The B.E. Journal of Macroeconomics

Contributions

Volume 9, *Issue* 1 | 2009 | *Article* 18

A Model of Sequential City Growth

David Cuberes*

*Clemson University, d.cuberes@sheffield.ac.uk

Recommended Citation
David Cuberes (2009) "A Model of Sequential City Growth," *The B.E. Journal of Macroeconomics*: Vol. 9: Iss. 1 (Contributions), Article 18.

A Model of Sequential City Growth*

David Cuberes

Abstract

Strong evidence indicates that in most countries cities tend to develop sequentially, with the initially largest cities growing first. This paper presents a model of city growth that rationalizes this pattern. Increasing returns to scale constitute the force that favors agglomeration of resources in a city, and convex costs associated with the stock of installed capital represent the congestion force that limits city size. The key to generating sequential city growth is the assumption of irreversible investment in physical capital. As expected, the presence of a positive external effect of aggregate city capital on individual firms makes the competitive equilibrium inefficient.

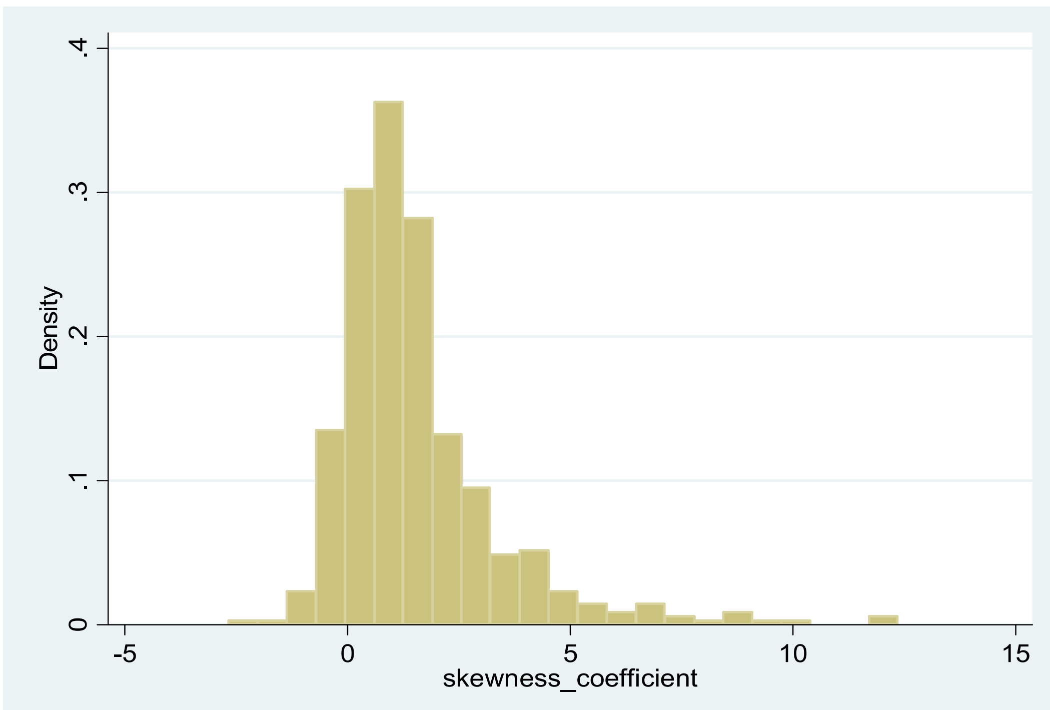
KEYWORDS: cities, city growth, Gibrat's law, increasing returns, congestion costs

*I would like to thank the editor, Esteban Rossi-Hansberg, for very helpful comments. I also thank Roc Armenter, Gary Becker, Jan Eeckhout, Oded Galor, François Gourio, Vernon Henderson, Michal Jerzmanowski, Aitor Lacuesta-Gabarain, Casey Mulligan, Rafael Repullo, Curtis Simon, Balazs Szentes, Robert Tamura, two anonymous referees and seminar participants at the University of Chicago, Clemson University, Universidad de Alicante, Bank of Spain, Universidad de Navarra, Universidad del País Vasco, Universitat Autònoma de Barcelona, and Universitat Pompeu Fabra. All remaining errors are mine.

1. Introduction

In a recent paper Cuberes (2008) presents extensive evidence on the fact that, in most countries, cities tend to grow in sequential order during the 1800-2000 time period. This pattern holds when one considers the administrative definition of a city or its metropolitan area. He first shows that in most decades and countries, the cross-sectional distribution of cities' growth rates is clearly skewed to the right, indicating that a few of them grow much faster than the rest. Figure 1 shows the histogram of these coefficients of skewness for cities.[1]

Figure 1: Histogram of the Coefficient of Skewness of Cities' Growth Rates

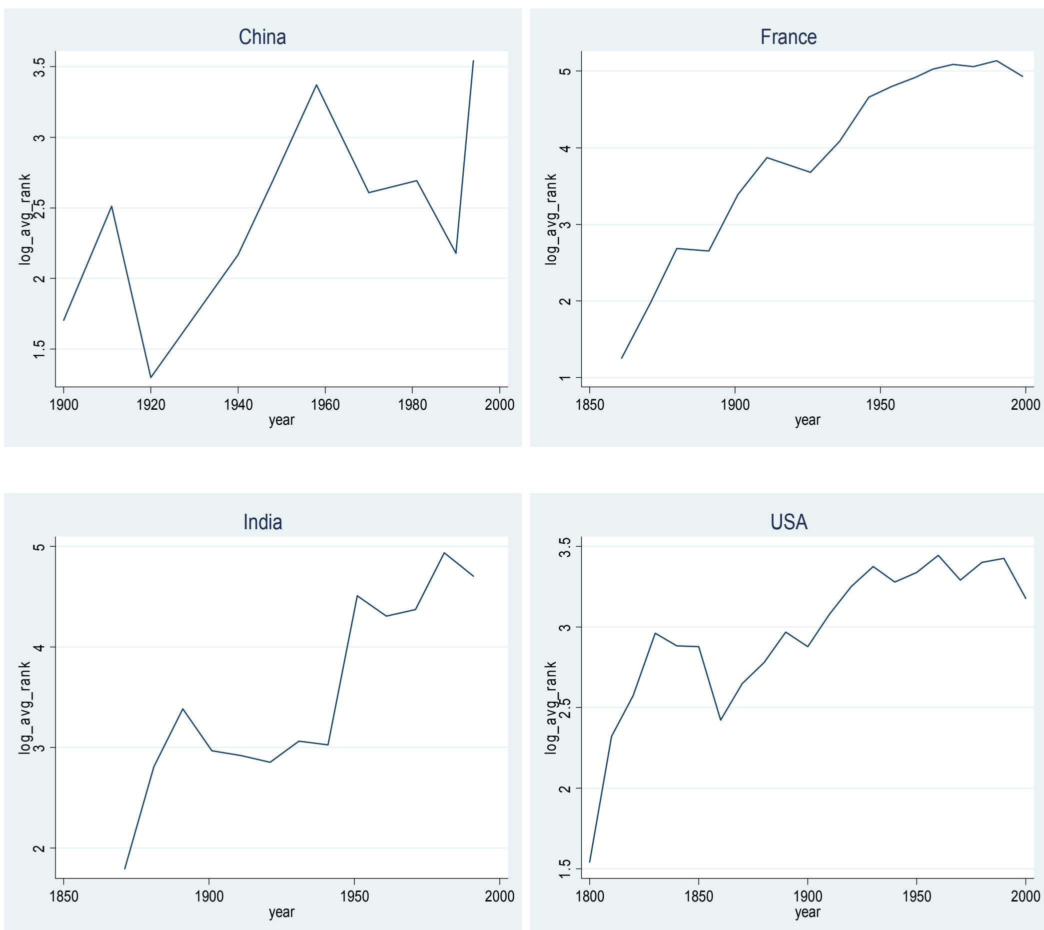


It is apparent from this graph that most of these coefficients are strictly positive. In his sample of 536 country-decades observations for cities, Cuberes finds that the coefficient of skewness is positive in 89% of the cases and significantly so in 73% of the observations. The figures for metropolitan areas are 77% and 63%, respectively.

[1] The corresponding graph for metropolitan areas is very similar. See Cuberes (2008).

Second, he documents that the average size rank (with the largest city having the lowest rank) of the cities that grow the fastest during each decade tends to increase over time. This suggests that a country's initially largest city is, in most cases, the first one to grow at a substantially higher rate than the rest of the cities. Eventually this city reaches a critical size and the initially second-largest city is then the one that grows most rapidly, then the third one, and so on. Figure 2 displays the evolution of the (log) average rank of the 25% fastest growing cities on each decade for four different countries.

Figure 2: Average Rank of the Fastest Growing Cities



The positive slope of this statistic is confirmed by the results of Table 1. This table shows the results of regressing the log of the average rank of the fastest-growing cities on each decade against the log of time and the log of the existing number of cities in the sample used by Cuberes. The time trend is significantly positive, indicating that cities tend to grow in a sequential order.

Table 1: A Regression of the Rank of the Fastest Growing Cities on Time and Number of Cities

	(1)	(2)
log time	0.668*** (0.07)	0.063** (0.03)
log number cities		0.923*** (0.03)
constant	1.317*** (0.11)	-0.478*** (0.06)
R^2	0.331	0.841
N	536	536

Note: The dependent variable is the log of the average rank of the 25% fastest growing cities. Clustered standard errors in parentheses. ** $p < 0.05$, *** $p < 0.01$.

Finally, Table 2 (also from Cuberes 2008) shows that this sequential process is faster during periods when the country's urban population is growing rapidly. Increases in the urban population's growth rate are associated with significant increases in the rank of cities that grow the fastest during each decade.[2]

Table 2: A Regression of the Growth Rate of the Average Rank of the Fastest Growing Cities on the Growth Rate of Urban Population and the Growth Rate in the Number of Cities

	(1)	(2)
growth rate of urban population	1.322*** (0.18)	0.598*** (0.22)
growth rate in number of cities		0.81*** (0.04)
constant	-0.002 (0.07)	-0.046 (0.09)
R^2	0.225	0.626
N	479	479

Note: The dependent variable is the growth rate of the average rank of the 25% fastest growing cities. Clustered standard errors in parentheses. *** $p < 0.01$.

