

MPRA

Munich Personal RePEc Archive

A Model of Sequential City Growth

Cuberes, David

Clemson University

16 February 2007

Online at <https://mpra.ub.uni-muenchen.de/2172/>

MPRA Paper No. 2172, posted 09 Mar 2007 UTC

A Model of Sequential City Growth

David Cuberes*

The John E. Walker Department of Economics
Clemson University

February 26, 2007

Abstract

There is strong evidence showing that in most countries cities develop sequentially, with the initially largest city being the first to grow. This paper presents a growth model of optimal city size that rationalizes this particular growth pattern. Increasing returns to scale is 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 generate sequential growth is the assumption of irreversible investment in physical capital. The presence of a positive external effect of aggregate city capital on individual firms makes the competitive equilibrium inefficient.

1 Introduction

In a recent paper, Cuberes (2006) presents extensive evidence on the fact that cities grow sequentially in most countries during the 1800-2000 time interval. He first documents that the concentration of population in the largest cities of most countries follows a very clear inverse U-pattern over time. As countries develop, the fraction of their population who locates in the largest cities increases for some years and, after reaching a maximum, it continuously declines. This is shown in Figure 1 of Appendix D for the case of the U.S. Second, he presents evidence that in most decades and countries the cross-section distribution of cities growth rates is clearly skewed to the right, indicating that a few cities grow much faster than the rest. Finally, the rank of cities that grow the fastest

*I would like to thank Roc Armenter, Scott Baier, Gary Becker, Oded Galor, Edward Glaeser, François Gourio, Vernon Henderson, Michal Jerzmanowski, Aitor Lacuesta-Gabarain, Casey Mulligan, Rafael Repullo, Balazs Szentes, Robert Tamura, Marc Teignier, 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, Universitat Pompeu Fabra, the 2005 North American Regional Science Council, the 2006 Summer Econometric Society, and the 2006 Southern Economic Association. All remaining errors are mine. E-mail: cuberes@clemson.edu

in each period increases over time in most of the cases (see Figure 2). These three observations imply that cities grow sequentially and that the initially largest cities are always the ones to start growing.

These facts are hard to explain using Gibrat's law¹, which states that the growth processes of cities follow a random walk. This law has been the cornerstone of a large part of the empirical and theoretical literature on cities, summarized in the next section. The goal of this paper is to develop a theoretical model that departs from Gibrat's law and can easily explain this sequential city growth.

I develop a general equilibrium growth model with two cities, modeled as Cobb-Douglas firms. These firms present increasing returns to scale but diminishing returns to each input, capital and labor. On the other hand, there are convex congestion costs associated with the stocks of capital installed in each city. Investment in a city is irreversible. In this setup, increasing returns is the force that enhances the concentration of resources, while congestion costs limits it.² To save space, only the competitive problem is solved in the main text. The problem of a benevolent social planner is then presented in Appendix B. The main difference between both setups is that the presence of a positive external effect of aggregate city capital on firms operating in that city, makes the competitive equilibrium suboptimal.

The model is able to generate sequential city growth. The initially large city grows alone for some periods. Eventually, congestion costs in that city become too large, and investment starts in the second largest city. Since these costs are fairly small in this second city, its investment rate is strictly larger than that of the initially largest one. Depending on the parameters of the model, the initially smaller city may completely catch up with the large one in terms of both labor and capital, in which case, both cities may grow at the same rate during some periods. The model predicts that, at each point in time, one city grows much faster than the rest, hence generating a distribution of growth rates that is skewed to the right. Moreover, the city that has the largest growth rate at a given point in time is the largest one, conditional to the fact that congestion costs have not yet reached a critical level. Therefore the rank of the fastest growing city increases over time, as it occurs in the data.

The paper is organized as follows. In Section 2, I summarize the existing literature on city size distribution and city growth. Section 3 presents the theoretical model and Section 4 concludes.

¹See Gibrat (1931) for an initial statement of the law and Sutton (1997) for some of its applications in economics.

²Congestion costs can come from consumer's disutility associated with crowded cities or from production costs caused by large stocks of accumulated capital. This paper follows the second approach.

2 Related Literature

The existing literature on the evolution of city size distribution has mainly focused on explaining Zipf's law³, and on analyzing steady state properties of a system of cities, paying little attention to other interesting aspects of their growth processes. In particular, not much research has been devoted to the transitional dynamics of a system of cities towards its steady state.⁴ Some of these empirical and theoretical papers are summarized below.⁵

Many authors have shown that Zipf's law is a good representation of the distribution of city sizes for most countries in different years. As noted in Lai (2001), the most successful theoretical models that generate this law are based on the assumption that the growth rate of a city is independent of its initial size, i.e. Gibrat's law. The link between these two laws was first established in Gabaix (1999a), where it is formally proved that Gibrat's law implies Zipf's law. Cordoba (2003) examines in detail the theoretical conditions needed for Zipf's law to hold, and concludes that "Gibrat's law is not *an* explanation [...] but it is *the* explanation". According to this study, Gibrat's law is a sufficient but also necessary condition for Zipf's law. However, from a purely theoretical point of view, it has often been argued that this assumption implies the absence of economies of scale, something that has been a fundamental building block of theories of urban development.⁶ A second, and perhaps stronger criticism to those models based on a strict version of Gibrat's law is that they do not satisfactorily explain the transition of cities to their steady state distribution.

A different strand of the literature has developed models that predict Zipf's law without assuming Gibrat's law or allowing for important deviations from it. Duranton (2002) presents a quality-ladder model of growth where city sizes evolve following industries' technological innovations. His model does not match Gibrat's law but, under some circumstances, it verifies Zipf's law. Axtel and Florida (2001) present a hybrid model of an urban system whose steady state is characterized by Zipf's law. In their model, firm formation is the mechanism which determines city formation. Firms' growth rates are Laplace-distributed

³This law describes a striking empirical fact in urban economics. It states that, the size of the largest city in a country is roughly twice the size of the second largest city, three times the size of the third one, and so on. See Zipf (1949), Rosen and Resnick (1981), Soo (2002), Brackman et al. (1999) and Axtel and Florida (2001) among others.

⁴In many cases, the main purpose of the studies has been to develop theoretical models that are able to generate Gibrat's law instead of testing it. An interesting aspect that has been ignored and is treated in Cuberes (2005) is the behavior of cities with declining population. The theoretical model developed here predicts a *smooth* decline of population in those cities, which is actually true in the data. This represents an important improvement relative to most of the existing models that study the evolution of city size distribution. In those models, once a city reaches a critical size, its population suddenly experiences a huge decrease.

⁵For a comprehensive review of the literature on city size distribution, see Gabaix and Ioannides (2004).

⁶Rossi-Hansberg and Wright (2003) show that this criticism does not always apply. In their model, it is possible to have economies of scale and obtain an equilibrium distribution of city size based on Gibrat's law that is compatible with Zipf's law.

and city sizes obey Zipf’s law. Finally, different papers have analyzed the distribution of city sizes using models based on economic geography. Some examples are Brakman et al. (2001), Fujita et al. (1999), Krugman (1996b) and Davis and Weinstein (2001).

