \lambda_2 (hsyr_{it})^2] \tag{2}$$

[6] In 1965, there were 38 countries (35%), in 1970, 33 countries (30%), in 1975, 32 countries (29%), in 1980, 30 countries (27%) and in 1985, 28 countries (25%).

Doing so, it appears that the marginal impact of urbanization on human capital accumulation is increasing (specification (1)), with an upper threshold value of 10 years when adding the square term of human capital accumulation (specification (2)). This threshold however exceeds the highest educational attainments of any country thus far (USA 1990: 6.92 years). Thus marginal impact of urbanization is always positively correlated to the level of education, as all countries are in the increasing part of the parabola.

Building on these results, we will consider urbanized areas as favorable places of human capital accumulation in the sequel model.

3 The Model

Consider an economy with N workers, who are ex ante identical. All workers are members of single-worker firms who live for one period, supply labor inelastically, and consume what they produce.[7] In the beginning of the period workers make a joint location and human capital investment decision. In the middle of the period they work, and at the end of the period they consume.

Workers may choose to reside in either of two locations. The two locations are distinct geographic entities which differ in their spatial structures and production technologies. The first location is a rural one, in which workers engage in a traditional method of production. Workers in this location have no incentive to invest in human capital, at least over and above some minimum skill level required for their craft. Formally, workers in the rural areas receive the following earning levels measured in utiles:

$$I_{rt} = G_r A_t \tag{3}$$

where A_t represents the economy's level of technology at time t, an G_r is an exogenous productivity factor, representing, for example, the quality of rural infrastructure.

The second location is the urban one. Workers in the urban location have access to different levels of public infrastructure and production technologies, and benefit from investment in education, which requires a supplementary effort given by P per unit of human capital.

[7] The model could easily be augmented to include physical capital. Here we choose to keep the model as tractable as possible and take for granted forces at the origin of cities and that have already been modelled, for instance in Henderson (1988).

This assumption is directly related to section 2 and merits some discussion. While there is substantial evidence that returns to human capital are higher in urban areas, this does not imply that there are no returns to education in rural areas. This particular assumption is made for simplicity and is not necessary - qualitative results of the model remain unchanged if workers in rural areas can also benefit from human capital investment. However, it is necessary that the marginal productivity of a given human capital investment is higher in the urban area.

An additional feature of the urban location is the existence of congestion externalities. There are numerous potential sources of external diseconomies of urban scale. As local populations grow, the degree of traffic congestion rises. There is evidence that there are negative health effects of urbanization, with increased infant mortality rates in large urban areas. This is true for developing areas, but this has also been true until the 19th century in Europe and the US. To some degree, these negative impacts are unpriced and are thus a negative externality. Individual migrants do not recognize the impact of their location decision on the rest of the population.

Formally, gross earnings received by a worker i who chooses to locate in the urban location are:

$$W_{ut} = G_u A_t h_{it}^{\alpha} \tag{4}$$

where G_u represents the quality of public infrastructure at the urban level and other productivity advantages it may possess, A_t is the current level of technology, and h_{it} is the amount of human capital possessed by individual i, which can be obtained through a per unit level of effor P. We assume decreasing returns to scale in education, $\alpha < 1$.[8]

In addition, the utility levels of workers in the urban location are negatively impacted by the size of the local population, capturing urban external diseconomies. We assume that increasing urban scale imposes a cost on urban residents, which for simplicity, can be measured in units of net earnings. The urban residents earnings net of educational efforts and congestion are then:

$$I_{ut} = W_{ut} - Ph_{it} - b(zN)^{\frac{1}{\delta}} \tag{5}$$

where $0 \leq z \leq 1$ is the proportion of the population that locates in the city, b is a constant, and $\delta > 1$ is a parameter that reflects the inverse intensity of the congestion externalities.

[8]Note that earnings have only private returns in the present framework. When introducing the equation of motion, we will however see that there are also social returns to education, but indirectly through the productivity variable.

3.1 Human Capital Investment and Technological Progress

The human capital investment decision faced by urban workers is:

$$\max_{h_{it}} I_u = G_u A_t h_{it}^{\alpha} - P h_{it} - b(zN)^{\frac{1}{\delta}}$$

providing:

$$h_{it}^{*} = (\alpha G_u P^{-1} A_t)^{\frac{1}{1-\alpha}} \tag{6}$$

Though the choice of h_t will impact on the equilibrium value of urban population, as the next section demonstrates, individual workers do not perceive the impact of their investment and location decisions on city size as they do not perceive their contribution to congestion when migrating to cities.

3.2 Locational Equilibrium

With costless migration, the proportion of the population that chooses to locate in the city, $\hat{z}_t$, adjusts up until the point where:

$$G_u A_t h_{it}^{*\alpha} - P h_{it}^{*} - b(\hat{z}_t N)^{\frac{1}{\delta}} = G_r A_t$$

or until $z_t = 1$, and the economy experience full urbanization. This condition can be rewritten as:

$$\hat{z}_t = \left[\frac{G_u A_t h_{it}^{*\alpha} - P h_{it}^{*} - G_r A_t}{b N^{\frac{1}{\delta}}}\right]^{\delta} \tag{7}$$

Proposition 1 *The city size which maximizes the economy's per capita output net of congestion losses is equal to* $\left(\frac{\delta}{\delta+1}\right)^{\delta} \hat{z}_t N$.

Proof. The proof is straightforward. A planner allocating workers in order to maximize per capita net output $\bar{I}_t$ in the economy faces the following problem:

$$\max_{z_t} \bar{I}_t = z[G_u A_t h_{it}^{*\alpha} - P h_{it}^{*} - b(z_t N)^{\frac{1}{\delta}}] + (1-z) G_r A_t$$

from the first order conditions we have:

$$z_t^{m} = \left(\frac{\delta}{\delta+1}\right)^{\delta} \left[\frac{G_u A_t h_{it}^{*\alpha} - P h_{it}^{*^2} - G_r A_t}{b N^{\frac{1}{\delta}}}\right]^{\delta} \tag{8}$$

■

This is a standard result deriving from the existence of a congestion externality. When choosing to locate in the city, individual workers do not take into account the impact of their location on the costs of living for all other workers. Thus, in the absence of any migration restrictions, cities will tend to be too large, at least from a static perspective. Note that if the size of the city is restricted to z^m, earnings in the city will be larger than that in the rural area, making city dwellers better off.