[2] Again, the results for metropolitan areas are qualitatively similar.

The goal of this paper is to develop a simple theoretical model that can explain these stylized facts. The model is standard in the sense that it combines increasing returns and congestion costs, as in many urban models. However, a key difference from most of the existing theoretical literature is the assumption that investment in physical capital in a given city is irreversible. I present the problem of a benevolent social planner and the decentralized one, and show that both of them can qualitatively rationalize the patterns described above.

The paper is organized as follows. In Section 2 I summarize the existing literature on city growth that is closest to my paper. Section 3 presents the optimal and the decentralized versions of the model. In Section 4 I extend the model to an arbitrarily large number of cities, and finally, Section 5 presents the conclusions of the paper.

2. Related Literature

The urban economics literature has long been interested in understanding the patterns of city formation and city development. However, most existing models analyze this problem in a static context. For instance, the system of cities approach (Henderson, 1974) aims to explain the existence of different types of cities at a given point in time.

Fujita (1976) was among the first to study spatial agglomeration in a dynamic context. Other important contributions are Anas (1978, 1992), Kanemoto (1980), Henderson and Ioannides (1981), Miyao (1981), Fujita (1982), Krugman (1992), Ioannides (1994), and Palivos and Wang (1996).[3]

One important limitation of these models is that they assume free mobility of all factors of production. Consequently, they predict large and rapid swings in the population of cities that reach a critical level. Moreover, when new cities form, their population jumps instantly to some arbitrarily large size. These two predictions are clearly counterfactual- the existing data on cities' population exhibit smooth fluctuations as countries urbanize. See Henderson and Venables (2009) for a more detailed discussion on this.

There have been different attempts to generate smooth changes in cities' population as countries urbanize. Brezis and Krugman (1997) develop a model in which cities grow sequentially following technological innovations. Upon the introduction of the new technology, there is a rapid drop in the original city's population, but after that, its population declines progressively as the new city's relative productivity increases with learning.[4] While this seems a plausible theoretical mechanism, it is hard to argue that the introduction of new

[3] An excellent review of this literature can be found in Duranton and Puga (2004) and Rossi-Hansberg and Wright (2007).
[4] See Desmet (2002) for a similar mechanism.

technologies is the main force driving the dynamics of cities' populations. Another strategy is to assume that one of the production factors is immobile. Here the limited existing literature typically focuses on the existence of irreversible investment in some capital good. Fujita (1978) describes an optimal dynamic equilibrium with capital immobility, but he does not present its decentralized version.[5]

My paper relates most closely to Henderson and Venables (2009), HV henceforth.[6] HV analyze city formation in an economy that experiences steady growth of its urban population over time. The presence of immobile housing and urban infrastructure, and the fact that agents are forward looking, generates a sequential pattern of city formation and predicts smooth changes in the population of cities, i.e. it avoids the abrupt swings implied by most of the previous literature. The paper also analyzes how different institutions affect the equilibrium city size, which may or may not be optimal.

While my setup is simpler, it complements HV in important ways. First, it emphasizes the role of physical capital in generating increasing returns at the city level. In HV the immobile durable good is residential capital, which is introduced in their model to avoid large and sudden changes in a city's population. Although the authors argue that this capital good can also be thought of as public infrastructure, increasing returns in their theory are entirely driven by workers' interaction with each other. The capital good plays no role in generating the centripetal force of the model.

In contrast, in my model, immobile physical capital is an input in each firm's production function, and it is hence a crucial ingredient to generate increasing returns at the city level. This distinction is important for two reasons: First, from a theoretical point of view, several papers show that it is reasonable to think that it is indeed the interaction of workers and urban infrastructure (i.e. physical capital in my model) that generates increasing returns at the city level. For instance, Duranton and Puga (2004) argue that increasing returns may be generated by the fact that consumers share indivisible goods and facilities.[7]

Second, and perhaps more relevant, the fact that the capital good is an input in the production function allows me to endogenize the interest rate. HV consider the problem of a small open economy and hence take the interest rate as given. In my model the economy is closed and households have concave utility functions, so that both savings and the interest rate are endogenous. This allows

[5] Brueckner (2000) and Helsley and Strange (1994) introduce immobile capital, but the former focuses on only one city, while the latter is a static model.

[6] The two papers were written independently and at the same time.

[7] One can interpret the capital stock in my model as partly representing such indivisible goods. Other papers that model cities as the outcome of large indivisibilities in production are Koopmans (1957), Mills (1967), and Mirrlees (1972).

one to have relevant predictions about how productive capital -infrastructure in general- is in different cities and at different points in time during the process of a country's development. These predictions could be tested using data on countries during their period of rapid urbanization. Although data on capital stocks at the city level may be difficult to obtain, one could attempt to use regional or state data. This historical data could be found for instance in Baier et al. (2008), who construct time series on stocks of capital at the state level for the United States during the 1840-2000 period. The idea would then be to construct estimates of the marginal product of capital in different regions/states and over time to analyze whether they can be matched with the theory proposed here.[8] This exercise –left for further research- would be important for academics and policymakers interested in countries that are currently experiencing an ongoing process of urbanization. For instance, Bennathan and Canning (2000) study the social rate of return on infrastructure investments in developing countries, though they lack a theoretical framework to interpret their findings.[9] Another interesting question that may be answered with this type of test is whether capital flows from big cities to small ones or vice versa, an analog of Lucas' (1990) paper on capital flows across countries.

Finally, an important methodological contribution of the model presented here is that it analyzes how the introduction of increasing returns and convex congestion costs modify the dynamics and predictions of the standard neoclassical growth model.

3. The Model

In this section I first present the problem of a benevolent social planner and then solve its decentralized version. The possibility of multiple equilibria in the latter setup is also discussed.

3.1. The Planner's Problem

Consider a closed economy with two cities[10] *A,B* that are modeled as firms with identical Cobb-Douglas production functions. In each city, production takes place using two inputs: capital and labor. Output Y^j in city j is given by

---

[8] Caselli and Feyrer (2007) conduct a similar exercise at the country level.

[9] Ioannides (2008) builds up an endogenous urban growth model in which governments invest in public infrastructure.

[10] Anas (1992) studies city development in a two-city perfect-foresight economy but without reproducible capital.

$$Y^j = \left(N^j\right)^{\alpha}\left(K^j\right)^{\beta}$$

where N^j and K^j represent the number of workers and the stock of capital used in production, respectively.

I assume that both cities face diminishing returns to the each of the two inputs K and N but increasing returns to scale, i.e. the production parameters satisfy the following inequalities: $0 < \alpha < 1$, $0 < \beta < 1$, and $\alpha + \beta > 1$. At the same time, each city is subject to congestion costs that are modeled as convex costs $g\left(K^j\right)$ associated with the stock of installed capital in that city.[11] This congestion cost function satisfies $g(0) = 0$, $g' > 0$, and $g'' > 0$.

Investment in the model is assumed to be irreversible. This is a reasonable assumption if one interprets physical capital - or at least a significant fraction of it- as infrastructure (buildings, urban facilities, etc.). It is also assumed that labor can migrate across cities at no cost. Finally, total population N is constant over time.[12]

The problem of a benevolent social planner is to allocate labor and capital across cities and over time so that the utility of a representative individual is maximized. To simplify notation I omit time subscripts except when there may be some confusion. Assuming households have a logarithmic utility, this problem can be written as:

$$\max \int_0^{\infty} e^{-\rho t} \ln(c)\, dt$$

$$\sum_{j=A,B} N^j = N$$

$$\sum_{j=A,B} I^j + \sum_{j=A,B} g\left(K^j\right) + C = \sum_{j=A,B} Y^j$$
$$Y^j = \left(N^j\right)^{\alpha}\left(K^j\right)^{\beta}, \forall j = A, B$$
$$\dot{K}^j = I^j - \delta K^j, \forall j = A, B$$
$$I^j \geq 0, \forall j = A, B$$
$$K_0^j \; given, \forall j = A, B$$

---

[11] Congestion costs can also arise from consumers' disutility associated to crowded cities. To make the model tractable, I assume that these costs are originated in the production side of the economy. It is possible to show that allowing consumers to bear a fraction of the costs does not change the model's qualitative results.

[12] In Section 3.1.4 I analyze the effect of an exogenous increase in N.

where $\rho \in (0,1)$ is the household's discount rate and C and $C \equiv \frac{c}{N}$ represent consumption and consumption per capita, respectively. Note that, for simplicity, consumption is assumed to be not city-specific. I^j and K_0^j respectively denote investment and the initial stock of capital installed in city j, and $\delta \in (0,1)$ is the depreciation rate of capital.

I proceed to solve this problem in two stages. In the first one, labor is allocated optimally between the two cities, taking the two stocks of capital as given. In the second stage, using the optimal labor allocation, the planner assigns capital across cities and over time to maximize utility.

3.1.1. Optimal Allocation of Labor

The planner's problem in this stage is static and reduces to maximize total national production, i.e.

$$\max_{N^A, N^B} \sum_{j=A,B} Y^j$$
$$\sum_{j=A,B} N^j = N$$
$$Y^j = \left(N^j\right)^\alpha \left(K^j\right)^\beta, \forall j = A, B$$
$$K^j \textit{ given}, \ ,\forall j = A, B$$

The solution to this problem is

$$\frac{N^A}{N^B} = \left(\frac{K^A}{K^B}\right)^{\frac{\beta}{1-\alpha}} \qquad (1)$$

This optimality condition states that the ratio of labor between the two cities is an increasing function of their ratio of capital stocks. Note that increasing returns imply $\frac{\beta}{1-\alpha} > 1$, and so at each point in time the city with the largest stock of capital also has the largest population. Moreover, increases in the (log) capital ratio are associated with more than proportional increases in the (log) population ratio.