Henderson (1974, 1988) are the first theoretical contributions to the literature on optimal city size. However, little further theoretical or empirical research on this concept has been developed since these seminal papers.⁷ The explanations I find for this are two-fold. First, most of the empirical work has only looked at a few countries and used very limited datasets, leading to controversial conclusions in favor of Gibrat’s law, a principle that is at odds with the idea of an optimal city size. Moreover, the methods used in such studies may have been misleading, as it is argued in Cuberes (2006). Second, in the theoretical arena, the success of the endogenous growth literature has diverted attention towards models that predict balanced growth paths in which each city grows at the same rate.⁸

Other papers have focused on trying to test the empirical validity of Gibrat’s law for cities. Henderson and Wang (2003) analyze the urbanization process⁹ for a set of countries in the period between 1960 and 2000. They present some evidence of the absence of unit roots¹⁰, which directly contradicts Gibrat’s law for cities. Ioannides-Overman (2003) show that deviations from Gibrat’s law are not statistically significant for the main U.S. metropolitan areas in the period 1900-1990. Dobkins and Ioannides (1998a) present a model of human capital accumulation and test it using data from the U.S. census on MAs from 1900 to 2000. They find that U.S. cities are characterized by divergent growth if one ignores spatial evolution, and by convergent growth in the presence of regional effects. Eeckhout (2004) shows that if one considers the whole support of U.S. cities instead of their truncated distribution -the one that selects the N largest cities- Gibrat’s law implies a lognormal distribution of cities, not a power law.¹¹

Different studies have analyzed the growth processes of cities without directly dealing with Gibrat’s law. Glaeser, Scheinkman and Shleifer (1995) look

⁷Somewhat related, Abdel-Rahman and Anas (2004) and Carlino and Chatterjee (1999) have studied the behavior of systems of cities.

⁸See, for instance, Black and Henderson (1999).

⁹In Cuberes (2005) I show that, while rural-urban migration is a key phenomenon in most countries during the analyzed time interval, it cannot, by itself, account for the inverse U-shaped pattern. A second process, namely the existing competition among cities to attract resources (or urban primacy), is also in place.

¹⁰Gibrat’s law can be formulated as the presence of a unit root in the growth process of cities, i.e. $\alpha = 1$ in the equation

$$\ln S_{it} = \alpha \ln S_{it-1} + \varepsilon_{it}$$

where S_{it} is the size of city i in period t and ε_{it} is a white noise. Cuberes (2005) provides different estimates for the United States and confirm the absence of unit roots.

¹¹Zipf’s law is a particular case of a power distribution, namely the one with parameter equal to one.

at the relation between certain urban characteristics in 1960 and urban growth of income and population during the period 1960-1990. They show that there is a negative relationship between growth rates of city population and their initial size, indicating the existence of regression to the mean in city size, i.e. a violation of Gibrat's law. However, in many cases, this relationship turns out to not be significant after controlling for different regressors such as initial schooling, initial employment, and initial share of population in the manufacturing sector. Eaton and Eckstein (1997) show that the relative populations of the largest 40 metropolitan areas of France and Japan remained constant during the time interval 1876-1990 for France and 1925-1985 in the case of Japan. They claim that this may suggest that urbanization has taken the form of parallel growth of cities (Gibrat's law), rather than convergence (regression to the mean) to an optimal city size or divergent growth (regression away from the mean). Glaeser and Shapiro (2001) analyze the dynamics of growth rates of U.S. cities during the 1990s, and conclude that, although this decade was an unusually good one for the largest American cities and, in particular, for the cities of the Midwest, urban growth looked extremely similar to the one experienced during the prior post-war decades. Finally, Williamson (1965) and Hansen (1990) briefly document the existence of an inverse U-shaped pattern in population concentration, somewhat related to the one described in Cuberes (2006).

My paper relates most closely to Henderson and Venables (2004).¹² They analyze city formation in an economy that experiences steady growth of urban population over time. The presence of immobile housing and urban infrastructure and the fact that agents are not myopic, generates a sequential pattern of formation of cities and predicts a smooth decay of cities, i.e. it avoids the population swings present in most of the previous literature. The paper also analyzes how different institutions affect the equilibrium city size, which may or may not be optimal.

My model has two advantages with respect to Henderson and Venables (2004). First, it is easier to interpret. The dynamics of the model are generated by savings, as in the well-understood neoclassical growth model, whereas in their case households have linear utility functions, and hence zero savings. The second advantage is that I formally prove that, under my assumptions, the equilibrium growth path of cities is sequential, with the smaller cities growing later than the bigger ones. In their model, sequentiality is just a conjecture that seems to be verified. Moreover, I provide a proof of existence and uniqueness of equilibrium. Another important difference between the two models is that theirs deals with the creation of cities when urban population grows, whereas mine explains city growth once cities have already been created, taking population as constant.

¹²Krugman and Brezis (1996) is also close in spirit to my paper. They develop a model in which cities grow sequentially following technological innovations.

3 The Model

This section presents a theoretical framework to explain the sequentiality in city development described above. For simplicity, the model consists of just two cities that are modeled as firms with identical Cobb-Douglas production functions.¹³ In each city, production takes place using two inputs: capital and labor. The technology of each firm in a given city exhibits diminishing returns to each input but increasing returns to scale. On the other hand, each city has congestion costs that are modeled as convex costs associated with the stock of installed capital in that city. It is assumed that, at the initial period, one of the two cities has a slightly larger stock of capital than the other one. Investment in the model is irreversible.¹⁴ On the other hand, labor can migrate across cities at no cost.

In this setup, if congestion costs at low levels of capital are low enough, resources are disproportionately allocated in the city with an initial larger stock of capital until those costs make the initially smaller city a more attractive place to invest. At that point, since congestion costs are much lower in the smaller city, the investment rate becomes strictly larger in that city than in the largest one, until the stock of capital in both cities is equated. It will be shown that depending on the parameters of the model, this complete catch-up scenario may never take place.

This section presents the decentralized equilibrium. In Appendix B it is shown how this problem differs from the one solved by a benevolent social planner. The presence of an external effect of average city capital on each firms' productivity makes the market solution inefficient since there is underinvestment in both cities, and inequality in city size is too large.

3.1 The Decentralized Economy

I proceed by describing the behavior of households and firms in the economy and then calculate the competitive equilibrium, which exists and is unique, as I show in Appendix B.

3.1.1 Households

Households' income sources are wage earnings and the return on assets. The competitive wage rate is ω . Let Z^j be the amount of assets invested in city j , $j = A, B$. The capital income of a household who holds some positive amount of assets in both cities is given by $\sum_{j=A,B} r^j Z^j$, where r^j represents the gross return

¹³An extension of this model to one with $J > 2$ cities is presented in Appendix C. All the results proved here still hold in the general case.

¹⁴This is a reasonable assumption if one interprets physical capital as infrastructure (buildings, marketplaces, urban structure, etc...) installed in a city.

to capital in city j . These resources can be used to consume and accumulate assets. The budget constraint of a representative household in per capita terms is therefore

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

where $z^j \equiv \frac{Z^j}{N}$ and $c \equiv \frac{C}{N}$ represent the per capita holdings of asset j , $j = A, B$ and per capita consumption, respectively. Assuming logarithmic utility, the problem of a representative household in per capita terms (omitting time subscripts) can then be written as:

$$\begin{aligned} & \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\ & \sum_{j=A,B} \dot{z}^j + c = \omega + \sum_{j=A,B} r^j z^j \\ & \dot{z}^j \geq 0, \forall j = A, B \\ & z_0^j, \forall j = A, B \text{ given} \end{aligned}$$

where c represents per capita consumption and, as stated above, households face the irreversibility constraints $\dot{z}^j \geq 0$, $j = A, B$.