3.3 The dynamic path of the economy

The maximized level of net earnings for urban workers is given by:

$$I_{ut}^{*} = (1-\alpha)B_1 A_t^{\frac{1}{1-\alpha}} - b(z_t N)^{\frac{1}{\delta}} \tag{9}$$

where $B_1 \equiv (\alpha^\alpha G_u P^{-\alpha})^{\frac{1}{1-\alpha}}$ and z will be determined by the locational equilibrium conditions $\hat{z}_t$ above.

As is common in the literature (e.g. Galor and Tsiddon, 1997), we assume that technological change is driven by human capital accumulation. Specifically, we assume that the equation of motion for the overall level of technology A_t as given by:

$$A_{t+1} = \max[A_t, (z_t h_{it})^\theta] \tag{10}$$

with no depreciation. The reason why we suppose that productivity is linked to average rather than absolute level of human capital is related to the fact that we have supposed no direct external effects to education, thus justifying the present assumption.[9]

The level of technology in the economy is increasing in the average level of human capital investment in the preceding period. Increased urbanization leads to greater average investment in human capital, which propels technological progress. Thus, urbanization is the engine of economic growth in this model. More urban residents investing in more human capital increase future productivity.

To solve for the dynamic evolution of the economy, first substitute $(z_t h_{it})^\theta$ into (6), adjusting time subscripts appropriately, to obtain:

$$h_{it} = \left[\frac{\alpha G_u (z_{t-1} h_{it-1})^\theta}{P}\right]^{\frac{1}{1-\alpha}} \tag{11}$$

[9] For a given N, results remain qualitatively unaltered even if productivity depends on absolute levels of human capital.

Hence urbanization acts through two channels on technological change: directly via z_t in expression (10), and indirectly through the impact of z_{t-1} on h_{it} (expression (11)), this latter variable interveining in (10).

Before beginning a more formal analysis of the dynamic path of the economy, an intuitive description of the process is appropriate. As the level of technology increases, individuals will invest in greater amounts of human capital. In addition, higher productivity encourages more people to migrate to the city, increasing the proportion of that population that invests in human capital. Depending on the values of the parameters in the model, there are variety of paths the economy may take. It will always be the case that for low (but positive) initial technology levels, the economy will remain in a development trap, with no economic growth.[10] It is possible that regardless of the initial level of technology, the economy will never experience economic growth. Or, if the initial level of technology is high enough, the economy may grow for a time and then converge to a steady state level of output, with full or partial urbanization, or may experience unbounded growth with full urbanization.

For the dynamic equilibrium path of technology $\{A_t\}_{t=0}^{\infty}$, three cases may be considered: no urbanization, partial urbanization, and full urbanization. In the trivial case, when $z_t = 0$ and $I_{ut} < I_{rt}$, there is no incentive for urbanization to occur. This will be the case whenever the initial level of technology $A^0 \leq \left(\frac{G_r}{(1-\alpha)B_1}\right)^{\frac{1-\alpha}{\alpha}}$. The economy remains in a development trap, experiencing no growth, with per capita earnings $G_r A_t = G_r A^0$ $\forall t$. For interior rates of urbanization, $0 < z_t < 1$, the dynamic path of technology can be found by combining (6), (7), and (10). Per capita earnings are simply equal to $\bar{I}_t = G_r A_t$. Finally, in the case of full urbanization, when $z = 1$, the evolution of technology is obtained by replacing (6) in (10), z_t being equal to 1. Per capita earnings are given by $\bar{I}_t = (1-\alpha)B_1 A_t^{\frac{1}{1-\alpha}} - bN^{\frac{1}{\delta}}$.

Leaving aside the trivial no urbanization case, the dynamic path of technology in the economy is characterized by the following equation of motion:

$$A_{t+1} = \gamma(A_t) = \begin{cases} \gamma_1(A_t) & \text{if} \quad z_t < 1 \text{ and } \gamma_1(A_t) > 0 \text{ and } (z_t h_{it})^{\theta} > A_t \\ \gamma_2(A_t) & \text{if} \quad z_t = 1 \text{ and } \gamma_2(A_t) > 0 \text{ and } (z_t h_{it})^{\theta} > A_t \\ A_t & \text{otherwise} \end{cases}$$

where

[10] Under the assumption that technology is never lost or forgotten, then the economy simply remains at its initial level of output forever.

$$\gamma_1(A_t) \equiv B_2 B_3[(1-\alpha)B_1 A_t^{\frac{\delta+1}{\delta(1-\alpha)}} - G_r A_t^{\frac{\delta(1-\alpha)+1}{\delta(1-\alpha)}}]^{\delta\theta} \tag{12}$$

$$\gamma_2\left(A_t\right) = B_2 A_t^{\frac{\theta}{1-\alpha}} \tag{13}$$

and $B_2 = \left(\frac{\alpha G_u}{P}\right)^{\frac{\theta}{1-\alpha}}$ and $B_3 = \left(\frac{1}{bN^{\frac{1}{\delta}}}\right)^{\delta\theta}$.

The function $\gamma_1(A_t)$ governs the dynamic evolution of A_t in case of partial urbanization. The evolution is governed by $\gamma_2(A_t)$ when there is full urbanization. It is shown in the appendix that at the point where $z_t = 1$, and $\gamma_1(A_t) = \gamma_2(A_t)$, the slope of $\gamma_2(A_t)$ is strictly less than that of $\gamma_1(A_t)$. The reason is given below.