3.1.2. Optimal Allocation of Capital

In this stage of the problem the planner takes into account the optimal labor allocation of the first sub-problem and allocates capital across cities and over time to maximize the flow of utility. Defining all city-*j* specific variables in per capita terms as $x^j \equiv \frac{X^j}{N}$, one can now write the problem as

$$\max \int_0^\infty e^{-\rho t} \ln(c) dt$$

$$\sum_{j=A,B} i^j + \frac{1}{N} \sum_{j=A,B} g\left(K^j\right) + c = f\left(k^A, k^B, N\right)$$

$$\dot{k}^j = i^j - \delta k^j, \forall j = A, B$$
$$i^j \geq 0, \forall j = A, B$$
$$k_0^j \; given, \forall j = A, B$$

where

$$f\left(k^A, k^B, N\right) \equiv \frac{F\left(k^A, k^B, N\right)}{N} = N^{\alpha+\beta-1} \Omega\left(k^A, k^B, N\right)^{1-\alpha} \qquad (2)$$

and

$$\Omega\left(k^A, k^B, N\right) \equiv \sum_{j=A,B} \left(k^j\right)^{\frac{\beta}{1-\alpha}} \qquad (3)$$

$F\left(k^A, k^B, N\right)$ is the value of total output that results from inserting the expression for the optimal allocation of labor (1) into total national production $Y^A\left(N^A, K^A\right) + Y^B\left(N^B, K^B\right)$ and $k^j \equiv \frac{K^j}{N}$ is the per-capita stock of capital in city *j*. This function is derived in the Appendix. The present-value Hamiltonian of this problem is

$$H = e^{-\rho t} \ln c + \lambda \left[f\left(k^A, k^B, N\right) - \sum_{j=A,B} i^j - \frac{1}{N} \sum_{j=A,B} g\left(K^j\right) - c \right] + \sum_{j=A,B} \mu^j \left(i^j - \delta k^j\right)$$

where $\lambda, \mu^j, j = A, B$ are Lagrange multipliers. The first order conditions with respect to consumption and the city-*j* specific variables are:

$$\frac{\partial H}{\partial c} = 0 \Leftrightarrow e^{-\rho t}\frac{1}{c} = \lambda \qquad (4)$$

$$\frac{\partial H}{\partial i^j} \leq 0 \Leftrightarrow -\lambda + \mu^j \leq 0 \qquad (5)$$

$$\frac{\partial H}{\partial k^j} = -\dot{\mu}^j \Leftrightarrow \lambda\left[f_j^p\left(k^A, k^B, N\right) - g'\left(K^j\right)\right] - \mu^j \delta = -\dot{\mu}^j \qquad (6)$$

and the transversality condition is $\lim_{t\to\infty} k^j(t)\mu^j(t) = 0$. The expression $f_j^p\left(k^A, k^B, N\right) \equiv \frac{\partial f\left(k^A, k^B, N\right)}{\partial k^j}$ represents the planner's gross (i.e. without taking into account congestion costs) marginal product of capital in city *j*:[13]

$$f_j^p\left(k^A, k^B, N\right) = \beta N^{\alpha+\beta-1}\Omega\left(k^A, k^B, N\right)^{-\alpha}\left(k^j\right)^{\frac{\beta}{1-\alpha}-1} \qquad (7)$$

Equation (7) states that the gross marginal product of capital (*MPK* henceforth) in city *j* depends on the two stocks of capital and on total population. This comes from the fact that labor allocates optimally across cities based on their stocks of capital (equation (1)). The following proposition shows that, due to the presence of increasing returns to scale, at any point in time, the gross *MPK* is strictly larger in the city with the largest stock of capital.

Proposition 1: $f_A^p\left(k^A, k^B, N\right) > f_B^p\left(k^A, k^B, N\right), \forall k^A > k^B > 0$

Proof: Assume, by contradiction, that $k^A > k^B > 0$ but

$f_A^p\left(k^A, k^B, N\right) \leq f_B^p\left(k^A, k^B, N\right)$. This implies $\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}-1} \leq 1$ which is a

contradiction since $k^A > k^B$ and $\alpha + \beta > 1$. *Q.E.D.*

The next two assumptions establish some initial conditions that initiate the dynamics of the model. Assumption 1 states that, at the initial date, city *A* has a slightly larger stock of capital than city B.[14]

[13] The superscript *p* stands for "planner". This notation helps me comparing this setup with the decentralized one.
[14] A similar assumption is typically made in models of uneven development. See Krugman (1981).

Assumption 1

$$K^A(0) = K^B(0) + \varepsilon$$

where ε is a small positive number.

It is now convenient to define the net *MPK* in city *j*, $\tilde{f}_j(k^A, k^B, N)$ as the difference between its gross *MPK* and its marginal congestion costs[15]:

$$\tilde{f}_j = f_j^p - g'(K^j)$$

Assumption 2 states that, at date zero, congestion costs in the city with the largest stock of capital, city *A*, are relatively small compared to the productivity gains associated with its large size. Moreover, it is assumed that investment is profitable in the smallest location, city *B*, i.e. its net *MPK* is strictly larger than the sum of the discount rate and the depreciation rate. Obviously, this assumption also ensures that in this economy investment is profitable in city *A* at date zero.

Assumption 2

$$\tilde{f}_A(0) > \tilde{f}_B(0) > \rho + \delta$$

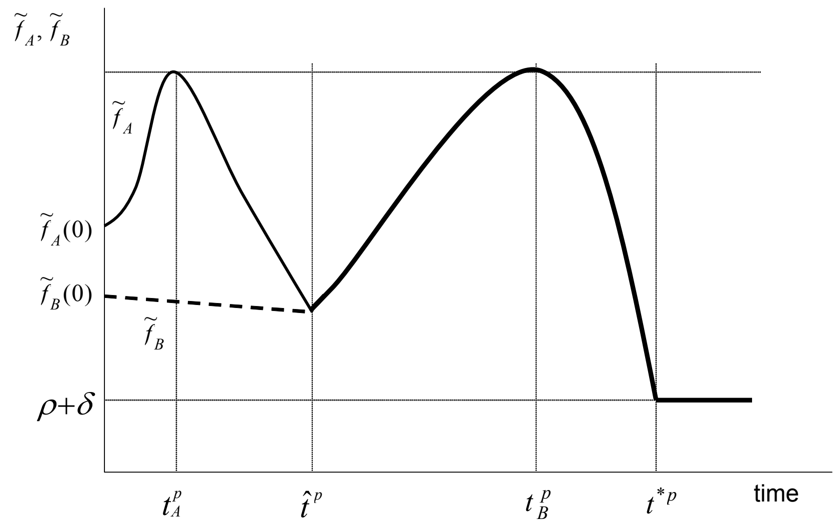
3.1.3. Equilibrium Transition and Steady State

The qualitative features of the unique equilibrium in this economy are displayed in Figures 3-5. Figure 3 plots the evolution of the equilibrium net *MPKs* in the two cities. In the time interval $\left(0, t_A^p\right]$, the gross *MPK* in the city with the largest stock of capital, city *A*, increases very fast.

---

[15] To simplify the notation, I will write $\tilde{f}_j(k^A, k^B, N)$ as $\tilde{f}_j$ omitting its dependence on the two stocks of capital k^A, k^B and on total population *N*.

Figure 3: Net *MPKs*



This is the case because, as a result of its initial advantage (see Assumptions 1 and 2) and the existence of increasing returns to scale, all the new investment in capital goes to city *A* and labor reallocates from city *B* to city *A* at an increasing rate. At period t_A^p, congestion costs in city *A* start dominating increasing returns, and so the *MPK* in that city begins to decline rapidly. Meanwhile, the *MPK* in city *B* decreases for two reasons: First, capital depreciates there at the constant rate δ because no investment takes place in that city. Second, labor migrates to city *A* hence further decreasing the *MPK* in city *B*.

Eventually, the economy reaches period $\hat{t}^p$, at which the two *MPKs* are equated. From that point on, non-arbitrage implies that these *MPKs* must remain equal to each other until the economy reaches its steady state. In order for this pattern to be optimal, one needs to make the additional assumption that in period $\hat{t}^p$, investment in city *B* is still profitable, i.e.[16]

[16] It is clear that, without this assumption, this would become a one-city model, which would not help us understand any of the empirical facts described in the introduction. Note that assuming $k_B(\hat{t}^p) > 0$ is necessary but not sufficient. The remaining stock $k_B(\hat{t}^p)$ must be large enough to make investment in city *B* socially optimal.

Assumption 3

$$\tilde{f}_B(\hat{t}^p) > \rho + \delta$$

The two *MPKs* increase after period $\hat{t}^p$ until congestion costs - now in the two cities- become critical again (period t_B^p). Finally, from period t_B^p until period t^{*p} the common *MPK* smoothly converges to its steady-state value $\rho + \delta$.

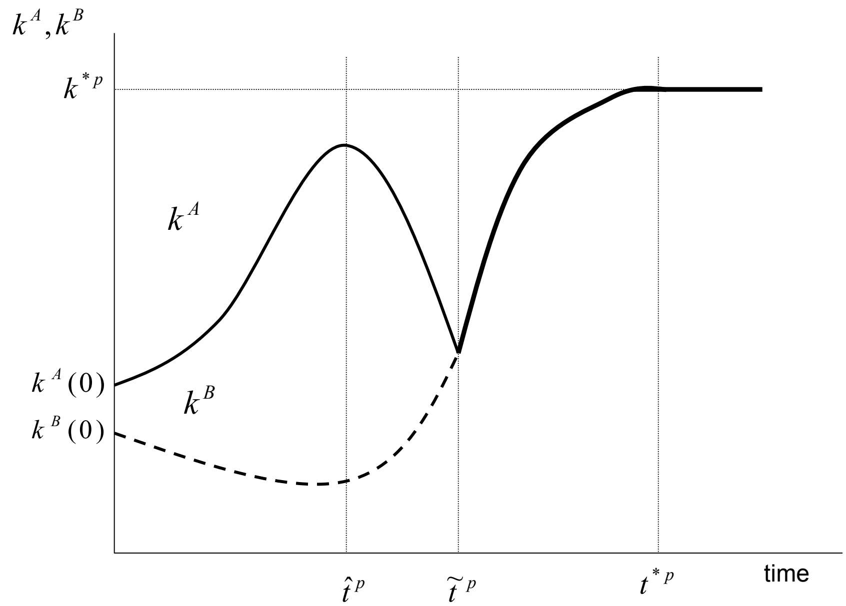
Figure 4 displays the paths of the two stocks of capital. As stated above, k^A increases between dates zero and $\hat{t}^p$, when the two *MPKs* are equalized. During this period the stock of capital in city *B* declines as a result of the depreciation. Non-arbitrage implies that from that point on, the pattern of growth of the two cities must be such that this equality holds. Since city *A* is relatively large and city *B* relatively small this can only happen if the stock of capital increases in *B* and falls in *A*. Therefore, for some periods, investment turns positive in *B* and zero in *A*. The decline in the stock of capital in *A* is again driven by the depreciation rate. At some point, the stocks of capital are equalized in the two cities. Let $\tilde{t}^p$ be the period in which this occurs. Note that this period is well defined because it is impossible to attain the steady state unless the two cities have already reached a critical level of congestion costs.[17] Once the two stocks of capital are the same, they grow together until the economy reaches its steady state (at period t^{*p}).[18]

Also note that in this last transition the common net *MPK* will eventually reach its maximum value (at period t_B^p). However, the lack of a better investment alternative makes the stock of capital in each city continue growing to its steady state level k^{*p} .

