ELSEVIER Regional Science and Urban Economics 27 (1997) 443–474

regional SCIENCE & urban ECONOMICS

Cities and growth: Theory and evidence from France and Japan

Jonathan Eaton[a,*], Zvi Eckstein[1,b]

[a]Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA
[b]Department of Economics, Tel-Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

Received 6 May 1995; revised 14 February 1996

Abstract

The relative populations of the top 40 urban areas of France and Japan remained very constant during these countries' periods of industrialization and urbanization, and are described quite well by the 'rank-size rule.' Moreover, projection of their future distributions based on past growth indicates that their size-distributions in steady state will not differ essentially from what they have been historically. Urbanization consequently appears to have taken the form of the parallel growth of cities, rather than convergence to an optimal city size or the divergent growth of the largest cities. We develop a model of urbanization and growth based on the accumulation of human capital consistent with these observations. Our model predicts that larger cities will have higher levels of human capital, higher rents and higher wages per worker, even though workers are homogeneous and free to migrate between cities. Cities grow at a common growth rate, with relative city size depending upon the environment that they provide for learning.

Keywords: Urbanization; Human capital; Endogenous growth

JEL classification: 018; R11; O41

*Corresponding author. Tel. (1-617) 353-4142; fax: (1-617) 353-4449; e-mail: jeaton@acs.bu.edu
[1] Tel.: (972-3) 640-9914; e-mail: eckstein@econ.tau.ac.il

0166-0462/97/$17.00
PII S0166-0462(96)02160-6

1. Introduction

A basic component of economic development is the movement of population from the countryside to cities.[2] In this paper we consider the particular experiences of France between 1876 and 1990 and Japan between 1925 and 1985. These countries experienced the phenomenon of substantially increased urbanization during these periods of industrialization.[3]

We find that urbanization took the form of similar growth rates across cities of different sizes ('parallel growth'), rather than either an increase in the population of larger cities relative to other cities ('divergent growth') or of the growth of smaller cities relative to larger cities ('convergent growth'). Moreover, the 'rank-size rule,' that cities' populations are proportional to the inverse of their rank, captures the behaviour of relative populations throughout the period quite well.

We then develop a theory of growth and urbanization consistent with the finding of parallel growth, in which urban growth is driven by the acquisition of human capital.[4] Observers of urban development have emphasized the role of human capital in the functioning of cities.[5] Lucas (1988), in particular, specifically relates what his model identifies as the driving process of economic growth, the acquisition of knowledge and the externalities associated with it, to the forces that lead to the development of cities:

"It seems to me that the 'force' we need to postulate to account for the central role of cities in economic life is of exactly the same character as the 'external human capital' I have postulated as a force to account for certain features of aggregate development" (Lucas, 1988, pp. 38–39).

[2] Kuznets (1966) historical data document the increased share of population living in urban areas during the economic development of a number of rich nations. Chenery and Syrquin (1975) cross-sectional evidence shows that the share of a nation's population that lives in cities rises with its per capita GNP.

[3] Between 1876 and 1990 the population of France grew from 36.9 million to 55.8 million, while the agglomeration of Paris grew from 2.6 million to 9.3 million. Between 1925 and 1985 Japan grew from 59.7 to 122.6 million people, while Tokyo grew from 6.5 million to 26.5 million. Hence we are considering each country for a period in which the total population roughly doubled, while the population of the largest urban areas (as well as the other urban areas in our sample) approximately quadrupled.

[4] Work on the determinants of economic growth has attempted to explain the secular growth of per capita GNP in terms of: (i) physical capital investment (Solow, 1956; Romer, 1986), (ii) the accumulation of human capital (Uzawa, 1965; Lucas, 1988), (iii) product and process innovation (Inada, 1969; Grossman and Helpman, 1991), and (iv) learning by doing (Arrow, 1962; Young, 1991).

[5] See, for example, the discussions in Jacobs (1969), (1984); Henderson (1988); Rauch (1993); Glaeser et al. (1992). Much of the urban literature has focused on the development of cities that specialize in the production of particular commodities (as in Henderson, 1988). Our approach, in contrast, focuses on cities that need not be specialized, whose productivity derives from the interaction of individuals with complementary forms of knowledge.