The path of the economy depends critically on the magnitude and shape of the γ_1 function in (12). The economy remains in a no-growth trivial steady state for all values of A_t such that $\gamma_1(A_t) < A_t$. γ_1 becomes positive for some $A^0 > 0$. Below this point, all levels of technology are trivial no-growth steady-states. As the level of technology increases, $\gamma_1(A_t)$ may intersect the 45^o line, where $\gamma_1(A_t) = A_t$, once or twice. Each intersection of this line corresponds to a potential non-trivial steady state. If there exists a single lower intersection, denoted $\bar{A}^1$, then this is a steady state that is unstable from above. Whether or not the economy experiences unbounded endogenous growth or converges to a steady state will then depend on the concavity or convexity of the function $\gamma_2(A_t)$. If it is concave, then the γ function will have an upper intersection with the 45^o line, corresponding to an upper steady state $\bar{A}^2$ that is stable from below (all points above are also trivial no-growth steady states). Otherwise, the convexity of both $\gamma_1(A_t)$ and $\gamma_2(A_t)$ implies the economy will experience unbounded growth. If γ_1 intersects the 45^o line twice, then there definitely will exist an upper state $\bar{A}^2$ with a corresponding level of per capita earnings $\bar{I}^2$.

Forgetting the trivial no urbanization case, we can summarize in the following proposition the dynamic evolution of technology:

Proposition 2 *Three possible dynamic paths of the economy exist:*
1. The economy always remains at its initial levels of urbanization and aggregate output, for all initial levels of technology.

2. There is a range of initial levels of technology such that the economy experiences growth and converges to a steady state with $\bar{A} > 0$, $\bar{h} > 0$, and $0 < \bar{z} \leq 1$. Outside of this range, the economy remains at its initial levels of technology, urbanization, human capital investment, and aggregate output.

3. If the initial level of technology is sufficiently advanced, the economy experiences unbounded growth with full urbanization in the long run.

Case 1 is represented in figure 2. In this case, parameters are such that γ_1 is concave and $\forall A_t, \gamma(A_t) < A_t$. Even if technology is initially quite high, not enough individuals are induced to invest in a high enough level of human capital in order for there to be technological progress. Regardless of the initial level of technology, the economy never manages to grow and remains at its initial levels of technology, earnings, and urbanization forever. Note that figure 2 is pictured assuming that the function $\gamma_1(A_t)$ is concave. This case can still potentially apply even if it is convex, as long as $\gamma_2(A_t)$ is concave and crosses $\gamma_1(A_t)$ below the 45° line.

Case 2 is represented in figures 3 and 4. There are two possible (non-trivial) steady state levels of technology. The steady state $\bar{A}^1$ is unstable. For low levels of initial technology providing earnings levels below this steady state, workers do not invest in enough human capital and the economy remains in a development trap. For an initial level of technology of precisely $\bar{A}^1$, the economy will remain at the lower, unstable steady state. If technology begins above $\bar{A}^1$, the process of urbanization and growth commences. In Case 2a the economy will grow, with human capital investment and the urbanization rate increasing until they reach steady state levels which provide a per capita earnings of $G_r\bar{A}^2$. Note that it is actually possible to "overshoot" this steady-state. If the level of technology begins above this point, the economy remains at initial levels as technology is not forgotten. In this case, the steady state level of urbanization $\bar{z} < 1$.

In Case 2b, at the upper stable steady state level of technology $\bar{A}^2$, the economy experiences full urbanization. The kink in the function describing the dynamical system at point A^f is at the point where technology is sufficiently advanced that the entire population chooses to invest in human capital and locate in the urban area. Above that point, the speed of growth slows, as only new human capital investment, and not a greater number of workers choosing to become educated, increases the average level of human capital investment and propels technological progress. Per capita earnings in the steady state are equal to $(1-\alpha)B_1(\bar{A}^2)^{\frac{1}{1-\alpha}} - bN^{\frac{1}{\delta}}$.[11]

Case 3 is represented in figure 5. A necessary condition for this case is that $\theta + \alpha \geq 1$. In this case, technology and output grow without bound, at least for high enough initial levels of technology. The economy progresses

[11] The inclusion of an actual agricultural sector in the model would generally preclude the potential for full urbanization in the long run.

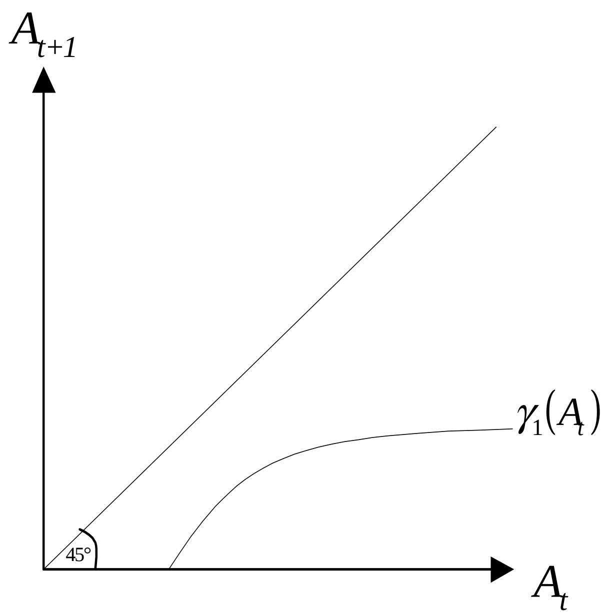

Figure 2: Global development trap (Case 1)

first towards full urbanization, at the point A^f, and then economic growth proceeds, with per capita earnings (and growth rates) rising forever. When $\theta+\alpha = 1$, upon reaching a level of full urbanization, the economy experiences constant growth rates of technology and incomes equal to $B_2 - 1 > 0$. Otherwise, in case 3 the economy experiences explosive growth with ever increasing growth rates, for high enough initial levels of technology.

How can an economy bring itself out of a development trap? The initial level of technology is taken as exogenous here. If the economy experiences a positive technology shock, it is possible that this will lead to urbanization and investment levels sufficient enough to bring the economy out of the trap. In addition, investment in urban infrastructure, by affecting the urban production technology parameter G_u, may provide the necessary incentives for the economy to move from stagnation to growth.