[17] *MPK* in city *B* increases in the $(\hat{t}^p, \tilde{t}^p]$ interval and, from Assumption 3, $\tilde{f}_B(\hat{t}^p) > \rho + \delta$. Since in steady state $\tilde{f}_B(t^{*p}) = \rho + \delta$, the steady state must be reached later than $\tilde{t}^p$.

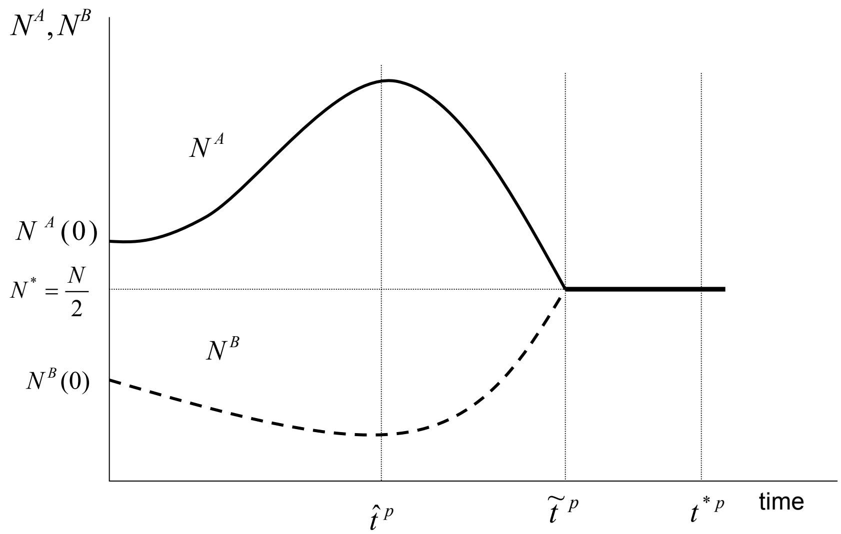
[18] Unlike in the neoclassical growth model, the presence of convex congestion costs implies that this steady state is reached in finite time so that t^{*p} is also well defined.

Figure 4: Capital



Finally, Figure 5 shows the evolution of the population in the two cities. It is clear that in the time interval $\left(0, \tilde{t}^p\right]$, population mimics the pattern of capital because the optimal allocation rule is given by $\frac{N^A}{N^B} = \left(\frac{K^A}{K^B}\right)^{\frac{\beta}{1-\alpha}}$. At period $\tilde{t}^p$, when the stocks of capital are the same in both cities, the incentives to migrate disappear and the two cities become identical in all respects. Therefore, the steady-state level of population in each city is $N^* = N\left(\tilde{t}^p\right) = \frac{N}{2}$.

Figure 5: Labor



The following proposition calculates the growth rate of per capita consumption in the different critical time intervals defined above.

Proposition 2: The growth rate of the per capita consumption is given by

$$\gamma_c^p \equiv \begin{cases} \tilde{f}_A - \rho - \delta, \forall t \in \left(0, \hat{t}^p\right] \\ \tilde{f}_j - \rho - \delta, \forall t \in \left(\hat{t}^p, t^{*p}\right], \forall j = A, B \\ 0, \forall t > t^{*p} \end{cases}$$

Proof: Consider the first order conditions (4)-(6). Suppose that in the time interval $\left(0, \hat{t}^p\right]$, $i^B > i^A = 0$. From (5) this implies that $\lambda = \mu^B$ and so (6) becomes

$$\frac{\dot{\lambda}}{\lambda} = \delta - \tilde{f}_B$$

Taking logs and derivatives in (4)

$$-\rho - \frac{\dot{c}}{c} = \frac{\dot{\lambda}}{\lambda}$$

Then, equating the last two expressions one has

$$\gamma_c^p = \widetilde{f}_B - \rho - \delta$$

From Proposition 1, the inequality $f_B^p < f_A^p$ is satisfied in this time interval. Moreover, as long as $t \leq \hat{t}^p$ congestion costs in city *A* are not large enough to make its net *MPK* equal to that of city *B*, so $\widetilde{f}_B < \widetilde{f}_A$. It follows that the proposed investment policy cannot be optimal. The same logic implies that any simultaneous positive investment in the two cities during this interval is also suboptimal. Therefore

$$\gamma_c^p = \widetilde{f}_A - \rho - \delta$$

in the time interval $\left(0, \hat{t}^p\right]$. In the time interval $\left(\hat{t}^p, t^{*p}\right]$ the two net *MPKs* are equated and so the growth rate of consumption is given by

$$\gamma_c^p = \widetilde{f}_j - \rho - \delta$$

where $\widetilde{f}_j$ is the *MPK* in the two cities. Finally, at any period $t > t^{*p}$ consumption is constant since the economy has reached the steady state. *Q.E.D.*

The steady state of the economy is then characterized by the following two equations:

$$\widetilde{f}_j\left(t^{*p}\right) = \rho + \delta \quad (8)$$

$$2\delta k^{*p} + \frac{2g\left(K^{*p}\right)}{N} + c^{*p} = f\left(k^{*p}, N\right) \qquad (9)$$

3.1.4. The Effect of an Exogenous Population Shock

To make the model tractable, I have so far assumed that total population *N* is constant over time. In this section I analyze how the dynamics of the model change when the economy experiences an exogenous increase in its population. One can interpret this experiment as a rapid increase in the country's urbanization process. Such rapid changes in urban population are not uncommon in the data, especially in developing countries.[19] The next proposition shows that increasing returns imply that the gross *MPK* in both cities increases as a result of the shock.

Proposition 3: An exogenous increase in population *N* increases the gross *MPK* in the two cities.

Proof: Using the fact that $k^j \equiv \frac{K^j}{N}$ the gross *MPK* in city *j* (7) can be expressed as

$$f_j^p = \beta N^{\alpha} \hat{\Omega}\left(K^A, K^B\right)^{-\alpha} \left(K^j\right)^{\frac{\beta}{1-\alpha}-1}$$

where

$$\hat{\Omega}\left(K^A, K^B\right) = \sum_{j=A,B} \left(K^j\right)^{\frac{\beta}{1-\alpha}}$$

Since the term $\hat{\Omega}\left(K^A, K^B\right)^{-\alpha} \left(K^j\right)^{\frac{\beta}{1-\alpha}-1}$ is independent of *N* one has that the effect of an increase in *N* on the *MPK* of city *j* is

$$\frac{\partial f_j^p}{\partial N} = \alpha\beta N^{\alpha-1} \hat{\Omega}\left(K^A, K^B\right)^{-\alpha} \left(K^j\right)^{\frac{\beta}{1-\alpha}-1} > 0$$

Q.E.D.

Proposition 4 states that increasing returns also imply that the population shock increases the gross return to capital more in the largest city than in the smallest one.

Proposition 4: An exogenous increase in population *N* increases the gross *MPK* in the largest city more than in the smallest one.

Proof: Suppose the opposite is true i.e. $k^A > k^B$ but $\frac{\partial f_A^p}{\partial N} \leq \frac{\partial f_B^p}{\partial N}$. This implies

$\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}-1} \leq 1$ which is a contradiction since $\alpha + \beta > 1$ and $k^A > k^B$. *Q.E.D.*

Finally, the next corollary shows that any increase in *N* is in general associated with a more pronounced process of sequential city growth.

[19] Henderson and Wang (2007) document that the urbanization rate is most rapid at low-income levels and then tails off as countries become "fully urbanized."

Corollary: An exogenous increase in population *N* at any period $t \leq \hat{t}^p$ accelerates the process of sequential city growth

Proof: Suppose that the shock occurs at $t \in \left(0, \hat{t}^p\right]$. In equilibrium, city *A* is the only one growing in this time interval and from Propositions 3 and 4 the shock makes city *A* grow faster. Moreover, its marginal congestion costs $g'\left(K^A\right)$ increase with *N*, since $K^A = k^A N$. These two forces imply that $\hat{t}^p$ must now be lower i.e. the period at which the population of city *B* starts growing is reached faster and so the process of sequential growth is more pronounced. *Q.E.D.*

In the context of two cities, if the shock takes place at any $t > \hat{t}^p$, the speed of sequential growth would not be affected. Consider first the case in which the shock occurs at $t \in \left(\hat{t}^p, \tilde{t}^p\right]$. Since in this time interval city *A* is too congested, most of the additional population (and hence investment) will locate in city *B* until this city reaches its steady state. Population would optimally be allocated between the two cities, but the model does not allow for the creation of a (potentially optimal) third city. So city *B* grows faster as a consequence of the shock; however this has no implications for sequential growth because there is no “rank 3” city to start growing. If total population increases after the economy’s steady-state has been reached, the additional population is split evenly between the two cities because at that point they are identical. Therefore, there is no change in the speed of sequential growth either.

If one extends the model to $M > 2$ cities (see Section 4), an increase in total population translates into faster sequential growth unless if this shock takes place after the $M-1$ city experiences its period of solo growth. For *M* large enough, the probability of the shock occurring in this time interval is arbitrarily small, and so the model predicts that, in most cases, an increase in total population implies faster sequential city growth.