3.1.2 Firms

Each firm faces a constant returns to scale technology. This ensures that profits are zero and that one can consider a representative firm in each city. Each firm is subject to a positive external effect similar to the one in Romer (1986), coming from the average amount of capital installed in the city where the firm operates. The production function of firm i located in city j is given by

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

where Y^{ij} , N^{ij} , and K^{ij} respectively represent production, labor, and capital in city i located in firm j . The parameter ψ is strictly positive, and \tilde{K}^j represents the average capital installed in city j , i.e. $\tilde{K}^j \equiv \frac{1}{M} \sum_{m=1}^M K^{mj}$, where M is a large number that indicates the number of firms operating in city j . As in Romer's paper, the assumption that firms treat \tilde{K}^j as given could be rationalized in a model with a continuum of agents.¹⁵

¹⁵See Lucas (1988) and Tamura (1991) for the important role of average human capital on growth.

Firms hire labor and capital at their competitive prices and sell their product in the market. Moreover, they have to pay a congestion cost¹⁶ $g(K^{ij})$ for each unit of capital they install in the city where they operate, where $g(K^{ij}) = (K^{ij})^\sigma$, and it is assumed¹⁷ that $\sigma > \frac{2\beta}{1-\alpha}$, where $\beta \equiv \psi - \alpha + 1$. Normalizing the price of the consumption good to one, profits for firm i , located in city j , are then given by

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

where $\delta \in (0, 1)$ is the rate at which capital depreciates.

The first order conditions of the problem faced by this firm are:

$$\frac{\partial \pi^{ij}}{\partial N^{ij}} = 0 \Leftrightarrow \alpha (N^{ij})^{\alpha-1} (K^{ij})^{1-\alpha+\psi} = \omega \quad (1)$$

$$\frac{\partial \pi^{ij}}{\partial K^{ij}} = 0 \Leftrightarrow (1-\alpha)(N^{ij})^\alpha (K^{ij})^{\psi-\alpha} = r^j + \delta + g'(K^{ij}) \quad (2)$$

where the symmetric equilibrium condition $K^{ij} = K^j, \forall i$ has been imposed. Since labor can costlessly move across cities, the wage rate is the same in both cities in equilibrium. This implies that the optimal allocation of labor is given by

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

where $k^j \equiv \frac{K^j}{N^j}$, $j = A, B$ represents per capita capital in city j . This in turn implies the following gross marginal product of capital (i.e. without considering the congestion costs) to the competitive firm in city j :

$$f_j = (1-\alpha)\Omega^{-\alpha}(k^j)^{\frac{\beta}{1-\alpha}-1} \quad (3)$$

where total population has been normalized to one and $\Omega \equiv \sum_{j=A,B} (k^j)^{\frac{\beta}{1-\alpha}}$.

The first order condition with respect to capital for a representative firm in city j can then be written as

$$\tilde{f}_j = r^j + \delta \quad (4)$$

where $\tilde{f}_j \equiv f_j - g'(K^{ij})$ is the *net* marginal product of capital in city j .

¹⁶In Appendix A I provide some microfoundations for these production and cost functions.

¹⁷As it is shown in Appendix B, this parameter restriction is sufficient to guarantee the existence and uniqueness of a solution to the problem. This restriction states that congestion costs have to be large enough relative to increasing returns to ensure that the problem is well-behaved.

Equilibrium

Since the economy is closed and there is no government, the only available asset is physical capital, i.e. $z^j = k^j, \forall j = A, B$. The following proposition shows that, due to the initial difference in stocks of installed capital and the presence of increasing returns to scale, the gross marginal product of capital is always strictly larger in the city with a larger stock of installed capital.

Proposition 1 $f_A > f_B, \forall k^A > k^B > 0$

Proof. See Appendix B. ■

Assumptions 1 and 2 set up some initial conditions to make the problem interesting. Assumption 1 establishes that, at the initial date, city A has a slightly larger stock of installed capital than city B .

Assumption 1

$$K_0^A = K_0^B + \epsilon$$

where ϵ is a small positive number.

Assumption 2 states that, at date zero, congestion costs in the city with the largest stock of capital are relatively small compared to the productivity gains associated with its larger size. This ensures that investment in city A is strictly more profitable than in city B in the initial date. Moreover, it is assumed that investment is profitable in the smallest location, city B , i.e. its net marginal product of capital is strictly larger than the discount rate plus depreciation rate.

Assumption 2

$$\tilde{f}_{A0} > \tilde{f}_{B0} > \rho + \delta$$

With these assumptions, the unique competitive equilibrium in this economy has the following features. Households start investing only in the city with initially largest stock of capital, city A , since it is the most profitable investment opportunity available. However, as the stock of installed capital in city A increases, congestion costs rise quickly so that eventually investment takes place also in city B . Moreover, to make the problem interesting, one needs to assume that at period \hat{t} , when investment starts in city B , investment in that city is still profitable i.e.

Assumption 3

$$\tilde{f}_{B\hat{t}} > \rho + \delta$$

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

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

Proof. See Appendix B. ■

Depending on the parameter values of the model, it is possible that the city with a lower initial stock of capital completely converges to the largest one before the economy reaches the steady state. I call this the *catch-up case*. If convergence does not occur before the net marginal product of capital equals the sum of intertemporal discount rate and the depreciation rate, the economy is in the *non catch-up* scenario, and cities have a non degenerate distribution in the steady state.¹⁸

The Catch-Up Case

This is a situation in which the initial gap in the stocks of capital vanishes out before the economy reaches its steady state. Define period \bar{t} as the period at which that occurs. The condition needed for this scenario is $\tilde{f}_A(\bar{t}) = \tilde{f}_B(\bar{t}) > \rho + \delta$, i.e. the net marginal product of capital converges before investment becomes unprofitable.

The likelihood of having catch-up depends on three sets of parameters: first, it depends negatively on the size of this initial gap. Second, the larger ρ i.e. the more impatient individuals are, the faster will the steady state be reached, and, for a given initial gap, the less likely is that such gap vanishes. Finally, the production parameters α, β and the cost parameter σ also affect the probability of convergence. Intuitively, the larger the degree of increasing returns $\alpha + \beta$, the larger the fraction of investment going to city A , making the catch-up case less likely. The opposite is true when one considers the congestion parameter σ : if congestion costs are very convex, investment in the initially largest city will stop earlier, and hence the likelihood of convergence must, ceteris paribus, increase.