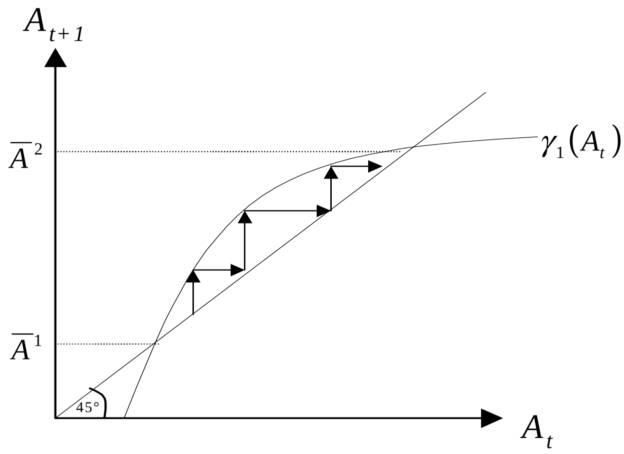

Figure 3: Steady state with partial urbanization (Case 2a)

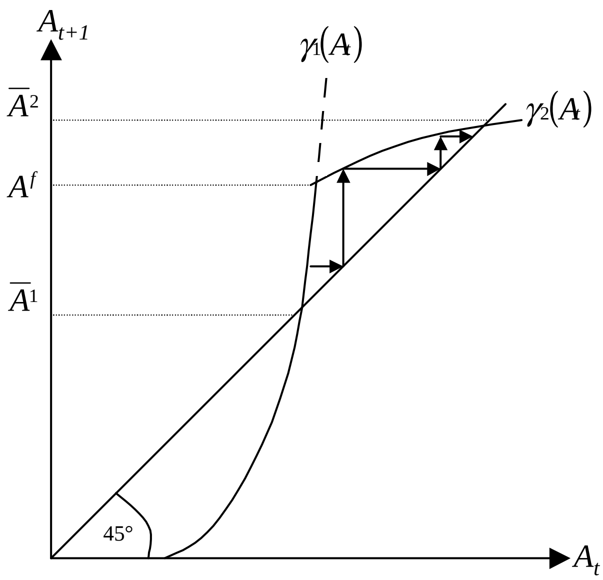

Figure 4: Steady state with full urbanization (Case 2b)

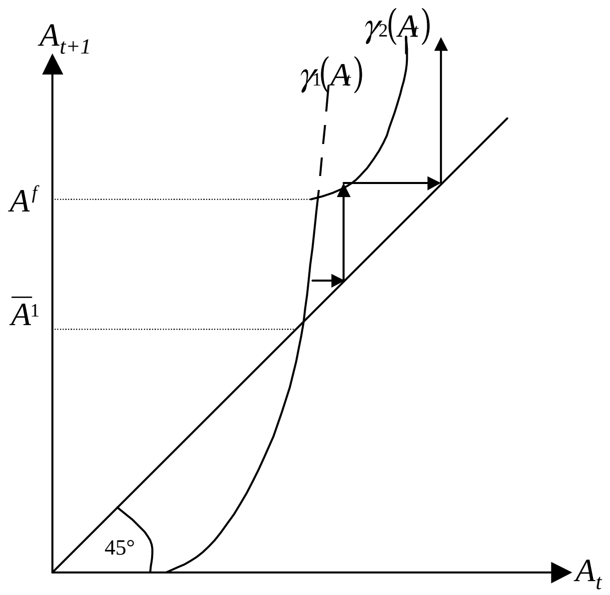

Figure 5: Unbounded growth (Case 3)

4 The Static Planner Case

Proposition 1 demonstrated that equilibrium city sizes are too large, from a static perspective. Workers arriving in the city do not take into account the added burden of congestion that they place on all other workers. Thus, for a planner able to allocate workers in order to maximize per capita earnings in the economy, the solution is simple - restrict urban entry. Of course, the success of any attempt to enforce such a policy will depend on a variety of factors, and may be limited in practice.

However, restricting urban entry also has additional dynamic effects. Limiting the number of residents in the city diminishes the number of workers who invest in human capital, without any impact on the investment levels of workers who are allowed entry. Thus, the aggregate and average level of human capital investment will fall, slowing the rate of technological progress. In addition to potentially slowing the rate of growth, such a policy will expand the range of initial technological levels for which the economy is stuck in a development trap.

Consider the impact of a myopic planner whose objective function is to maximize per capita incomes in each period on the dynamical system. Assume that this policy is announced in the beginning of each period, before each new generation has made its investment and location decisions. A lottery system allocates city dwelling permits to a set number of workers, who then make their human capital investment decisions.

From Proposition 1, we know that the planner will allocate z^m workers in each period, an amount which will depend on the current level of technology and other parameters. The impact of this on the dynamical system is as follows:

$$A_{t+1}^m = \begin{cases} \left(\frac{\delta}{\delta+1}\right)^{\theta\delta} \gamma_1\left(A_t^m\right) & \text{if } z_t^m < 1 \text{ and } \gamma_1\left(A_t^m\right) > 0 \text{ and } \left(z_t^m h_{it}\right)^\theta > A_t \\ \gamma_2\left(A_t\right) & \text{if } z_t^m = 1 \text{ and } \gamma_2\left(A_t^m\right) > 0 \text{ and } \left(z_t^m h_{it}\right)^\theta > A_t \\ A_t & \text{otherwise} \end{cases}$$

The intuitive reason for this is as follows. The myopic planner influences z, the fraction of the population locating in the urban area, and consequently investing in human capital. The myopic solution is a fraction of the optimal solution, however by definition z cannot be greater than one. Thus, when the myopic optimum is a corner solution, with $z^m = 1$, it is equivalent to the equilibrium outcome. Thus, the economy's growth, which is now purely driven by an increase in the quantity of education obtained, and not by any change in the number of those choosing to become educated, matches that of the equilibrium growth path.