Nevertheless, this literature has provided little formal link between the processes of economic growth and of urbanization.[6]

Our theory draws upon two literatures, that of the endogenous determinants of economic growth, in particular the Lucas (1988) model of human capital accumulation, and that of circular cities, developed by Mills (1967); Arnott (1979); Helpman and Pines (1980); Henderson (1987), (1988), among others. Existing work on urbanization and growth (e.g., Miyao, 1987; Henderson, 1987), predicts that urbanization takes the form of the creation of new cities, whose size converges to an optimum city size.[7] In our alternative approach urbanization involves the parallel expansion of a given number of cities.

In Section 2 we consider the basic empirical question of the extent to which the process of urbanization associated with development is primarily extensive, taking the form of the creation of new cities, or intensive, involving the growth of existing cities. We examine these issues with data on 39 French urban areas from 1876 to 1990 and on 40 Japanese urban areas from 1925 to 1985. We find that the Lorenz curves of populations across urban areas in France and Japan remained almost identical for the entire period. Moreover, the curves for France and for Japan have very similar shapes. We also find that the 'rank-size rule', which holds that city populations are proportional to the inverse of their rank-order, describes the data quite well throughout the period.[8] Another finding is that (in the French data) that wages and housing rents are highly correlated with city size.

To obtain further evidence on the extent to which urban growth is parallel we

[6] An exception is the recent paper by Palivos and Wang (1994) who model the endogenous growth of a single city.

[7] Henderson's (1987) analysis relates new cities to new industries. The steady-state implication is that "the economy grows by churning out new cities at the rate of population growth." (p. 950). Ioannides (1994); Glaeser et al. (1992), on the other hand, find that diversified cities tend to grow faster.

[8] The notion that relative city size does not change as urban populations grow has an old tradition in the regional science literature. The 'rank-order rule,' also known as 'Zipf's Law' (Zipf, 1949) asserts that the product of a city's population and its rank in population is constant across cities and time. Beckmann (1958); Rodwin (1970); Henderson (1988); Gell-Mann (1994) discuss the rule and its history. Rosen and Resnick (1980) extensively analyse international evidence on the rank size rule as it applies to cities as defined by their political boundaries. They find that for most of the 44 countries in their sample (including Japan and France) cities are distributed more evenly than the rank size rule would predict. For France (and the 5 other countries for which they have data), however, switching from a political to a metropolitan definition of a city leads to much better performance of the rank size rule. With the exception of the analysis in Henderson (1988), which concerns the growth of specialized cities, the rule does not appear to have played a role in current theories of urban growth. The rule does not appear to be mentioned, for example, in the Handbook of Regional and Urban Economics (1987). Wheaton and Shishido (1981); Ades and Glaeser (1995) provide other evidence on the relationship between urban concentration and development. Their analysis is cross-sectional, however, and the second paper focuses solely on the largest or 'primate' city.

apply the Quah (1993) Markov transition probability model to our French and Japanese data. We find that the asymptotic distribution of cities is very close to what the distribution is now and has been historically. That is, if cities continue to change their positions relative to the mean as they have in the past, the size distribution that ultimately emerges will not differ much from what it is now.

We then develop a model of urbanization and growth consistent with the parallel steady-state growth of cities, but with possibly different short-term growth rates and changes in the ranking of individual cities.[9]

Section 3 provides a model of a growing system of cities with the implication that, in steady state, cities will maintain the same relative populations even as they grow in size over time. Land is a factor of production, and total productivity within a city declines with the distance of production from the city center. This last assumption is meant to capture in a simple way the contribution of urban agglomeration and proximity to productivity.

We relate total factor productivity in a city to its average level of human capital, as in Lucas (1988). A basic characteristic of a city is the environment that it provides for acquiring human capital (which can either be city-specific or general in terms of its applicability). Cities are linked together in terms of how their human capital stocks contribute to learning, much as the human capital stocks of different countries jointly contribute to learning in Lucas (1993). The interaction of the human capital stocks of different cities implies that, in the long run, city populations will grow at common rates.

The dynamics of the model determine the growth and the distribution of human capital. Migration provides the link between the growth and distribution of human capital among cities and their relative populations. We analyze migration between cities of different relative levels of human capital. The model implies that cities where time spent acquiring human capital is more productive will have larger populations, higher wages, higher land rents and higher levels of human capital per worker, correlations found in the data.[10]

In Section 4 we calibrate the model to our French and Japanese data. We find that under plausible parameter values that fit the data, relatively small differences

[9] Parallel growth of per capita income levels has been observed among countries. The literature on international technology diffusion has models with this implication. Examples are Parente and Prescott (1994); Benhabib and Spiegel (1994); Eaton and Kortum (1994). Except for the last, these papers assume that knowledge flows are unidirectional. Here we allow for very general knowledge spillovers among cities. Moreover, since these other models apply to countries, they treat workers as immobile. A purpose of our model here is to determine city size on the basis of migration opportunities. Our model also explains why migration may not eliminate differences in real wages among cities, even though individuals in our model are ex ante identical.