Define t^* to be the period at which $\tilde{f}_j = \rho + \delta, \forall j = A, B$, i.e. the period at which the economy reaches its steady state. The three periods \hat{t}, \bar{t}, t^* are well-defined because, unlike in the neoclassical growth model, convergence to a given

¹⁸A casual look at the distribution of city size in different countries suggests that the non catch-up case is the empirically relevant one. Under the convergence scenario, one would predict that in developed countries (which, according to the model should be closer to their steady state distribution) many of their largest cities should have identical size, which is clearly counterfactual.

stock of capital is reached in finite time. This is so because the Inada condition $\lim_{k^j \rightarrow \infty} \tilde{f}_j = -\infty$, ensures that the growth rate of the economy decreases fast enough. The next corollary summarizes the optimal investment policy in the catch-up case:

Corollary 1 *Optimal investment policy:*

$$\begin{cases} i^A > i^B, \forall t \in (0, \hat{t}] \\ i^A < i^B, \forall t \in (\hat{t}, \bar{t}] \\ i^A = i^B, \forall t > \bar{t} \end{cases}$$

The Non Catch-Up Case

Under this scenario, the initial gap in the stocks of capital does not disappear before the economy reaches its steady state. Corollary 2 summarizes the optimal investment policy in this case:

Corollary 2 *Optimal investment policy:*

$$\begin{cases} i^A > i^B, \forall t \in (0, \hat{t}] \\ i^A < i^B, \forall t \in (\hat{t}, t^*) \\ i^A = i^B, \forall t \geq t^* \end{cases}$$

The evolution of the stocks of per capita capital in each city in the catch-up case is displayed in Figure 3 in Appendix D. As assumed above, the initial stock of per capita capital is larger in city A at period zero. During the time interval $(0, \hat{t}]$ the stock of per capita capital increases in city A , and declines in city B as a result of the constant depreciation. Once the net marginal product of capital is the same in both cities, at period \hat{t} , the stock of per capita capital grows faster in city B than in city A . This is the case until city B catches up with city A . At that point, both stocks grow at the same rate and, finally, none of them grows once the steady state is reached, at period t^* . Figure 4 shows the evolution of population shares under this scenario. The share of population in city A increases during the time interval $(0, \hat{t}]$, since investment only occurs in that city. After period \hat{t} , A 's share decreases since investment becomes larger in city B . In the time interval $(\bar{t}, t^*]$, since the investment pattern is identical in both cities, and the stock of per capita capital is the same, population shares are equal to each other. Population in city B mirrors the behavior of that in city A .

Figures 5 and 6 display the non catch-up case. The only difference with respect to the previous scenario is that now, the level of per capita capital and the population share of city B never catches-up completely with those of the city with an initially larger stock of capital, city A .

4 Concluding Remarks

In this paper I present a model that explains why and how cities grow sequentially. Additionally, the theory is able to reproduce the inverse U-shaped pattern in population of the largest cities of a country over time and the right-skewness of cities growth rates at any point in time.

The model is one of optimal city size and contains the main elements necessary to study cities as economic entities. For simplicity, cities are modeled as Cobb-Douglas firms with increasing returns to scale but diminishing returns to each input, labor and capital. Congestion costs in the model are introduced as convex costs in the stock of installed capital. In this framework, each 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. Depending on the parameters of the model, the city that starts with a low level of capital completely catches up with the other one. Finally, increasing returns are external to private firms, and this fact generates two inefficiencies: too low of a growth rate of per capita consumption and steady state level of per capita capital and a variance in city sizes that is too large in the sense that these cities generate excessive congestion costs.

Appendix A

5 Microfoundations

5.1 A Model of Sharing

5.1.1 Indivisible Goods

Following Duranton and Puga (2003), one reason why cities exhibit increasing returns is that they allow consumers to share indivisible goods and facilities.¹⁹ This "large indivisibility" argument motivates urban increasing returns by directly assuming increasing returns at the aggregate level. In my model, the variable \tilde{K}^j may represent the average amount of installed indivisible goods in city j . At the same time, as the size of the community using these indivisible good grows, the facility will be subject to increasing crowding. Crowding in the model comes from the fact that, as the population of the city grows, new facilities need to be located somewhere and, in a standard monocentric linear model, this implies that they will be located further away from the city center. In this setup, it is reasonable to justify that private firms have to pay for these congestion costs. As the number of indivisible amenities grows, consumers start moving from the central business district (where all firms are located) to the suburbs and so, in order to sell their products, firms have to pay transport costs. this is the case since the irreversibility in investment precludes them from changing their location. Large indivisibilities in the provision of some public good are just one possible motivation for this. A common alternative is to assume large indivisibilities in some production activity. There is indeed a long tradition of modelling cities as the outcome of large indivisibilities in production (Koopmans (1957), Mills (1967), Mirrlees (1972)).

5.1.2 Common Marketplaces

One could also interpret K^{ij} for a given firm as infrastructure capital (for instance, buildings and machinery needed to produce). In this case, private firms benefit from the average infrastructure in a city, which is related to what in the literature is called *common marketplaces*- each firm benefits from being close to other firms because they can share a larger labor market. Alternatively, if one introduces intermediate goods needed for production (for instance, using a Dixit-Stiglitz production function), locating nearby other cities reduces the cost of acquiring these inputs. Consider again a monocentric city model where now it is assumed that consumers have a fixed location at the central business district (as in Wang (1990)). As the stock of urban structure grows, additional investments have to locate further away from the city business district and so firms have to pay higher transport costs to sell their products to consumers.

¹⁹A classical example is an ice hockey rink, which represents a facility that can be shared by many users.

Some other models on these lines are Berliant and Wang (1993), Wang (1993), Berliant and Konishi (2000), and Konishi (2000).

5.2 A Model of Land Rents

Consider a production function for firm i in location j that uses three inputs: capital (K^{ij}), labor (N^{ij}), and land (L^{ij}). Suppose there are constant returns to scale in K^{ij} and N^{ij} but diminishing returns to both inputs. Additionally, the supply of L^{ij} is limited. This can be modeled as a supply for land that becomes increasingly inelastic as the stock of capital and population grow (the simplest example is a Leontieff production function that requires a minimum amount of land to produce output). In this model, congestion costs naturally arise from the cost of land rents. An initially large city will grow very fast but, eventually, as it runs out of land (or the price of land becomes prohibitive) investment shifts to the second largest city, and so on, as in my model. Such a framework is qualitatively similar to one where land is omitted and we consider only the reduced form of costs of installation of capital.²⁰ Another closely related model would be one in which it is assumed that there is a TFP component that is initially increasing in K (or Y , i.e. K and L) and eventually becomes a decreasing function.

5.3 Knowledge Spillovers

Another possibility is to interpret K in my model as a combination of immobile capital structure and human capital. This formulation would have the advantage that is straightforward to justify the presence of increasing returns due to knowledge spillovers. On the other hand, investment would only be irreversible for the fraction of capital goods that represent infrastructure. Human capital (embodied to people) will freely move across cities. Congestion costs could be modeled as land rents or transport costs that are directly linked to the total stock of physical capital installed in a city, as in the models sketched above.²¹

²⁰ Arnott and Stiglitz (1979) is a good example of a model based on land rents.

²¹ Tamura (2002, 2006) analyzes the role of human capital in generating increasing returns in production.

Appendix B

The Social Planner's Problem

The problem of a benevolent social planner is to allocate labor and capital across cities and over time so that individuals' utility is maximized. To simplify notation, I will omit time subscripts in most of the remaining expressions:

$$\begin{aligned} & \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\ & \sum_{j=A,B} N^j = N \\ & \sum_{j=A,B} I^j + \sum_{j=A,B} g(K^j) + C = \sum_{j=A,B} Y^j \\ & \dot{K}^j = I^j - \delta K^j, \forall j = A, B \\ & I^j \geq 0, \forall j = A, B \\ & K_0^j \text{ given}, \forall j = A, B \end{aligned}$$

where $Y^j = (N^j)^\alpha (K^j)^\beta$, $0 < \alpha < 1$, $0 < \beta < 1$, and $\alpha + \beta > 1$, for any $j = A, B$. The constant N represents total national population, I^j and K_0^j respectively denote the investment rate and the initial stock of capital installed in city j , $j = A, B$, and $\delta \in (0, 1)$ is the depreciation rate of capital. In order to solve this problem, I divide it in two stages.