3.1.5. Can the Model Match the Stylized Facts?

The model is clearly able to generate the three facts documented in Cuberes (2008). Population in the initially large city grows alone in the time interval $\left(0, \hat{t}^p\right]$. Eventually, congestion costs in this city become too large, and investment starts in the second-largest city, whose population is the only one that increases between the periods $\hat{t}^p$ and $\tilde{t}^p$.[20] Therefore, in most periods, the population of one city grows much faster than the rest and cities grow in a sequential order.

[20] Between periods $\tilde{t}^p$ and t^{*p}, the population of the two cities grows at the same rate (zero), i.e., Gibrat’s law holds.

Moreover, it is always the largest city that grows the fastest, conditional on not having reached a critical level of congestion costs. These two predictions correspond with the first two facts documented in Cuberes (2008). Finally, in the model, increases in population are associated with a more rapid process of sequential growth, which is the third empirical finding in Cuberes' paper.

3.2. The Decentralized Economy

In this section I show that, if one introduces a market structure in the model, the same qualitative results as in the planner's setup would hold. I also show that, because households do not internalize the positive externality generated by larger cities, the rate of investment is suboptimal in the decentralized equilibrium. Finally, I rule out the possibility of multiple equilibria in two different ways.

3.2.1. Households

There are two cities (*A,B*) in the model and a continuum of atomistic households that can invest in physical capital employed in these cities to produce final goods.

The sources of income of a representative household are wage earnings and the return on assets. Let ω be the equilibrium wage rate in this economy.[21] z^j is the amount of assets invested by this household in city *j*, *j=A,B*. As in the planner's problem, investment in physical capital is assumed to be irreversible. The household's capital income is given by $\sum_{j=A,B} r^j z^j$, where r^j represents the net return to capital (net of congestion costs) in city *j* and his budget constraint is therefore

$$\sum_{j=A,B} i^j + c = \omega + \sum_{j=A,B} r^j z^j$$

where i^j is investment in city *j* and *c* denotes consumption. Assuming logarithmic utility, the problem of a representative household (omitting time subscripts) can then be written as:

[21] It is shown below that in equilibrium there is a unique national wage rate.

$$\max \int_0^\infty e^{-\rho t} \ln(c) dt$$

$$\sum_{j=A,B} i^j + c = \omega + \sum_{j=A,B} r^j z^j$$

$$i^j \geq 0, \forall j = A, B$$

$$z_0^j \; given, \forall j = A, B$$

where $\rho \in (0,1)$ is the discount rate, which is assumed to be the same as the social planner's one. As stated above, households face the irreversibility constraints $i^j \geq 0, j = A, B.$ Finally, z_0^j is the initial stock of assets in city *j*.

3.2.2. Firms

Firms use a constant-returns-to-scale technology to produce goods. However, each firm is subject to a positive external effect à la Romer (1986), which comes from the total stock of capital installed in the city where it operates.[22] The production function of firm *i* located in city *j* is given by

$$Y^{ij} = \left(N^{ij}\right)^{\alpha} \left(K^{ij}\right)^{1-\alpha} \left(K^j\right)^{\psi}$$

where Y^{ij}, N^{ij}, and K^{ij}, respectively, represent the firm's production, labor, and capital. K^j is the total stock of capital installed in city *j*, i.e. $K^j \equiv \sum_{i=1}^{I} K^{ij}$, and *I* is the number of firms operating in that city. The parameter ψ captures the size of the external effect of aggregate city capital on any firm that operates in the city. I assume that ψ is strictly positive, and, to make this problem comparable with the one of the social planner, I impose $\psi = \alpha + \beta - 1$.

Firms hire labor and capital at their competitive prices and sell their product in the market. Moreover, they must pay a fraction $\frac{1}{I}$ of the congestion cost $g\left(K^j\right)$ generated by the total stock of capital installed in the city where they operate, where *g(.)* is the same function as the one used in the planner's problem. Normalizing the price of the consumption good to one, profits of firm *i* located in city *j*, are then given by

$$\pi^{ij} = \left(N^{ij}\right)^{\alpha} \left(K^{ij}\right)^{1-\alpha} \left(K^j\right)^{\psi} - (r^j + \delta) g\left(K^{ij}\right) - \frac{1}{I} g\left(K^j\right) - \omega N^{ij}$$

[22] Other models in which centripetal forces are generated by external economies of scale include Chipman (1970), Abdel-Rahman (1990) and Krugman (1992).

where $\delta \in (0,1)$ is the rate at which capital depreciates. The first order conditions of the firm with respect to labor and capital are, respectively:

$$\alpha\left(N^{ij}\right)^{\alpha-1}\left(K^{ij}\right)^{1-\alpha+\psi} = \omega \qquad (10)$$

$$(1-\alpha)\left(N^{ij}\right)^{\alpha}\left(K^{ij}\right)^{\psi-\alpha} = r^{j} + \delta + g'\left(K^{j}\right) \qquad (11)$$

where the symmetric equilibrium condition $K^{ij} = K^{j}, \forall i$, has been imposed.[23]

As in the planner's problem, it is assumed that labor can migrate across cities at no cost and that there is no unemployment. Therefore, in equilibrium there is a unique national wage rate and from (10) labor allocates across cities to satisfy

$$\frac{N^{A}}{N^{B}} = \left(\frac{K^{A}}{K^{B}}\right)^{\frac{1+\psi-\alpha}{1-\alpha}} \qquad (12)$$

Note that, since it is assumed that $\psi = \alpha + \beta - 1$, this allocation rule is identical to the one chosen by the planner's (see equation (1)). From (11) one has

$$r^{j} = f_{j}^{m} - \delta - g'\left(K^{j}\right) (13)$$

where

$$f_{j}^{m} \equiv (1-\alpha)\left(N^{j}\right)^{\alpha}\left(K^{j}\right)^{\psi-\alpha} \qquad (14)$$

represents the decentralized[24] gross *MPK* in city *j*.

3.2.3. Equilibrium Transition and Steady State

Since the economy is closed and the only available asset is physical capital one has $z^{j} = k^{j}, \forall j = A, B$. The equilibrium is given by equations (12)-(14) and the market clearing condition

[23] Since there are constant returns to scale at the firm level, the number and size of firms are indeterminate, and thus one can focus on the problem of a representative firm in each city and normalize *I* to one.
[24] The superscript *m* stands for "market".

$$N^A + N^B = N \qquad (15)$$

along with the normalization $k^j \equiv \dfrac{K^j}{N}, \forall j = A, B$ which was already used in the planner's problem. In the Appendix it is shown that, combining these equations, the equilibrium gross *MPK* in city *j* is given by

$$f_j^m(k^A, k^B, N) = (1-\alpha)N^{\psi}\Omega\left(k^A, k^B, N\right)^{-\alpha}\left(k^j\right)^{\frac{\beta}{1-\alpha}-1} \qquad (16)$$

where

$$\Omega\left(k^A, k^B, N\right) \equiv \sum_{j=A,B}\left(k^j\right)^{\frac{\beta}{1-\alpha}} \qquad (17)$$

Note that, again since $\psi = \alpha + \beta - 1$, this expression is an analog of the corresponding *MPK* of the planner's problem (see equation (7)). The following proposition shows that, as expected, the decentralized equilibrium is inefficient because firms do not internalize the external effect they receive from the aggregate stock of capital in the city where they operate. Consequently, the transition to the steady state is too slow.

Proposition 5: The gross *MPK* of the city that experiences a positive investment is always lower in the decentralized economy than in the planner's one.

Proof: From (7) and (16) the market-planner's ratio of the gross *MPK* of city *j* is $\dfrac{1-\alpha}{\beta}$ which is less than one due to the presence of increasing returns to scale. *Q.E.D.*

The assumptions made in the decentralized economy are essentially the same as the ones of the planner's problem. As I explain in the next section, Assumption 1 requires a further restriction on the size of the initial gap ε between the stocks of capital of the two cities to guarantee uniqueness of equilibrium. Assumption 2 is now stated as:

Assumption 2'

$$r^A(0) > r^B(0) > \rho$$

Finally, Assumption 3 needs to be modified as follows:

Assumption 3'

$$r_B\left(\hat{t}^m\right) > \rho$$

Since the two *MPKs* differ only by a multiplicative constant, Proposition 1 also holds in the decentralized economy. Proposition 2 is essentially the same as in the planner's problem, although the critical periods are now different. Finally, Propositions 3-4 and the corollary also hold in this setup.[25] The only difference is that now the effect of an exogenous population shock is smaller than in the planner's case. Once again, the reason for this is that firms do not internalize the positive externality generated by the stock of capital installed in the city where they produce. Finally, the equations that characterize the steady state are the analog of (8)-(9) with the corresponding superscript "m".

3.2.4. History versus Expectations

In this section I explore the possibility that the decentralized economy has a different equilibrium than the one described above. Suppose everyone believes that the initially small city will grow first. Under this scenario increasing returns imply that it may be rational for all investors to initially invest in this city only. In principle, it is possible that any deviant would experience a net loss from this strategy and so this may indeed be an equilibrium.[26] Assume one generalizes the production function of a representative firm *i* in city *j* as:

$$Y^{ij} = A^{ij}\left(N^{ij}\right)^{\alpha}\left(K^{ij}\right)^{\beta}$$

where A^{ij} represents total factor productivity (*TFP*) of firm *i* in city *j*. In the previous exposition of the model I assumed $A^{ij} = 1, \forall i = 1,...,I, \forall j = A, B$ because this paper abstracts from the effect of technology on city growth. Here I consider *TFP* explicitly since it helps exploring the possible existence of multiple equilibria.

Assume, as in the previous sections, that all firms within a city are identical so that one can omit the superscript *i*. Any firm in city *j* produces output according to:

---

[25] In the corollary again $\hat{t}^p$ must be replaced by $\hat{t}^m$.

[26] It is obvious that this would be an inefficient equilibrium, but in principle there is no reason why it would fail to exist.

$$Y^j = A^j \left(N^j\right)^{\alpha} \left(K^j\right)^{\beta}$$

Consider in two different cases. In the first one, *TFP* is the same in the two cities. In the second one, each city has a different *TFP* parameter.