[10] We find a correlation between city population and wages and population and price level in our French data. Rauch (1993) finds a significant positive correlation between levels of human capital and city size in U.S. cities. Henderson (1988) discusses other evidence on the correlation of education levels and city size.

in city characteristics can imply large differences in city populations. Section 5 concludes.

2. Evidence on the size distribution of cities

We have already mentioned that economic growth and urbanization are parallel processes. We now consider the question of how cities of different sizes grow during the process of development. One possibility is that urbanization occurs as new cities develop, and as smaller cities catch up with larger ones, in which case the size distribution of cities would become more even over time. At the other extreme, urbanization could take the form of the expansion of the largest cities, so that the size distribution would become more unequal.

To examine this issue we look at historical data on urban agglomerations from France and Japan. We choose France for several reasons. First, since it is a high-income country we can observe the evolution of its urban structure during the process of industrialization. Second, it has constituted an intact nation-state more or less within its current borders throughout the industrial revolution. Third, its total land area was settled at the origin of the industrial revolution. Fourth, it is geographically large enough to contain a number of distinct, large metropolitan areas.[11] Japan shares the first three characteristics but not the fourth. However, we have constructed data on agglomerations for Japan that seem to be consistent over the period for which we have data.[12]

We have collected data on the population of 39 urban agglomerations in France for the years: 1876, 1911, 1936, 1954, 1962, 1982 and 1990.[13] Our criterion for selection is a 1911 population of at least 50 000 inhabitants. Only two agglomerations (Grasse–Cannes–Antibes and Bethune) not in our sample rank among the top 35 cities in 1990 (rank 17 and 19) in 1990. The smallest agglomeration in our sample (Hagondage) ranks 50 in 1990. Hence, there are almost no new urban

[11] While Great Britain shares these characteristics, because of its much greater population density, metropolitan areas have had a much greater tendency to blend into each other in the process of urbanization. We could not find historical data based on definitions of urban areas that remained consistent during the period of interest.

[12] A common feature of these countries is that they have had well developed urban systems with dominant primate cities for several centuries. These features are shared by most countries that have occupied a relatively stable area and have been fully settled over a long period of time, such as Spain, Portugal, the United Kingdom and the Scandinavian countries. Countries for which this characterization is not accurate, such as Germany, Italy, Canada, Australia and the United States, were either unified or settled relatively recently. In Switzerland language and religious differences have hampered migration. Even in countries of recent settlement, where cities in newly settled regions have tended to grow relative to those in regions settled earlier, the distribution of city sizes has tended to obey Zipf's Law; see Mills (1972).

[13] The data are from INSEE, Annuaire Statistique de la France, various issues. They are reported in Table A1 of the Working Paper version of this paper (Eaton and Eckstein, 1994).

agglomerations since 1911 and no urban agglomeration has fallen drastically in its relative size; i.e., no city that was (relatively) big in 1911 has 'died.'[14]

For Japan we have organized data for the largest 40 agglomerations from 1925 to 1985, for every 5 years.[15] We included all agglomerations that had a population of 250 000 and more in 1965. As in France, we find only a few (3) cities that become marginally larger than the cities that are included in the sample. But there is no 'new' city in that every city in the top 30 cities in Japan in 1990 was in our sample. Similarly, none of the cities in the sample 'died', in the sense that none of the cities in the sample ranks below 50 in 1985.

As we discussed in the introduction, the urban economics literature addresses the issue of new cities and the optimal size of a city (e.g., Henderson, 1988), in both static and dynamic contexts. The evidence from France and Japan is that there are no new cities. Even the tourist cities in France, which may be viewed as new, existed and had a moderate size by the beginning of the 19th century.

2.1. Lorenz curves

We compute Lorenz curves for population for each year in our sample for France and every second year for Japan. These are displayed in Figs. 1 and 2.[16] The French data demonstrate starkly how the size distribution of cities has not changed noticeably during the most spectacular period of growth of population, movement of population from rural to urban areas, and growth of income per capita. While the population of Paris nearly quadrupled between 1876 and 1990 (while that of France as a whole did not quite double) its share of the total population of our sample of cities remained stable at 40–43%. The change of ranking among the cities (up and down within the sample) is more frequent for small cities. That is, the relative size of a city in the sample is more stable among the largest cities.