Stage 1: Optimal Allocation of Labor

In this stage, taken as given the levels of installed capital, the social planner chooses the optimal amount of labor to be allocated in each city. This subproblem is static and reduces to maximize total national production i.e.

$$\begin{aligned} & \max \{Y^A + Y^B\} \\ & \sum_{j=A,B} N^j = N \\ & Y^j = (N^j)^\alpha (K^j)^\beta \\ & K^j \text{ given} \end{aligned}$$

for any $j = A, B$. The solution to this problem is

$$\frac{N^A}{N^B} = \left(\frac{k^A}{k^B} \right)^{\frac{\beta}{1-\alpha}} \quad (5)$$

where $k^j \equiv \frac{K^j}{N}$, $j = A, B$ represents the stock of per capita capital installed in city j . Similarly, I define the following per capita variables: $y^j \equiv \frac{Y^j}{N}$, $i^j \equiv \frac{I^j}{N}$. This solution implies that the optimal ratio of labor between the two cities must be proportional to their ratio of installed capital. Substituting this optimal labor allocation in the aggregate production, one obtains the following value function:

$$f \equiv \sum_{j=A,B} y^j (N^{j*}) = \Omega^{1-\alpha}$$

where total population has been normalized to one. The variable $y^j \equiv \frac{Y^j}{N}$, $j = A, B$ represents per capita output in city j , and $\Omega \equiv \sum_{j=A,B} (k^j)^{\frac{\beta}{1-\alpha}}$.

Define the marginal product of capital in city j as

$$f_j \equiv \frac{\partial f}{\partial k^j} = \beta \Omega^{-\alpha} (k^j)^{\frac{\beta}{1-\alpha}-1} \quad (6)$$

Stage 2: Optimal Allocation of Capital

In this stage of the problem, capital is optimally allocated across cities and over time to maximize the flow of per capita utility. This dynamic maximization problem can be written in per capita terms as

$$\begin{aligned} & \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\ & \sum_{j=A,B} i^j + \sum_{j=A,B} g(k^j) + c = f(k^A, k^B) \\ & \dot{k}^j = i^j - \delta k^j \\ & i^j \geq 0 \\ & k_0^j \text{ given} \end{aligned}$$

for any $j = A, B$. Note that the value function obtained in stage 1 is used on the right-hand side of the resources constraint.

The Hamiltonian of this subproblem is

$$H = e^{-\rho t} \ln c + \lambda [f - \sum_{j=A,B} i^j - \sum_{j=A,B} g(k^j) - c] + \sum_{j=A,B} \mu_j [i^j - \delta k^j]$$

where λ , μ_j , $j = A, B$ are Lagrange multipliers. This problem has three control variables (c, i^A, i^B) and two state variables (k^A, k^B) . The first order conditions for city j are:

$$\frac{\partial H}{\partial c} = 0 \Leftrightarrow e^{-\rho t} \frac{1}{c} = \lambda \quad (7)$$

$$\frac{\partial H}{\partial i^j} \leq 0 \Leftrightarrow -\lambda + \mu_j \leq 0 \quad (8)$$

$$\frac{\partial H}{\partial k^j} = -\dot{\mu}_j \Leftrightarrow \lambda[f_j - g'(k^j)] - \mu_j \delta = -\dot{\mu}_j \quad (9)$$

and the transversality condition is

$$\lim_{t \rightarrow \infty} k_t^j \mu_{jt} = 0 \quad (10)$$

As in the decentralized problem, it is clear that, due to the presence of increasing returns to scale, capital is more productive in the largest city, i.e. $f_A > f_B, \forall k^A > k^B > 0$.

Assumptions 1-3 from the decentralized problem are also used here. The only two differences are first, that the convergence period \hat{t} is different from the one in the competitive equilibrium. Second, as it will be shown below, the gross marginal product of capital in city j , differs between the two problems.

As in the decentralized economy, depending on the parameters of the model, the economy's steady state may be characterized by a degenerate distribution of city size where all cities have equal population and capital (catch-up case) or by a city structure with cities of different size (non catch-up case). The qualitative evolution of the economy is then displayed, as in the decentralized economy, in Figures 3-6

Equivalence between both problems

To make both problems comparable we need the following restriction on the externality parameter ψ , I must impose the restriction²²:

$$0 < \psi < \alpha$$

The presence of a positive externality of aggregate capital in a city on each of the firms operating there implies a lower investment rate in each city at each point in time. This low investment rate generates two types of inefficiencies in the economy. The first one is the standard inefficiency à la Romer (1986). This can be easily seen in the expressions for the marginal product of capital in the

²²In the planner's problem, β , the capital share coefficient, is restricted between 0 and 1. Therefore, since $\beta = \psi - \alpha + 1$, one needs to assume $0 < \psi < \alpha$.

two problems (equations (3) and (6)). Since $1 - \alpha < \beta$, firms underinvest in the decentralized economy, lowering the growth rate of the economy.

The second inefficiency has to do with the congestion costs. Since these costs are convex, it is optimal to construct cities as similar in size as possible. In the non catch-up scenario, it is clear that $\left(\frac{k^A}{k^B}\right)^{market} > \left(\frac{k^A}{k^B}\right)^{planner}$, whereas in the case of convergence between cities these two ratios are equal, although the market ratio has been strictly larger than the planner's one for at least one period.

Proofs

Proposition 1: $f_A > f_B, \forall k^A > k^B > 0$

Proof: $f_A > f_B$ implies

$$(k^A)^{\frac{\beta}{1-\alpha}-1} > (k^B)^{\frac{\beta}{1-\alpha}-1}$$

or

$$\left(\frac{k^A}{k^B}\right)^{\frac{\psi}{1-\alpha}} > 1$$

which is true as long as $k^A > k^B > 0$, since increasing returns to scale imply $\psi > 0$.

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

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

Proof: Rewrite the household's problem as follows:

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

$$\begin{aligned} \sum_{j=A,B} i^j + c &= \omega + \sum_{j=A,B} r^j z^j \\ \dot{z}^j &= i^j, \forall j = A, B \\ i^j &\geq 0, \forall j = A, B \\ z_0^j &, \forall j = A, B \text{ given} \end{aligned}$$

where i^j represents the net investment rate in asset j . The Hamiltonian of this problem is

$$H = e^{-\rho t} \ln c + v(\omega + \sum_{j=A,B} r^j z^j - \sum_{j=A,B} i^j - c) + \sum_{j=A,B} \theta^j i^j$$

where $v, \theta^j, \forall j = A, B$ are Lagrange multipliers. The first order conditions are:

$$\frac{\partial H}{\partial c} = 0 \Leftrightarrow e^{-\rho t} \frac{1}{c} = v \quad (11)$$

$$\frac{\partial H}{\partial i^j} = 0 \Leftrightarrow -v + \theta^j \leq 0 \quad (12)$$

$$\frac{\partial H}{\partial k^j} = -\dot{\theta}^j \Leftrightarrow v r^j = -\dot{\theta}^j \quad (13)$$

Suppose in the time interval $(0, \hat{t}]$, $i^B > i^A = 0$. This implies that $v = \theta^B$ and so from (13)

$$\frac{\dot{v}}{v} = \delta - \tilde{f}_B \quad (14)$$

Taking logs and derivatives in (11)

$$-\rho - \frac{\dot{c}}{c} = \frac{\dot{v}}{v} \quad (15)$$

Then, equating (14) and (15) one has

$$\gamma_c = \tilde{f}_B - \rho - \delta$$

since in this time interval $\tilde{f}_B < \tilde{f}_A$ this investment policy is not optimal for consumers. The same logic implies that any investment in city B during this interval is also suboptimal. In the time interval $(\hat{t}, t^*]$ the two net marginal

products of capital are equated and so the growth rate of consumption is given by

$$\gamma_c = \tilde{f}_j - \rho - \delta, \forall j = A, B$$

since investment is strictly positive in both cities. Finally, at any period $t > t^*$, consumption is constant since the steady state has been reached.