Case 1: $A^A = A^B$

This is a situation in which *TFP* is the same in the two cities, as in the main presentation of the model. It is argued there that, in this case, when the two cities converge in terms of capital and labor, they effectively become identical. Suppose that city *A* has an initial advantage in the stock of capital i.e. $K^A(0) > K^B(0)$. However, everyone believes that city *B* will be the first one to grow, and thus all investors invest in that city first. Now consider the gain of an atomistic investor who decides to deviate and invest in city *A*. The net return of capital for him is strictly larger than for the rest of the investors until the two cities have accumulated the same stock of capital. The stock of capital increases in *B* very quickly and labor migrates from *A* to *B* following this investment pattern. Capital in city *A* accumulates very slowly because only the deviant agent is investing there. So, inevitably, at some point the two stocks of capital will be the same and hence population will also be identical in the two cities. The two cities are effectively identical from that point on. This implies that subsequent investment will be the same in the two cities and that the economy will be headed to its steady state. Let $\tilde{t}$ be the period in which the two cities become identical. In the time interval $\left(0, \tilde{t}\right]$ the deviant obtains a larger return on his investment than any of the other investors. After that, the returns are the same for everyone.

This shows that in this case, it is always optimal to deviate, and therefore this situation cannot be an equilibrium. Figures 6 and 7 show the pattern of the two interest rates and the two stocks of capital, respectively.

Figure 6: Interest Rates in the Inefficient Potential Equilibrium $\left(A^A = A^B\right)$

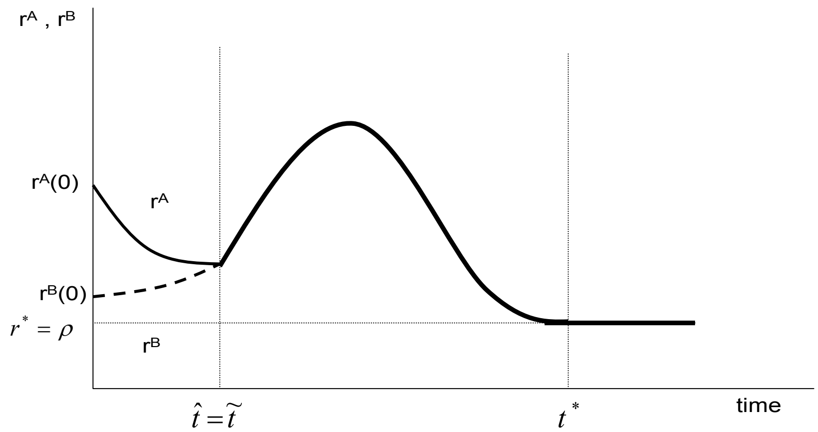
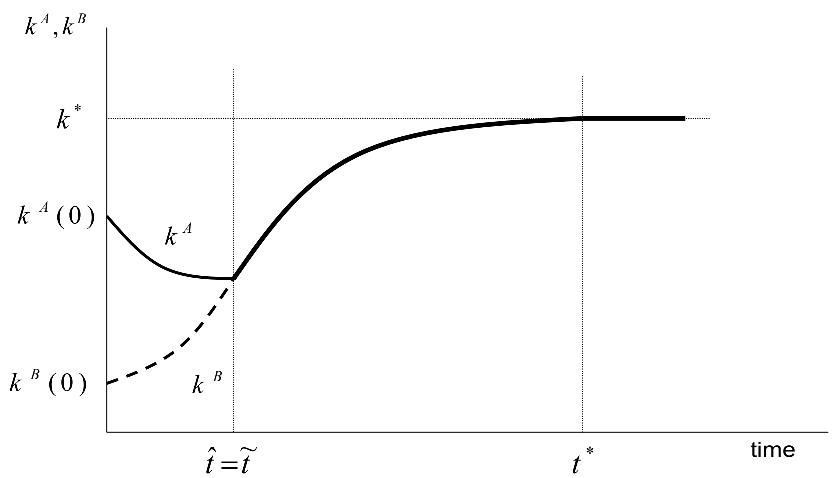


Figure 7: Capital in the Inefficient Potential Equilibrium ($A^A = A^B$)



Case 2: $A^A < A^B$

Assume now that the two *TFPs* A^A, A^B differ from each other ($A^A < A^B$) and that this is common knowledge among investors. As before, $K^A(0) > K^B(0)$.[27] Assume also that these investors prefer to invest in the city with the largest *TFP* even if it is the small one.[28] In this case, the two net *MPKs* will be equated when city *A* still has a larger stock of capital than city *B* (see Figures 8 and 9). In other words, the lower *TFP* in city *A* implies that this city needs more capital to be as productive as city *B*.

But note that after the two *MPKs* have been equated (at $\hat{t}$), investors keep investing in city *B* because they think it has the largest potential to grow. So at period $\tilde{t}$ there is complete convergence in the stocks of capital. As a result, the gross *MPK* is now strictly larger in city *B* because of the technology gap and, therefore, investment continues in city *B* until congestion costs there become critical (at period $\breve{t}$). At that point, investment in *B* stops until the gross *MPKs* in the two cities are equated again (at period $\hat{t}'$). Of course, the technological advantage if city *B* implies that the stock of capital in *B* never equals that of city *A* again. Finally, after period $\hat{t}'$ the two stocks of capital grow until they reach the two different steady states that satisfy $k^{*B} > k^{*A}$. Note that after period $\hat{t}'$ the two interest rates must always be equated. This is a non-arbitrage condition that has to be satisfied until the steady state is reached and $r^A = r^B = \rho$.

---

[27] The case $A^A > A^B$ is less interesting because city *A* would then have an advantage in both the initial stock of capital and *TFP*.

[28] One reason why they may want to follow this strategy is that they (correctly) anticipate that having a larger *TFP* implies a larger steady-state level of capital.

Figure 8: Interest Rates in the Inefficient Potential Equilibrium $\left(A^{A} > A^{B}\right)$

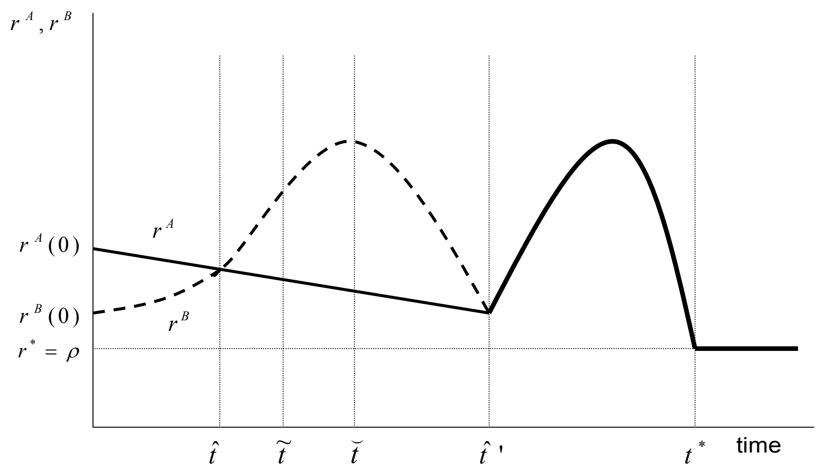
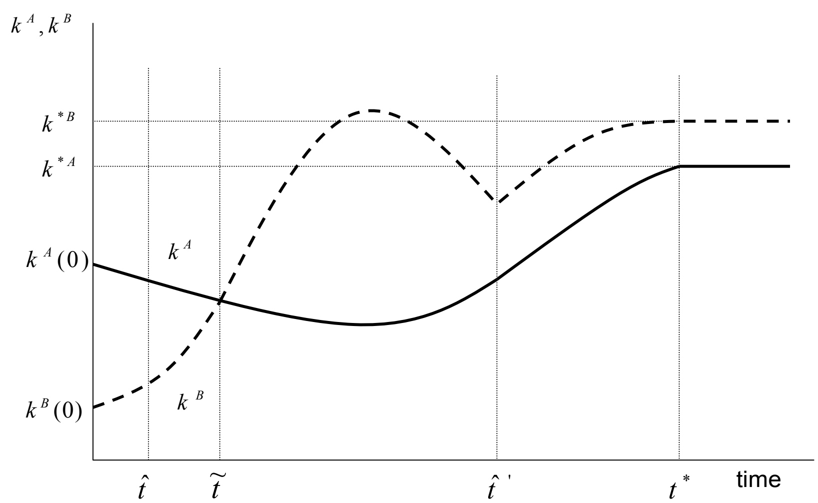


Figure 9: Capital in the Inefficient Potential Equilibrium$\left(A^{A} > A^{B}\right)$



In this situation it is unclear whether an investor will choose to deviate and start investing in *A* when the rest of the investors accumulate capital in *B*. In the time interval $(0,\hat{t}]$ this deviant would make a positive gain, while in $(\hat{t},\hat{t}']$ he would be losing money. The assumption one needs to make to ensure that any deviant would make a net positive gain is the following:

$$\int_0^{\hat{t}}\left(r^A(t)-r^B(t)\right)dt > \int_{\hat{t}}^{\hat{t}'}\left(r^B(t)-r^A(t)\right)dt$$

This condition states that the sum of the deviant gains in the time interval $(0,\hat{t}]$ must be strictly larger than the losses he would incur in $(\hat{t},\hat{t}']$. It is clear that $\hat{t}$ is a positive function of the initial gap $\varepsilon = k^A(0)-k^B(0)$. This is because a large gap implies that it would take more time to equate the two interest rates and hence it would take longer for city *B* to start growing. Moreover, $\hat{t}'-\hat{t}$ is independent of this initial gap. Therefore there must exist a critical gap ε^* such that the condition above is satisfied with equality. If one imposes the constraint $\varepsilon \geq \varepsilon^*$ any investor will gain from deviating and so this situation could not be an equilibrium.

An alternative way to rule out this equilibrium is to impose some behavioral assumptions on investors' beliefs. One would need to assume that all investors have beliefs that satisfy the following:

1. They have common initial beliefs about the magnitude of the parameters A^A and A^B.
2. These beliefs are a positive function of the initial stocks of capital in each city, $k^A(0)$ and $k^B(0)$, and these stocks are perfectly observable by all investors.
3. Although the relation between the beliefs and the initial stocks of capital may be noisy, the initial gap $\varepsilon = k^A(0)-k^B(0)$ is large enough to infer that $A^A > A^B$.

These beliefs would then imply that there is no reason why investors should start investing in the small city *B* instead of city *A*, and they ensure that a situation where all investors initially invest in *B* cannot be a rational equilibrium.