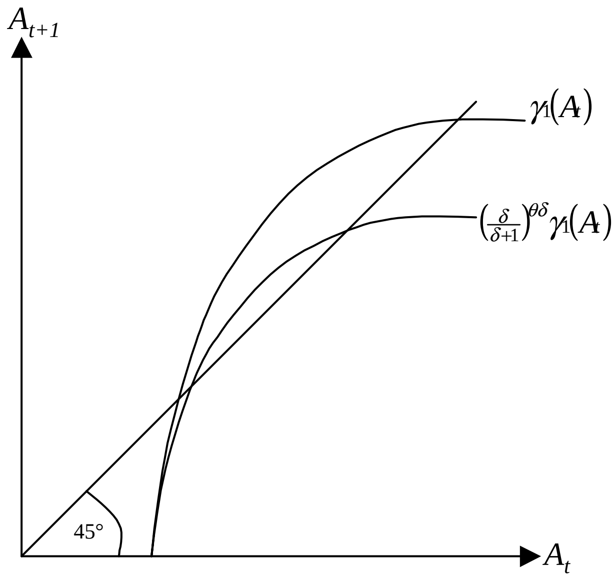

Figure 6: Planner case

Figure 6 gives a graphical representation of the impact of a restriction on urban entry, assuming parameters are such that the economy follows the dynamics of Case 2. Restricting urban entry rotates the equations describing the dynamical system. This has two major impacts. First, the range over which the economy is stuck in a development trap (any technology level below the lower steady state) is increased. Secondly, the upper steady state level of technology is also lower than in the equilibrium case.

Of course, what is important is not technology per se, but earnings. In comparing steady state levels of earnings under the equilibrium versus myopic planner's outcome, it is possible that they may actually be higher in steady state under the myopic planner's outcome. However, it is important to remember that the steady state level of technology determines the level of potential per capita earnings. In early stages of development, at least, the dynamic benefits provided by human capital accumulation may offset the static congestion losses. This is especially true when parameters are such that the economy would otherwise remain in a development trap. In later stages of development, for higher levels of technology, the dynamic benefits are more likely to be offset by static losses. And, at the steady state then for the equilibrium outcome, nothing is lost by restricting urban size to that which is statically optimal.

5 Conclusion

In this paper we presented a dynamic model of rural-urban migration with growth resulting from technological progress due to human capital accumulation. At low levels of technology, a development trap may occur, with the levels of human capital and urbanization being insufficient for growth to occur. In equilibrium, urbanization rates are too high from a static perspective, due to the existence of a congestion externality. However, as urbanization encourages human capital accumulation, there are dynamic benefits of static over-urbanization.

Myopic policies designed to reduce the degree of over-urbanization by limiting urbanization will tend to have an adverse impact on economic growth, lowering an economy's steady-state level of technology and potentially leaving the economy stuck in a development trap. This suggests that policies designed to remedy potential over-urbanization may have adverse dynamic effects. In addition, spatial redistribution, rather than a curtailing of an economy's urban population may remedy the costs of over-urbanization without these negative dynamic effects. However, a full understanding of this requires in-depth knowledge of the costs of infrastructure investments required for urban population decentralization.

References

[1] Ades, A. and E. Glaeser (1995), "Trade and Circuses: Explaining Urban Giants", Quarterly Journal of Economics, 110(1),195-227

[2] Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", Review of Economic Studies, 58, pp. 277-297

[3] Bairoch, P. (1988), "Cities and Economic Development: >From Dawn of History to the Present", (English translation by Christopher Braider), University of Chicago Press

[4] Barro, Robert J. (2001) "Human Capital and Growth", American Economic Review, Papers and Proceedings, 91(2), pp. 12-17

[5] Christaller, W. (1933), "Die zentralen Orte in Süddeutschland", Jena: Gustav Fischer. (Translated (in part), by Charlisle W. Baskin, as Central Places in Southern Germany. Prentice Hall 1966)

[6] Duranton, G. (2000), "Urban Structure, Urbanization and Growth", in J.-M. Huriot and J.-F. Thisse (Eds.), "Economics of Cities: theoretical Perspectives", Cambridge University Press

[7] Eaton, J. and Z. Eckstein (1997), "Cities and Growth: Theory and Evidence from France and Japan", Regional Science and Urban Economics 27, 443-474

[8] Fujita, M., P. Krugman and A. Venables (1999), "The Spatial Economy", Cambridge: MIT Press

[9] Funkhouser, E. (1998), "Changes and the Returns to Education in Costa Rica", Journal of Development Economics 57(2) ,289-317

[10] Galor, O. and D. Tsiddon (1997), "The Distribution of Human Capital and Economic Growth", Journal of Economic Growth 2(1), 93-124

[11] Galor, O. and J. Zeira (1993), "Income Distribution and Macroeconomics", Review of Economic Studies 60, 35-52

[12] Hanson, G. (1995), "The Effects of Offshore Assembly on Industry Location: Evidence from US Border Cities", NBER WP5400

[13] Harris, J.R. and M.P. Todaro (1970), "Migration, Unemployment and Development: a Two-Sector Analysis", American Economic Review 60, 126-142

[14] Henderson, J.V. (1974), "The Sizes and Types of Cities", American Economic Review 64, 640-656

[15] Henderson, J.V. (1988), "Urban Development: Theory, Facts and Illusion", Oxford University Press

[16] Henderson, J.V. (2000), "The Effects of Urban Concentration on Economic Growth", NBER WP7503

[17] Jacobs, J. (1969), "The Economy of Cities", New York: Random House

[18] Lucas, R. (1988), "On the Mechanics of Economic Development", Journal of Monetary Economics 22, 3-42

[19] Romer, P. (1986), "Increasing Returns and Long Run Growth", Journal of Political Economy 94, 1002-1037

[20] Simmons, A.B. (1979), "Slowing Metropolitan City Growth in Asia: Policies, Programs and Results", Population and Development Review 5(1), 87-104

[21] Temple, Jonathan (2001) "Growth Effects of Education and Social Capital in the OECD Countries" OECD Economic Studies, forthcoming

[22] Williamson, J.G. (1988), "Migration and Urbanization", in Chenery, H and T.N. Srinivasan (Eds.) "Handbook of Development Economics, Vol.1", New York, Elsevier

[23] World Development Report (2000), Oxford University Press, World Bank Publication

[24] World Urbanization Prospects (2001), World Urbanization Prospects: the 1999 Revision, United Nations Publications, http://www.un.org/esa/population/publications/wup1999/wup99.htm