Proof of Existence and Uniqueness

Proposition 3: The equilibrium in the economy exists and is unique if $k^j(t) \geq 1$, $j = A, B$, $\forall t$ and $\sigma > \frac{2\beta}{1-\alpha}$.

Proof: For simplicity, I only present the proof for the planner's problem. It is straightforward to show that the results also apply in the competitive economy. The original optimal problem is:

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

$$\begin{aligned} \sum_{j=A,B} i^j + \sum_{j=A,B} g(k^j) + c &= f(k^A, k^B) \\ \dot{k}^j &= i^j - \delta k^j, j = A, B \\ i^j &\geq 0, j = A, B \\ k_0^j, j &= A, B \text{ given} \end{aligned}$$

where $g(k^j) = (k^j)^\sigma$, $\sigma > 1$

and

$$f \equiv \Omega^{1-\alpha}$$

$$\text{where } \Omega \equiv \sum_{j=A,B} (k^j)^{\frac{\beta}{1-\alpha}}.$$

Using the Arrow-Kurz (1970) method, I will express the Hamiltonian as a function of the states and co-states of the problem and show that it is concave on the two states of the economy, k^A and k^B . Substituting c in the objective function one gets the following current-value Hamiltonian:

$$H = \ln(f - \sum_{j=A,B} i^j - \sum_{j=A,B} g(k^j)) + \sum_{j=A,B} \lambda_j (i^j - \delta k^j)$$

The first order conditions with respect to the controls i^j , $j = A, B$ are:

$$\frac{\partial H}{\partial i^j} = 0 \Leftrightarrow \frac{1}{c} = \lambda_j, j = A, B$$

Therefore

$$\lambda_A = \lambda_B \equiv \lambda = \frac{1}{c}$$

The new Hamiltonian is then

$$H = \ln\left(\frac{1}{\lambda}\right) + \lambda\left(\sum_{j=A,B} i^j - \delta \sum_{j=A,B} k^j\right)$$

One can then express the sum of the two controls as a function of the states and the co-state of the problem:

$$\sum_{j=A,B} i^j = f - \sum_{j=A,B} g(k^j) - \frac{1}{\lambda}$$

Finally,

$$H = \ln\left(\frac{1}{\lambda}\right) + \theta\left(f - \sum_{j=A,B} g(k^j) - \frac{1}{\lambda} - \delta \sum_{j=A,B} k^j\right)$$

To ensure the problem is concave, one needs to show that the new Hamiltonian is concave with respect to k^A and k^B . This is the case if its Hessian is negative semidefinite, i.e. $\frac{\partial^2 H}{\partial (k^A)^2} < 0$ and $\frac{\partial^2 H}{\partial (k^A)^2} \frac{\partial^2 H}{\partial (k^B)^2} - \left(\frac{\partial^2 H}{\partial k^A \partial k^B}\right)^2 \geq 0$. The second order condition of the Hamiltonian with respect to k^A is:

$$\frac{\partial^2 H}{\partial (k^A)^2} = \theta(f_{AA} - g''(k^A))$$

For this to be negative one requires $f_{AA} < g''(k^A)$. Define the following functions: $\xi(k^A, k^B) \equiv f(k^A, k^B)^{\frac{1}{1-\alpha}} - (k^A)^\sigma - (k^B)^\sigma$ and $h(x) \equiv x^{\frac{\beta}{1-\alpha}} - x^\sigma$. I will first prove that the function $h(x)$ is concave in R^+ . This is the case if

$$\ln\left(\frac{\frac{\beta}{1-\alpha}\left(\frac{\beta}{1-\alpha} - 1\right)}{\sigma(\sigma - 1)}\right) < \left(\sigma - \frac{\beta}{1-\alpha}\right) \ln x$$

which is verified if $\sigma > \frac{\beta}{1-\alpha}$ and $x \geq 1$. Since $f(k^A, k^B)^{\frac{1}{1-\alpha}} = (k^B)^{\frac{\beta}{1-\alpha}} + (k^A)^{\frac{\beta}{1-\alpha}}$ this also shows that $\frac{\partial^2 \xi(k^A, k^B)}{\partial (k^A)^2} < 0$ and, by symmetry, $\frac{\partial^2 \xi(k^A, k^B)}{\partial (k^B)^2} < 0$.

Hence the Hessian of $\xi(k^A, k^B)$ is given by $D^2 \xi \equiv \begin{pmatrix} \frac{\partial^2 \xi}{\partial (k^A)^2} & 0 \\ 0 & \frac{\partial^2 \xi}{\partial (k^B)^2} \end{pmatrix}$. therefore, the second condition to have a negative semidefinite Hessian is satisfied

since $\frac{\partial^2 \xi}{\partial (k^A)^2} \frac{\partial^2 \xi}{\partial (k^B)^2} > 0$. This concludes the proof that $\xi(k^A, k^B)$ is convex in k^A and k^B .

The original function is $\phi(k^A, k^B) \equiv f(k^A, k^B) - (k^A)^\sigma - (k^B)^\sigma$. The only difference between the functions $\xi(k^A, k^B)$ and $\phi(k^A, k^B)$ is that in the former, the function $f(k^A, k^B)$ is elevated to the power of $\frac{1}{1-\alpha}$, whereas in the latter this exponent is equal to one. It is obvious that the exponent $\frac{1}{1-\alpha}$ increases the absolute value of the diagonal terms of the Hessian.

To check the second condition for the Hessian to be negative semidefinite, one needs to numerically evaluate the matrix at different parameter values. The result of this exercise is that, if $\sigma > \frac{\beta}{1-\alpha}$ the second condition is verified in 62%-73% of the cases. The stronger assumption $\sigma > \frac{2\beta}{1-\alpha}$ ensures that the condition is verified in all cases. As an example, the combination of parameters $\alpha = 0.2$, $\beta = 0.9$, and $\sigma = 2.3$ satisfies this restriction.

Appendix C

Multiple Cities

In this section I sketch a model with $J > 2$ cities and argue that the behavior of the economy will also be characterized by a sequential growth of cities as long as cities differ in size in the initial period.

Households

Households' budget constraint can be written as

$$\sum_{j=1}^J \dot{z}^j = \omega + \sum_{j=1}^J r^j z^j - c$$

where $z^j \equiv \frac{Z^j}{N}$ represent the per capita holdings of asset j , $j = 1, \dots, J$. Assuming logarithmic utility, the problem of a representative household in per capita terms is:

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

$$\begin{aligned} \sum_{j=1}^J \dot{z}^j &= \omega + \sum_{j=1}^J r^j z^j - c \\ \dot{z}^j &\geq 0, \forall j = 1, \dots, J \\ z_0^j, \forall j &= 1, \dots, J \text{ given} \end{aligned}$$

where c represents per capita consumption and, as stated above households face the irreversibility constraints $\dot{z}^j \geq 0$, $j = 1, \dots, J$.