Ruling out equilibria by imposing a restriction on the initial conditions of the problem is a current practice in the literature of "history versus expectations." Krugman (1991) refers to this condition as "being out of the overlap region." Other related papers are Matsuyama (1991) and Fukao and Benabou (1993).

4. A Model with Multiple Cities

In this section I extend the model to $M > 2$ cities and show that its predictions do not change. I showed above that the planner's problem and the decentralized one are qualitatively similar. Therefore, to save space, I describe here the effect of extending the model to a larger number of cities using only the planner's setup. I start by presenting the model with three cities and then argue that the same logic applies when one allows for an arbitrarily large number of cities.

Suppose the economy consists of three cities: *A, B,* and *C*. The planner's problem in the first stage is now

$$\max_{N^A, N^B, N^C} \sum_{j=A,B,C} Y^j$$

$$\sum_{j=A,B,C} N^j = N$$
$$Y^j = \left(N^j\right)^\alpha \left(K^j\right)^\beta, \forall j = A,B,C$$
$$K^j \; given, \; , \forall j = A,B,C$$

From the first order conditions one gets

$$\left(N^A\right)^{\alpha-1}\left(K^A\right)^\beta = \left(N^B\right)^{\alpha-1}\left(K^B\right)^\beta = \left(N^C\right)^{\alpha-1}\left(K^C\right)^\beta$$

which implies that for any two cities j and j' one has:

$$\frac{N^j}{N^{j'}} = \left(\frac{K^j}{K^{j'}}\right)^{\frac{\beta}{1-\alpha}} \qquad (18)$$

As before, the economic intuition behind this condition is that increasing returns to scale $(\alpha + \beta > 1)$ implies that cities with a larger stock of capital also attract more population. Furthermore, increases in the (log) capital ratio of the two cities still lead to more than proportional increases in their (log) population ratio. I next define the following function:

$$F\left(K^A, K^B, K^C, N\right) = \sum_{j=A,B,C} Y^j\left(N^j, K^j\right) \qquad (19)$$

As in the two-city problem, this function represents the value of total output that results from inserting the expression for the optimal allocation of labor into total national production. Combining the first order conditions and the market clearing equation $N = \sum_{j=A,B,C} N^j$ it is easy to show that this function can be expressed as

$$F\left(k^A, k^B, k^C, N\right) = N^{\alpha+\beta}\Omega\left(k^A, k^B, k^C, N\right)^{1-\alpha}$$

where

$$\Omega\left(k^A, k^B, k^C, N\right)^{1-\alpha} \equiv \sum_{j=A,B,C}\left(k^j\right)^{\frac{\beta}{1-\alpha}}$$

It seems clear that there is nothing particular about $M = 3$, so the corresponding expression for an arbitrarily large number of cities $M > 3$ (in per capita terms) is

$$f\left(k^1,...,k^M, N\right) \equiv \frac{F\left(k^1,...,k^M, N\right)}{N} = N^{\alpha+\beta-1}\Omega\left(k^1,...,k^M\right)^{1-\alpha} \tag{20}$$

where

$$\Omega\left(k^1,...,k^M\right) \equiv \sum_{j=1}^{M}\left(k^j\right)^{\frac{\beta}{1-\alpha}} \tag{21}$$

Note this is the *M-city* version of equations (2) and (3). Having derived the function $f\left(k^1,...,k^M, N\right)$ the second-stage problem with *M* cities reads as

$$\max\int_0^\infty e^{-\rho t}\ln(c)dt$$

$$\sum_{j=1}^{M} i^j + \frac{1}{N}\sum_{j=1}^{M} g\left(K^j\right) + c = f\left(k^1,...,k^M, N\right)$$

$$\dot{k}^j = i^j - \delta k^j, \forall j = 1,...,M$$
$$i^j \geq 0, \forall j = 1,...,M$$
$$k_0^j\ given, \forall j = 1,..., M$$

The associated present-value Hamiltonian of this problem is

$$H = e^{-\rho t}\ln c + \lambda\left[f\left(k^A,...,k^M,N\right) - \sum_{j=1}^{M} i^j - \frac{1}{N}\sum_{j=1}^{M} g\left(K^j\right) - c\right] + \sum_{j=1}^{M}\mu^j\left(i^j - \delta k^j\right)$$

where $\lambda, \mu^j, j = 1,...,M$ are Lagrange multipliers. The first order conditions of this problem are the analog as those of the two-city case (equations 4-6).[29]

Proposition 1 also holds in this setup because nothing fundamental has been altered; increasing returns to scale and free labor mobility imply that the gross *MPK* is larger in the city with the largest stock of capital. Assumptions 1 and 2 now require that, at date zero, it is possible to rank the *M* cities with respect to their stock of capital and their net *MPK*. To simplify, I assume that the difference in capital stocks between any two cities is the same for all consecutively ranked cities. So if one lets cities 1 and *M* be the initially largest and smallest ones, respectively, this assumption reads:

Assumption 1''

$$K^1(0) > K^2(0) > ... > K^M(0)$$

and

$$K^j(0) - K^{j+1}(0) = \varepsilon, \forall j = 1,...,M$$

where ε is a small positive number.[30] Assumption 2 can be rewritten as

Assumption 2''

$$\tilde{f}_1(0) > \tilde{f}_2(0) > ... > \tilde{f}_M(0) > \rho + \delta$$

---

[29] The *MPK* in city *j* is now given by $f_j^p\left(k^1,...,k^M,N\right) \equiv \frac{\partial f\left(k^1,...,k^M,N\right)}{\partial k^j} = \beta N^{\alpha+\beta-1}\Omega\left(k^A,...,k^M,N\right)^{-\alpha}\left(k^j\right)^{\frac{\beta}{1-\alpha}-1}$. As before, note that this expression depends on the stock of capital in all cities.

[30] Note that if two cities have the same stock of capital on the initial date, the date-zero ranking of cities would not be perfect and, in equilibrium, these two cities will start growing at the same time and at the same rate. However, as long as there is enough dispersion in the initial distribution of capital stocks (and hence population) across cities -as the data clearly indicate- the main implications of the model would not change.

where, as before, $\tilde{f}_j(0)$ is the date-zero net *MPK* in city $j, j = 1,..., M$.

$$\tilde{f}_j(0) \equiv f_j^p(0) - g'(K_j(0))$$

Finally, the analogous of Assumption 3 is:

Assumption 3''

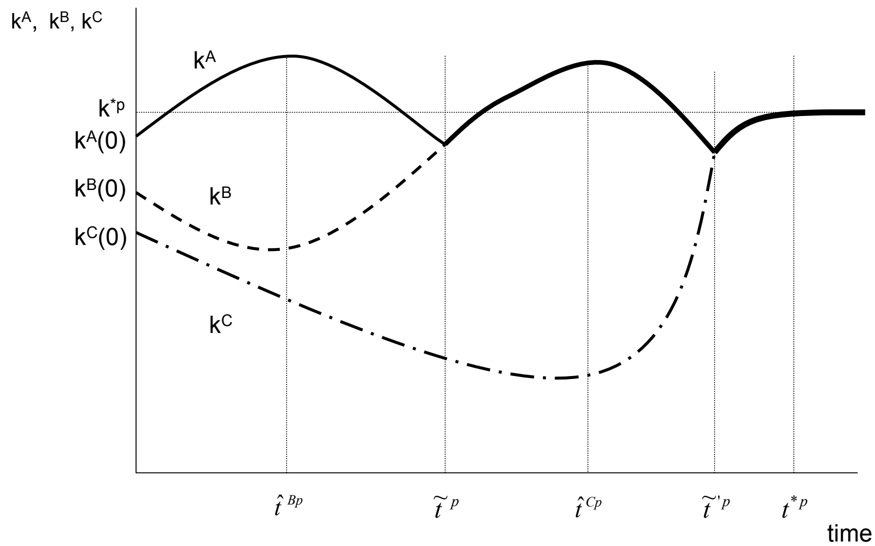
$$f_j(\hat{t}^{jp}) > \rho + \delta, \forall j > 1$$

where $\hat{t}^{jp}$ is the period at which city $j > 1$ starts its phase of solo growth.[31] The dynamics of the stock of capital in the three-city case are displayed in Figure 10. In the interval $(0, \hat{t}^{Bp}]$, the stock of capital of city *A* grows alone, while both cities *B* and *C* see their stock shrink (due to depreciation). This is the equilibrium path because during this time interval investment is most profitable in city *A*.[32]

[31] If the initial net *MPK* of a particular city does not satisfy this inequality, the model predicts that this city will never experience a period of solo growth and will eventually disappear. The data certainly shows cases of cities that have never experienced a rapid growth in population. Moreover, Henderson and Ioannides (1981) claim that it is not unusual to see sporadic deaths of towns in the data.

[32] Since the optimal allocation of labor implies $\frac{N^A}{N^B} = \left(\frac{K^A}{K^B}\right)^{\frac{\beta}{1-\alpha}}$ and $\frac{N^A}{N^C} = \left(\frac{K^A}{K^C}\right)^{\frac{\beta}{1-\alpha}}$, the model predicts migration flows from cities *B* and *C* to city *A*, with the latter flow being larger because the productivity gap between cities *A* and *C* is bigger than the one between cities *A* and *B*.

Figure 10: Capital with Three Cities



At period $\hat{t}^{Bp}$, city *A* becomes too congested and investment starts in the next best available location. Increasing returns imply that this location must be city *B*, the second-largest city. Capital keeps depreciating in city *C* (I assume that this stock remains positive throughout the transition of the economy to its steady state) and now it also depreciates in city *A*.

The stocks of capital in *A* and *B* are equated at period $\tilde{t}^{p}$. After that, capital increases at the same rate in cities *A* and *B* (they are identical cities at this point) until congestion costs in the two cities are too large again. That occurs at period $\hat{t}^{Cp}$. City *C* then attracts all the new investment, and at some point $\left(\tilde{t}^{'p}\right)$, the three cities end up with the same stock of capital. Finally, the steady state of capital is reached at period t^{*p}. As in the two-city case, the evolution of population is the same as that of the stock of capital with the only difference that after period $\tilde{t}^{'p}$, the incentives to migrate across cities disappear since they are identical in all respects. This of course means that the steady-state value of population in each city is $N^{*}=\dfrac{N}{3}$. Finally, it is straightforward to show that Propositions 2-4 and the corollary also hold in this setup since nothing fundamental has been changed.