[25] Yap, L.Y. (1977), "The Attraction of Cities: A Review of the Migration Literature", Journal of Development Economics 4, 239-264

6 Appendix

6.1 Econometric specification and issue

The estimated functional form is as follows:

$$\begin{aligned}\underbrace{\Delta\left(human\ capital\right)_{it}}_{NT\times 1} &= \alpha+\beta\underbrace{\left(human\ capital\right)_{i,t-1}}_{NT\times 1} \\ &\quad +\delta\underbrace{\left(urbanizat^{\circ}\right)_{i,t-1}}_{NT\times 1}+\underbrace{\left(controls\right)_{i,t-1}}_{NT\times K}\underbrace{\beta}_{K\times 1}+\mu_i+u_{it}\end{aligned} \tag{14}$$

where u_{it} is iid with $E\left(u_{it}\right)=0$ and $e\left(u_{it}u_{js}\right)=\begin{cases}\sigma_t^2 & \text{if } i=j \\ \sigma_i^2 & \text{if } t=s \\ 0 & \text{otherwise}\end{cases}$.

Specification (14) amounts to a conditional convergence pattern. μ_i is a country specific effect, allowing for a different constant per country. However, slope coefficients are identical across countries. In the above specification, μ_i is supposed to be uncorrelated with the error term u_{it}. Nevertheless, OLS estimation will lead to biased estimators as $\Delta\,(human\ capital)_{it}$ is correlated with the country fixed effect, which immediately implies that $(human\ capital)_{i,t-1}$ is a function of μ_i. The same is true for fixed effect estimation, as $(human\ capital)_{i,t-1} - \overline{(human\ capital)}_i$ is by construction correlated with $u_{it} - \bar{u}_i$. Following Arellano and Bond (1991), we thus wipe out fixed effects by first differencing specification (14):

$$\begin{aligned} & \Delta\,(human\ capital)_{it} - \Delta\,(human\ capital)_{i,t-1} \\ = \;& (X_{i,t-1} - X_{i,t-2})\,\varphi + (u_{it} - u_{i,t-1}) \end{aligned} \tag{15}$$

The first RHS expression can be instrumented by $X_{i,t-2}$, which is highly correlated with $(X_{i,t-1} - X_{i,t-2})$ but uncorrelated with $(u_{it} - u_{i,t-1})$ as long as u_{it} are serially uncorrelated. For next period, $X_{i,t-1}$ is a valid instrument for $(X_{i,t} - X_{i,t-1})$, but $X_{i,t-2}$ still too, and so forth for the following periods. Hence typically for dynamic panels, we have different numbers of instruments for different periods, leading to an overidentification issue. The literature suggests to ways out: either leave aside some instruments, but at the same time, lose some information, or estimate the overidentified system via GMM. The latter solution has been adopted in the regressions presented in the results table. The following tests are also provided in our output table: Sargan's test of overidentifying restrictions, asymptotically distributed according as χ^2. First-order and second-order serial correlation: first differencing the model induces MA(1) serial correlation if the time-varying component of the error term in levels is a serially uncorrelated disturbance. This test is distributed as a standard normal variable.

6.2 Data Source Appendix

Data used in this study were assembled from different sources by James Davis and Vernon Henderson. National and urban population data stem respectively from UN World Urbanization Prospects Table A.3 (Values for Czechoslovakia, Yugoslavia and USSR are summed over former constituent republics. Values for West Germany are calculated using the World Bank WDI urban percent of the Penn5.6 population figure.) and National population. UN World Urbanization Prospects Table A.5 (Values for West

Germany, Czechoslovakia, Yugoslavia and USSR are taken from the Penn World tables 5.6 (1995 are missing observations).

The Barro-Lee (Barro, R.J., and J.W.Lee, International Comparisons of Educational Attainment, Journal of Monetary Economics, 32, 363-394, 1993; International Measures of Schooling Years and Schooling Quality, American Economic Review, 86(2), 218-223, 1996; International Measures of Schooling Years and Schooling Quality online data, World Bank Economic Growth Research Group, Washington D.C.: World Bank, 1996.) datasets provided human capital variables. Census and survey figures primarily from UNESCO Statistical Yearbooks and UN Demographic Yearbooks fill two fifths of the observations. The remaining values are estimated using UNESCO school enrollment data and a perpetual inventory method. The data are not adjusted for quality of education or length of school day or year.

Finally, control variables were taken from Penn World Tables Mark 5.6 for *rgdpch*. 1995 values are estimated using World Bank WDI yearly real GDP growth figures on 1990 values. World Bank WDI provided data for *LlifeWBa*. Czechoslovakia, Yugoslavia and USSR are averages over the constituent republics weighted by population share. The values for united Germany are substituted for West Germany.

6.3 Econometric results

The following table provides the econometric results with fixed effect and dynamic panel estimators:

	(1)		(2)		(3)		(4)		(5)	
	GMM	*Fixed effects*	*GMM*	*Fixed effects*	*GMM*	*Fixed effects*	*GMM*	*Fixed effects*	*GMM*	*Fixed effects*
urbanization	**0.60*** (0.37)	**0.91***** (0.304)	**0.73*** (0.39)	**0.97***** (0.379)	**-2.02***** (0.42)	**-1.66**** (0.675)	**1.39***** (0.28)	**0.79**** (0.378)	**1.09***** (0.25)	**0.43** (0.400)
human capital	**-0.22***** (0.04)	**-0.31***** (0.031)	**-0.37***** (0.03)	**-0.41***** (0.035)	**-0.45***** (0.02)	**-0.45***** (0.035)	**-0.64***** (0.06)	**-0.67***** (0.084)	**-0.74***** (0.03)	**-0.71***** (0.084)
***Log* (GDP/capita)**			**0.32***** (0.07)	**0.22***** (0.055)	**0.06** (0.06)	**0.19***** (0.054)	**0.28***** (0.05)	**0.21***** (0.054)	**0.25***** (0.05)	**0.19***** (0.055)
***Log* (life expectancy)**			**-0.90***** (0.27)	**-1.61***** (0.384)	**-0.30** (0.24)	**-0.817**** (0.413)	**-0.45**** (0.22)	**-1.15***** (0.401)	**-0.49***** (0.19)	**-0.96**** (0.406)
(urbanization)^2					**3.41***** (0.44)	**2.77***** (0.596)				
urbanization× human capital							**0.35***** (0.09)	**0.41***** (0.115)	**0.60***** (0.07)	**0.68***** (0.154)
urbanization× (human capital)^2									**-0.03***** (0.004)	**-0.04***** (0.017)
Time dum.	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Sargan test [d.f.; p-value]	31.19 [24;0.2]		40.56 [36;0.3]		55.70 [48;0.2]		59.76 [48;0.1]		67.05 [60;0.3]	
1st order serial corr.	-3.02*** [0.003]		-2.78*** [0.005]		-2.8*** [0.005]		-2.86*** [0.004]		-2.85*** [0.004]	
2nd order serial corr.	-0.782 [0.434]		-0.913 [0.361]		-0.698 [0.485]		-0.860 [0.390]		-0.744 [0.457]	
R^2 within		0.25		0.29		0.32		0.31		0.32
# obs.	400	659	400	615	400	615	400	615	400	615

Fixed effect and GMM results

6.4 Analysis of γ function.

6.4.1 Crossing condition between γ_1 and γ_2

Begin by proving that at the point $z_t = 1$, and $\gamma_1(A_t) = \gamma_2(A_t)$, $\frac{\partial \gamma_1(A_t)}{\partial A_t} < \frac{\partial \gamma_2(A_t)}{\partial A_t}$. Assume not. We can write:

$$\begin{aligned} \frac{\partial \gamma_1(A_t)}{\partial A_t} &= \delta\theta B_2 B_3 [(1-\alpha) B_1 A_t^{\frac{\delta+1}{\delta(1-\alpha)}} - G_r A_t^{\frac{\delta(1-\alpha)+1}{\delta(1-\alpha)}}]^{\delta\theta - 1} \\ &\times [\frac{\delta+1}{\delta} B_1 A_t^{\frac{1/2+\alpha}{1-\alpha}} - \left(\frac{\delta(1-\alpha)+1}{\delta(1-\alpha)}\right) G_r A_t^{\frac{1}{\delta(1-\alpha)}}] \end{aligned} \quad (16)$$

At the intersection of the two functions:

$$\frac{\partial \gamma_2(A_t)}{\partial A_t} = \frac{\theta}{1-\alpha} \gamma_2(A_t) A_t^{-1} = \frac{\theta}{1-\alpha} \gamma_1(A_t) A_t^{-1} \quad (17)$$

Comparing (16) with (17) we find that having the former be less than the latter requires:

$$(\delta+1)B_1A_t^{\frac{1+\delta\alpha}{\delta(1-\alpha)}} - \left(\frac{\delta(1-\alpha)+1}{1-\alpha}\right)G_rA_t^{\frac{1}{\delta(1-\alpha)}} < B_1A_t^{\frac{1+\delta\alpha}{\delta(1-\alpha)}} - \frac{1}{1-\alpha}G_rA_t^{\frac{1}{\delta(1-\alpha)}}$$

which after rearranging requires:

$$B_1A_t^{\frac{1+\delta\alpha}{\delta(1-\alpha)}} < G_rA_t^{\frac{1}{\delta(1-\alpha)}}$$

or:

$$B_1A_t^{\frac{1}{1-\alpha}} < G_rA_t \tag{18}$$

Given that $B_1A_t^{\frac{1}{1-\alpha}} > I_{ut}^* = I_{rt}^* = G_rA_t$, (18) never holds.

6.4.2 Shape of the γ_1 and γ_2 functions

The dynamic evolution of the economy's level of technology depends on the shape of the function $\gamma(A_t)$.

$$A_{t+1} = \gamma(A_t) = \begin{cases} \gamma_1(A_t) & \text{if} \quad z_t < 1 \text{ and } \gamma_1(A_t) > 0 \text{ and } (z_th_{it})^\theta > A_t \\ \gamma_2(A_t) & \text{if} \quad z_t = 1 \text{ and } \gamma_2(A_t) > 0 \text{ and } (z_th_{it})^\theta > A_t \\ A_t & \text{otherwise} \end{cases}$$

where

$$\gamma_1(A_t) \equiv B_2B_3[(1-\alpha)B_1A_t^{\frac{\delta+1}{\delta(1-\alpha)}} - G_rA_t^{\frac{\delta(1-\alpha)+1}{\delta(1-\alpha)}}]^{\delta\theta} \quad \text{and} \quad \gamma_2(A_t) \equiv B_2A_t^{\frac{\theta}{1-\alpha}}$$

$\gamma_1(A_t) \leq 0$ when $A_t \leq \left[\frac{G_r}{(1-\alpha)B_1}\right]^{\frac{1-\alpha}{\alpha}}$ and is strictly positive otherwise. Taking the first derivative, we get expression (16). Over the relevant range where $\gamma_1(A_t)$ and z_t are both non-negative, the first expression in brackets is positive. The second expression in brackets is positive as long as $A_t > \left[\frac{(\delta(1-\alpha)+1)G_r}{(\delta+1)(1-\alpha)B_1}\right]^{\frac{1-\alpha}{\alpha}}$, which is always the case in the relevant range. The function reaches a minimum at $A_t = \left[\frac{(\delta(1-\alpha)+1)G_r}{(\delta+1)(1-\alpha)B_1}\right]^{\frac{1-\alpha}{\alpha}}$ and is monotonically increasing thereafter.

The evolution of the economy will depend critically on the curvature of the γ_1 function. Specifically, the number of times that this function intersects the 45^o line, where $\gamma_1(A_t) = A_t$, will be a crucial factor for the dynamic path of the economy. Recognizing that when $\gamma_1(A_t) = A_t$, $\gamma_1(A_t)^{\frac{1}{\delta\theta}} = A_t^{\frac{1}{\delta\theta}}$, we can simplify the analysis by defining $\hat{A}_t = A_t^{\frac{1}{\delta\theta}}$ and $\hat{\gamma}_1(.) = \gamma_1(.)^{\frac{1}{\delta\theta}}$. We can then analyze the shape of this somewhat simpler function, and the dynamics of the transformed variable $\hat{A}_t$.