Firms

The production function of firm i located in city j is identical to the one described in the two-city case. Free labor mobility implies

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

which in turn implies the following gross marginal product of capital (i.e. without considering the congestion costs) for city j :

$$f_j = \chi(k^1, k^2, \dots, k^J)$$

where total population N has been normalized to one.

Equilibrium

As in the two-city case it is straightforward to show that $f_j > f_{j'}, \forall k^j > k^{j'} > 0$. Consider the following two assumptions about the initial city size distribution, which, as in the two-city case, make the dynamics of the problem non trivial:

Assumption 1'

$$k_0^1 > k_0^2 > \dots > k_0^J$$

Assumption 2'

$$\tilde{f}_1(0) > \tilde{f}_2(0) > \dots > \tilde{f}_J(0) > \rho + \delta$$

where $\tilde{f}_j(0) \equiv f_j(0) - g'(k_j(0)), \forall j = 1, 2, \dots, J$

With these assumptions the unique competitive equilibrium in an economy with J cities has the same characteristics as the simplest two-case developed in the main text. At the initial date, households only invest in the city with initially largest stock of capital, city 1, since it is the most profitable investment opportunity. When the stock of installed capital in city 1 reaches a critical level (dictated by congestion costs), investment starts at city 2. The same process continues until investment takes place in the initially smallest city J . As before, to ensure that some investment takes place in every city we need to assume that the productivity of capital in all cities is high enough to make investment profitable i.e.

Assumption 3'

$$\tilde{f}_{j, \hat{t}_j} > \rho + \delta, j = 1, \dots, J$$

where \hat{t}_j represents the period at which city $j-1$ has reached its critical level of congestion costs. The exact shape of the steady state distribution of city size depends on the parameter values that are specified. The two extreme cases are

one in which all cities end up with an identical size (*pure catch-up*) and one in which all cities have a different size (*pure non catch-up*), with a large range of intermediate solutions in between. The existence and uniqueness of a solution to this problem requires that the Hessian of the transformed Hamiltonian (using the Arrow-Kurz method as above) is negative semidefinite. See Mas-Colell, Winston and Green (1995) for specific details.

Appendix D

Figure 1: Population Shares in the U.S.

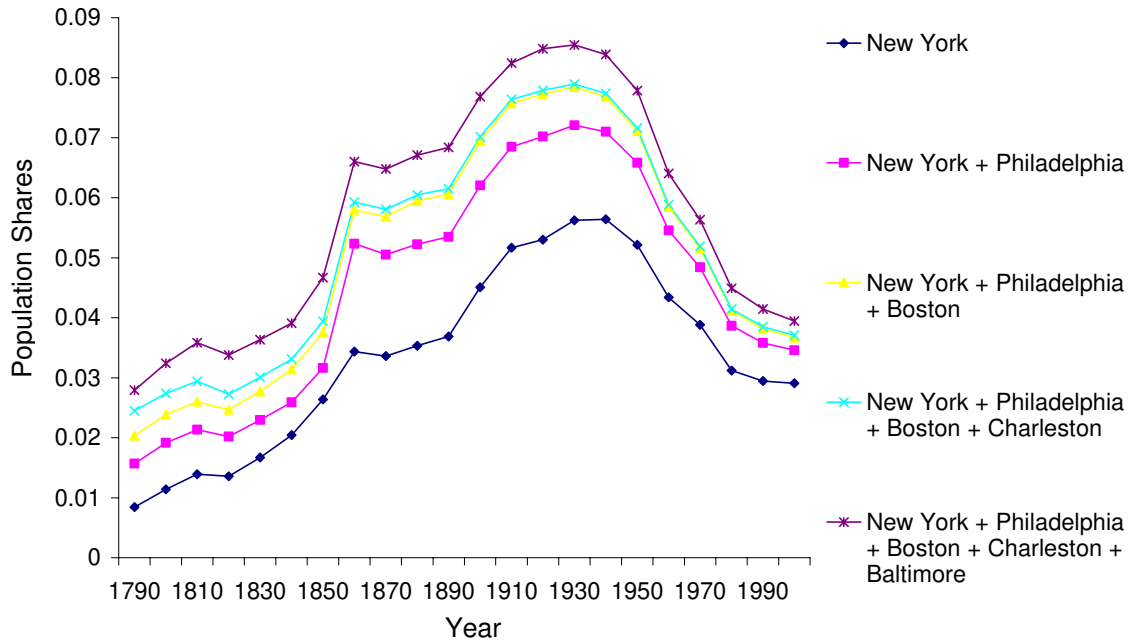
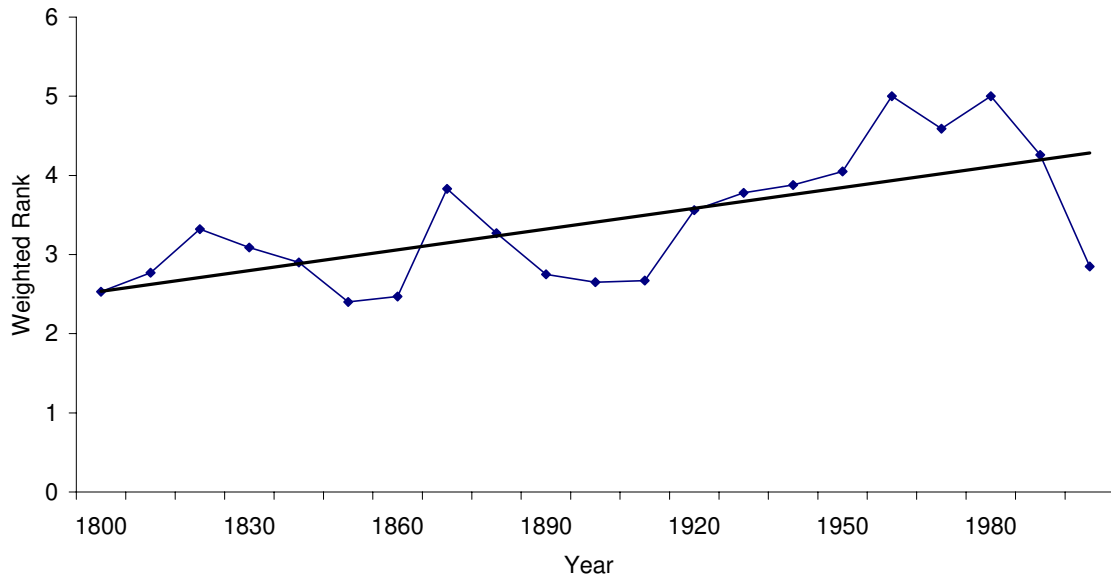


Figure 2: Weighted Rank of the Five Fastest Growing Cities in the U.S.