5. Conclusions

This paper presents a dynamic growth model of optimal city size that offers a simple mechanism by which cities grow sequentially. The model is able to rationalize the three new empirical findings on city growth described in Cuberes (2008).

Cities are modeled as Cobb-Douglas firms with diminishing returns to each input -labor and capital- but increasing returns to scale in the two inputs. Congestion costs in the model are introduced as convex costs in the stock of capital installed in a given city, which is crucially assumed to be immobile. In this framework, the initially largest city experiences a solo growth until it reaches a critical size. After this happens, the initially second-largest city is the one that grows alone. Therefore, at any point in time, one city grows much faster than the rest and the rank of the fastest growing cities (with the largest city having rank one) increases as time goes by. Finally, it is shown that exogenous increases in population boost the gross *MPK* in all cities but more so in the largest ones. Therefore, the model also predicts that sequential city growth is faster in periods of rapid growth of the urban population.

I solve the planner's problem and the decentralized one and show that, since increasing returns are external to firms, investment is suboptimal and the transition to the steady state is too slow in the latter setup. Uniqueness of equilibrium in the market economy is discussed and, finally, I show that extending the model to an arbitrarily large number of cities does not change its predictions.

Appendix

1. Derivation of the Function $F\left(k^A, k^B, N\right)$

Define the following aggregate production function

$$F\left(K^A, K^B, N^A, N^B\right) \equiv Y^A\left(N^A, K^A\right) + Y^B\left(N^B, K^B\right)$$

From (1) one has

$$N^A = N^B \left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}$$

and so

$$F\left(k^A,k^B,N\right)=\left(N^B\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}\right)^{\alpha}\left(k^A N\right)^{\beta}+\left(N^B\right)^{\alpha}\left(k^B N\right)^{\beta}$$

Using the market-clearing condition $N = N^A + N^B$ and (1) leads to

$$N^A = N\left(\frac{\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}}{1+\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}}\right)$$

Let $n\left(k^A,k^B\right)\equiv\left(\frac{\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}}{1+\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}}\right)$. One can then write $N^A = n\left(k^A,k^B\right)N$ and $N^B = \left(1-n\left(k^A,k^B\right)\right)N$. Using the last two expressions and following simple algebraic steps it is possible to show that

$$F\left(k^A,k^B,N\right)=N^{\alpha+\beta}\Omega\left(k^A,k^B,N\right)^{1-\alpha}$$

where $\Omega\left(k^A,k^B,N\right)\equiv\sum_{j=A,B}\left(k^j\right)^{\frac{\beta}{1-\alpha}}$ which corresponds to equations (2) and (3) in the text.

2. The Decentralized MPKs

From equation (14) one has $f_j^m \equiv (1-\alpha)\left(N^j\right)^{\alpha}\left(K^j\right)^{\psi-\alpha}$

The optimal allocation of labor is given by $\frac{N^A}{N^B} = \left(\frac{k^A}{k^B}\right)^{\frac{1+\psi-\alpha}{1-\alpha}}$ and so for city A one has

$$f_A^m\left(k^A, k^B, N\right) \equiv N^{\psi-\alpha}(1-\alpha)\left(N^B\left(\frac{k^A}{k^B}\right)^{\frac{1+\psi-\alpha}{1-\alpha}}\right)^{\alpha}\left(k^A\right)^{\psi-\alpha}$$

where I used the fact that $K^A = k^A N$. Since $N = N^A + N^B$ one can then write $N^B = \frac{N}{1+\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}}$. This implies

$$f_A^m\left(k^A, k^B, N\right) \equiv (1-\alpha)N^{\psi-\alpha}\left(\frac{N}{1+\left(\frac{k^A}{k^B}\right)^{\frac{\beta}{1-\alpha}}}\left(\frac{k^A}{k^B}\right)^{\frac{1+\psi-\alpha}{1-\alpha}}\right)^{\alpha}\left(k^A\right)^{\psi-\alpha}$$

Simple algebraic manipulation leads to

$$f_A^m(k^A, k^B, N) = (1-\alpha)N^{\psi}\Omega\left(k^A, k^B, N\right)^{-\alpha}\left(k^A\right)^{\frac{\beta}{1-\alpha}-1}$$

where

$$\Omega\left(k^A, k^B, N\right) \equiv \sum_{j=A,B}\left(k^j\right)^{\frac{\beta}{1-\alpha}}$$

which corresponds to equations (16) and (17) in the text. Similar steps for city B lead to

$$f_B^m(k^A, k^B, N) = (1-\alpha)N^{\psi}\Omega\left(k^A, k^B, N\right)^{-\alpha}\left(k^B\right)^{\frac{\beta}{1-\alpha}-1}$$

References

Abdel-Rahman, Hesham M. (1990), "Agglomeration Economies, Types, and Sizes of Cities." *Journal of Urban Economics* 27, pp. 25-45.

Anas, Alex (1978), "Dynamics of Urban Residential Growth." *Journal of Urban Economics* 5, pp. 66-87.

Anas, Alex (1992), "On the Birth and Growth of Cities." *Regional Science and Urban Economics* 22, pp. 243-258.

Baier, Scott, Sean Mulholland, Robert Tamura, and Chad Turner (2008), "How Important Are Human Capital, Physical Capital and Total Factor Productivity for Determining State Economic Growth in the United States, 1840-2000?" working paper, Clemson University.

Bennathan, Esra, and David Canning (2000), "The Social Rate of Return on Infrastructure Investments." *Policy Research Working Paper Series 2390*, The World Bank.

Brezis, Elise S., and Paul R. Krugman, (1997) "Technology and the Life Cycle of Cities." *Journal of Economic Growth*, vol. 2(4), December, pp. 369-83.

Brueckner, Jan (2000), "Urban Growth Models with Durable Housing." in *Economics of Cities*, Jean-Marie Huriot and Jacques-François Thisse (eds), Cambridge University Press.

Caselli, Francesco, and James Feyrer (2007), "The Marginal Product of Capital." *Quarterly Journal of Economics*, vol. CXXII(2), May, pp. 535-68.

Chipman, John S. (1970), "External Economies of Scale and Competitive Equilibrium." *Quarterly Journal of Economics*, vol. 84, No. 3, August, pp. 347-85.

Cuberes, David (2008), "Sequential City Growth: Empirical Evidence." working paper, Clemson University.

Desmet, Klaus (2002), "A Simple Dynamic Model of Uneven Development and Overtaking." *Economic Journal*, 112(482), pp. 894-918.

Duranton, Gilles, and Diego Puga (2004), "Micro-foundations of Urban Agglomeration Economies." in *Handbook of Regional and Urban Economics*, J. Vernon Henderson and Jacques-François Thisse (eds.), vol. 4, Amsterdam: North-Holland, pp. 2063-117.

Fujita, Masahisa (1976), "Spatial Patterns of Urban Growth: Optimum and Market." *Journal of Urban Economics* 3, pp. 209-41.

Fujita, Masahisa (1978), *Spatial Development Planning*, Chapter 6 and Appendix F, Amsterdam: North Holland.

Fujita, Masahisa (1982), "Spatial Patterns of Residential Development." *Journal of Urban Economics* 12, pp. 22-52.

Fukao, Kyoji, and Roland J. Benabou (1993). "History versus Expectations: A Comment." *Quarterly Journal of Economics*, vol. 108(2), May, pp. 535-42.

Helsley, Robert W., and William C. Strange (1990), "City Formation with Commitment." *Regional Science and Urban Economics*, 24, pp. 373-90.

Henderson, J. Vernon (1974), "The Sizes and Types of Cities." *American Economic Review*, LXIV, pp. 640-56.

Henderson, J. Vernon, and Yannis M. Ioannides (1981), "Aspects of Growth in a System of Cities." *Journal of Urban Economics* 10, pp. 117-39.

Henderson, J. Vernon, and Anthony J. Venables (2009), "Dynamics of City Formation." *Review of Economic Dynamics*, vol. 12(2), April, pp. 233-54.

Henderson, J. Vernon, and Hyoung. G. Wang (2007), "Urbanization and City Growth: The Role of Institutions." *Regional Science and Urban Economics*, vol. 37(3), May, pp. 283-313.

Ioannides, Yannis M. (1994), "Product Differentiation and Endogenous Growth in a System of Cities." *Regional Science and Urban Economics*, 24, pp. 461-84.

Ioannides, Yannis M. (2008), "Intercity Trade and Convergent versus Divergent Urban Growth." working paper, Tufts University.

Kanemoto, Yoshitsugu (1980), *Theories of Urban Externalities*, North-Holland, Amsterdam.

Koopmans, Tjalling C. (1957), *Three Essays on the State of Economic Science*, New York: McGrawHill.

Krugman, Paul R. (1981), "Trade, Accumulation, and Uneven Development." *Journal of Development Economics*, vol. 8, pp. 149-61.

Krugman, Paul R. (1991), "History Versus Expectations", *Quarterly Journal of Economics*, vol. 106(2), May, pp. 651-67.

Krugman, Paul R. (1992), "A Dynamic Spatial Model." *NBER working paper 4219*.

Lucas, Robert E. Jr. (1990), "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review*, LXXX, pp. 92-6.

Matsuyama, Kiminori (1991), "Increasing Returns, Industrialization, and Indeterminancy of Equilibrium." *Quarterly Journal of Economics*, vol. 106(2), May, pp. 617-50.

Mills, Edwin S. (1967), "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review Papers and Proceedings*, 57(2), pp. 197-210.

Mirrlees, James A. (1972), "The Optimum Town." *Swedish Journal of Economics*, 74(1), pp. 114-35.

Miyao, Takahiro (1981), *Dynamic Analysis of the Urban Economy*, Academic Press, New York, NY.

Palivos, Theodore, and Ping Wang (1996), "Spatial Agglomeration and Endogenous Growth." *Regional Science and Urban Economics* 26, pp. 645-69.

Romer, Paul M. (1986), "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94, pp. 1002-37.

Rossi-Hansberg, Esteban, and Mark L. J. Wright (2007), "Urban Structure and Growth." *Review of Economic Studies*, 74, pp. 597-624.