We have:

$$\hat{\gamma}_1(\hat{A}_t) = (B_2B_3)^{\frac{1}{\delta\theta}}[(1-\alpha)B_1\hat{A}_t^{\frac{\theta(\delta+1)}{1-\alpha}} - G_r\hat{A}_t^{\frac{\theta[(\delta(1-\alpha)+1)]}{1-\alpha}}]$$

where $B_2 = \left(\frac{\alpha G_u}{P}\right)^{\frac{\theta}{1-\alpha}}$ and $B_3 = \left(\frac{1}{bN^{\frac{1}{\delta}}}\right)^{\delta\theta}$.

$\hat{\gamma}_1(\hat{A}_t) > 0$ as long as $\left[\frac{G_r}{(1-\alpha)B_1}\right]^{\frac{1-\alpha}{\delta\alpha\theta}} \equiv \hat{A}^0$.

Taking the first derivative and simplifying notation by setting $v \equiv \delta + 1$

$$\frac{\partial\hat{\gamma}_1}{\partial\hat{A}_t} = (B_2B_3)^{\frac{1}{\delta\theta}}[\theta vB_1\hat{A}_t^{\frac{\theta v-1+\alpha}{1-\alpha}} - \left(\frac{\theta[(v-\delta\alpha)]}{1-\alpha}\right)G_r\hat{A}_t^{\frac{\theta[(v-\delta\alpha)]-1+\alpha}{1-\alpha}}]$$

from which it is easy to show that $\hat{\gamma}_1(\hat{A}_t)$ has a minimum where $\hat{A}_t = \left[\frac{(v-\delta\alpha)G_r}{v(1-\alpha)B_1}\right]^{\frac{1-\alpha}{\alpha}}$ and is monotonically increasing thereafter.

Taking the second derivative:

$$\begin{aligned}\frac{\partial^2\hat{\gamma}_1}{\partial\hat{\gamma}_1^2} &= (B_2B_3)^{\frac{1}{\delta\theta}}[\frac{\theta v[\theta v-1+\alpha]}{1-\alpha}B_1\hat{A}_t^{\frac{\theta v-2+2\alpha}{1-\alpha}} \\ &\quad - \left(\frac{\theta[(v-\delta\alpha)][\theta[(v-\delta\alpha)]-1+\alpha]}{(1-\alpha)^2}\right)G_r\hat{A}_t^{\frac{\theta(v-\delta\alpha)-2+2\alpha}{1-\alpha}}\end{aligned}$$

In order to sign this derivative, we need to distinguish between 3 sets of parameter values.

Case a: $\theta v - 1 + \alpha < 0$ In this case, $\theta(v-\delta\alpha) - 1 + \alpha < 0$ as well. Then, the second derivative is positive, $\hat{\gamma}_1(\hat{A}_t)$ is convex, when $\hat{A} < \left[\frac{(v-\delta\alpha)(\theta(v-\delta\alpha)-1+\alpha)G_r}{(\theta v-1+\alpha)(1-\alpha)}\right]^{\frac{1-\alpha}{\theta\delta\alpha}} < \hat{A}^0$, and concave otherwise. Thus, $\forall\hat{A}_t$ such that $\hat{\gamma}_1(\hat{A}_t) > 0$, the slope of the monotonically increasing function is decreasing.

Case b: $\theta(v-\delta\alpha) - 1 + \alpha < 0 < \theta v - 1 + \alpha$.

In this case, $\forall\hat{A}_t, \frac{\partial^2\hat{\gamma}_1}{\partial\hat{\gamma}_1^2} > 0$.

Case c: $\theta v - 1 + \alpha > 0$

In this case, $\hat{\gamma}_1(\hat{A}_t)$ is convex when $\hat{A} > \left[\frac{(v-\delta\alpha)(\theta(v-\delta\alpha)-1+\alpha)G_r}{(\theta v-1+\alpha)(1-\alpha)}\right]^{\frac{1-\alpha}{\theta\delta\alpha}} < \hat{A}^0$. So, for $\forall\hat{\gamma}_1 > 0$, $\frac{\partial^2\hat{\gamma}_1}{\partial\hat{\gamma}_1^2} > 0$.

To summarize, when $\theta v - 1 + \alpha < 0$, $\hat{\gamma}_1$ is concave $\forall\hat{\gamma}_1 > 0$. When $\theta(v-\delta\alpha) - 1 + \alpha < 0 < \theta v - 1 + \alpha$, the slope is increasing $\forall\hat{A}_t$. When

$\theta v - 1 + \alpha > 0$, the slope is decreasing over some range of positive values of $\hat{A}_t$, but increasing for all positive values of $\hat{\gamma}_1$.

In case (a), given concavity of $\hat{\gamma}_1(\hat{A}_t)$, the potential exists for there to be either 0, 1, or 2 values of $\hat{A}_t$ such that $\hat{\gamma}_1(\hat{A}_t) = \hat{A}_t$. Parameters may be such that $\forall \hat{A}_t$, $\hat{\gamma}_1(\hat{A}_t) < \hat{A}_t$. There may be a single tangency point where $\hat{\gamma}_1(\hat{A}_t) = \hat{A}_t$. Or, a range of values may exist such that $\hat{\gamma}_1(\hat{A}_t) > \hat{A}_t$, with two intersection points.

In cases (b) and (c), given convexity, there will always be a single intersection point where $\hat{\gamma}_1(\hat{A}_t) = \hat{A}_t$.

Analysis of $\hat{\gamma}_2(\hat{A}_t) = B_2 \hat{A}_t^{\frac{\theta}{1-\alpha}}$ is simpler. From the first and second derivatives, it is straightforward to determine that the function is monotonically increasing, and is globally concave if $\theta + \alpha < 1$ and globally convex otherwise.