References

- [1] Abdel-Rahman, H., and A. Anas (2004), "Theories of Systems of Cities," in J.V. Henderson and J-F. Thisse (eds), *Handbook of Urban and Regional Economics*, volume 4, Cities and Geography, North Holland, in press.
- [2] Alonso, W. (1980), "Five Bell Shapes in Development." Papers of the Regional Science Association, 45, 5-16
- [3] Arnott, R. and J. Stiglitz (1979), "Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size." *Quarterly Journal of Economics*, 93, 471-500
- [4] Axtell, R. L., and R. Florida (2001), "Emergent Cities: a Microeconomic Explanation of Zipf's Law," paper presented at the Society of Computational Economics, Yale University
- [5] Berliant, M., and H. Konishi.(2000), "The endogenous formation of a city: Population agglomeration and marketplaces in a location specific production economy." *Regional Science and Urban Economics*, 30(3):289—324.
- [6] Berliant, M., and P. Wang (1993), "Endogenous formation of a city without agglomerative externalities or market imperfections: Marketplaces in a regional economy." *Regional Science and Urban Economics*, 23(1):121—144.
- [7] Black, D. and J.V. Henderson (1999), "A Theory of Urban Growth." *Journal of Political Economy* 107:252-284
- [8] Brakman, S., H. Garretsen, C. Van Marrewijk, and M. van den Berg (1999), "The return of Zipf: a further understanding of the rank-size distribution", *Journal of Regional Science* 39, pp. 183-213
- [9] Brakman, S., H. Garretsen, C. Van Marrewijk (2001), *An Introduction to Geographical Economics*, Cambridge University Press, Cambridge and New York
- [10] Davis, D. R., and D. E. Weinstein (2002), "Bones, bombs, and break points: the geography of economic activity", *American Economic Review* 92, pp. 1269-1289
- [11] Carlino, G. and S. Chatterjee (2001), "Aggregate Employment Growth and the Deconcentration of Metropolitan Employment." *Journal of Monetary Economics*, Vol. 48, Number 3, December
- [12] Cordoba, J. C (2003), "On the Distribution of City Sizes," forthcoming, *Journal of Urban Economics*
- [13] Cuberes, D. (2006), "The Rise and Decline of Cities," *revised and resubmitted*, *Journal of Economic Dynamics and Control*

- [14] Davis, J, and V. Henderson (2003), "Evidence on the Political Economy of the Urbanization Process." *Journal of Urban Economics* 53, 98-125
- [15] Dobkins, L., and Y. Ioannides (1998a), "Dynamic Evolution of the U.S. City Size Distribution," in J.M. Huriot and J.F. Thisse eds., *The Economics of Cities* (New York, NY: Cambridge University Press,)
- [16] Dobkins, L., and Y. Ioannides (2001), "Spatial Interactions Among U.S. Cities: 1900-1990." *Journal of Economic Literature*
- [17] Duranton, G. (2002), "City Size Distribution as a Consequence of the Growth Process." Department of Geography and Environment, London School of Economics
- [18] Duranton, G., and D. Puga (2004), "'Micro-foundations of urban agglomeration economies," in J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, Vol. 4, 2004. Amsterdam: North-Holland, 2063-2117
- [19] Eaton, J., and Z. Eckstein (1997), "Cities and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics*, XXVII, 443-474
- [20] Eeckhout, J. (2004), "Gibrat's Law for (all) Cities." *American Economic Review*, December, vol. 94, No 5, 1429-1451(23)
- [21] Fujita, M., P. Krugman, and A. Venables (1999), *The Spatial Economy* (Cambridge, MA: MIT Press)
- [22] Gabaix, X. (1999a), "Zipf's Law and the Growth of Cities." *American Economic Review Papers and Proceedings*, LXXXIX, 129-132
- [23] Gabaix, X. (1999b), "Zipf's Law for Cities: an Explanation." *Quarterly Journal of Economics* 114:739-767
- [24] Gabaix X., and Y. Ioannides (2004), "The Evolution of City Size Distributions," in Henderson J. V., and J.-F. Thisse, eds. *Handbook of Regional and Urban Economics*, North-Holland, Amsterdam, 2341-2378.
- [25] Gibrat, R.(1931), *Les inégalités économiques* (Paris, France: Librairie du Recueil Sirey)
- [26] Glaeser, E., J. Shapiro (2001), "Is There a New Urbanism? The Growth of U.S. Cities in the 1990s." NBER Working Paper No W8357
- [27] Glaeser, E., J. Scheinkman, and A. Shleifer (1995), "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, XXXVI, 117-143
- [28] Grossman, G. M. and E. Helpman (1991), "Quality Ladders in the Theory of Growth." *Review of Economic Studies* 58:43-61

- [29] Henderson, J. (1974), "The Sizes and Types of Cities." *American Economic Review*, LXIV, 640-656
- [30] Henderson, J. (1988), "Urban Development: Theory, Fact, and Illusion." Oxford University Press
- [31] Henderson, J. and Hyoung Gun Wang (2003), "Urbanization and City Growth," working paper
- [32] Henderson, J. and A. J. Venables (2004), "The Dynamics of City Formation: Finance and Governance," working paper
- [33] Ioannides, Y. M., and H. G. Overman (2003), "Zipf's Law for Cities: An Empirical Examination." *Regional Science and Urban Economics* 33: 127-137
- [34] Konishi, H. (2000), "Formation of hub cities: Transportation cost advantage and population agglomeration." *Journal of Urban Economics* 48(1):1-28
- [35] Koopmans, T. C. (1957), "Three Essays on the State of Economic Science.", New York: McGrawHill
- [36] Krugman, P. (1996b), "Confronting the Mystery of Urban Hierarchy." *Journal of the Japanese and the International Economies* 10: 399-418
- [37] Krugman, P., and E.S. Brezis (1997), "Technology and Life Cycle of Cities." *Journal of Economic Growth*, 369-383.
- [38] Lai, J. (2001), "Using Random Growth Models to Explain Zipf's Law." unpublished paper, University of Texas, Austin
- [39] Mas-Colell, A., M. D. Whinston, and J. R. Green (1995), *Microeconomic Theory*. Oxford University Press
- [40] Mills, E S. (1967), "An aggregative model of resource allocation in a metropolitan area." *American Economic Review Papers and Proceedings*, 57(2):197-210
- [41] Mirrlees, J. A. (1972), "The optimum town." *Swedish Journal of Economics*, 74(1):114-135
- [42] Romer, Paul M. (1986), "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94, 5 (October), 1002-1037
- [43] Rosen, K. and M. Resnick (1980), "The Size Distribution of Cities: an Examination of the Pareto Law and Primacy." *Journal of Urban Economics* 8:165-186
- [44] Rossi-Hansberg, E., and M. L. J. Wright (2003), "Urban Structure and Growth." NBER Working Paper No W11262

- [45] Soo, K. T. (2005): "Zipf's Law for Cities: a Cross Country Investigation." *Regional Science and Urban Economics*, vol. 35, Issue 3, pp. 239-263
- [46] Sutton, J. (1997), "Gibrat's Legacy.", *Journal of Economic Literature*, XXXV, pp. 40-59
- [47] Tamura, R. (2002), "Human Capital and the Switch From Agriculture to Industry." *Journal of Economic Dynamics and Control*, 27, 207-242
- [48] Tamura, R. (2006), "Human Capital and Economic Development," forthcoming, *Journal of Development Economics*
- [49] Wang, P. (1990), "Competitive equilibrium formation of marketplaces with heterogeneous consumers." *Regional Science and Urban Economics*, 20 (2): pp. 295—304
- [50] Wang, P. (1993), "Agglomeration in a linear city with heterogeneous households." *Regional Science and Urban Economics*, 23 (2): pp. 291—306
- [51] Zipf, G. (1949), *Human Behavior and the Principle of Least Effort* (Cambridge, MA: Addison-Wesley)