The Japanese Lorenz curves show more movement towards less equal distribution of size. The share of the larger cities went up and, in particular, Tokyo's

[14] Moreover, the two towns that did grow substantially faster than the other cities and became agglomerations that rank among the top 20 in France are tourist centers with a significant locational advantage. If the demand for leisure and tourism is highly income elastic then the substantial increase of French income per capita over the last 130 years can easily explain the high growth of these two cities. Since our focus is on aggregate growth and the size distribution of cities that produce a common good, it seems reasonable to ignore those cities whose location led to higher growth due to a higher income elasticity of demand for their specialized product.

[15] Akiko Tamura constructed the Japanese urban agglomeration data from historical city population census data provided in Kanketsu Showa-kokusei Soran (Vol. 1), based on data provided by the Statistics Bureau of the Management and Coordination Agency of the Japanese government. In the appendix to the working paper version of this paper (Eaton and Eckstein, 1994), she explains the aggregation procedures. Tables A2 of the working paper present the data.

[16] Table A3 and Table A4 of the working paper version (Eaton and Eckstein, 1994) provide the raw data.

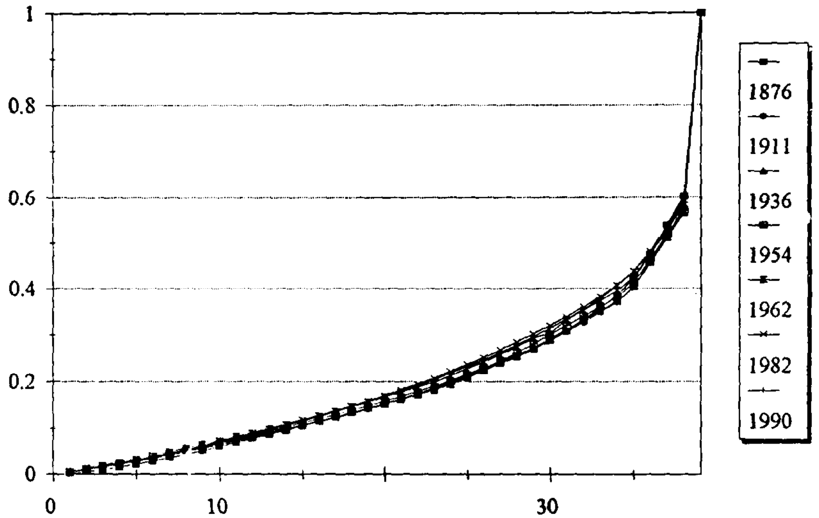


Fig. 1. Lorenz curves for French cities.

share increases for the whole population as well as among the top cities. It is interesting to note that the Lorenz curve for France is the same as the Japanese Lorenz curves for the early years. Hence, the two countries, which are significantly different in their geographical structure, have a very similar size distribution of cities.

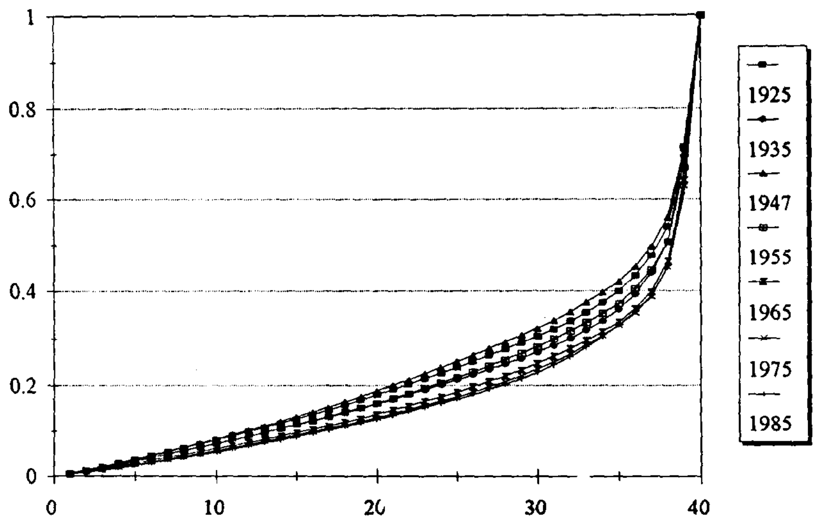
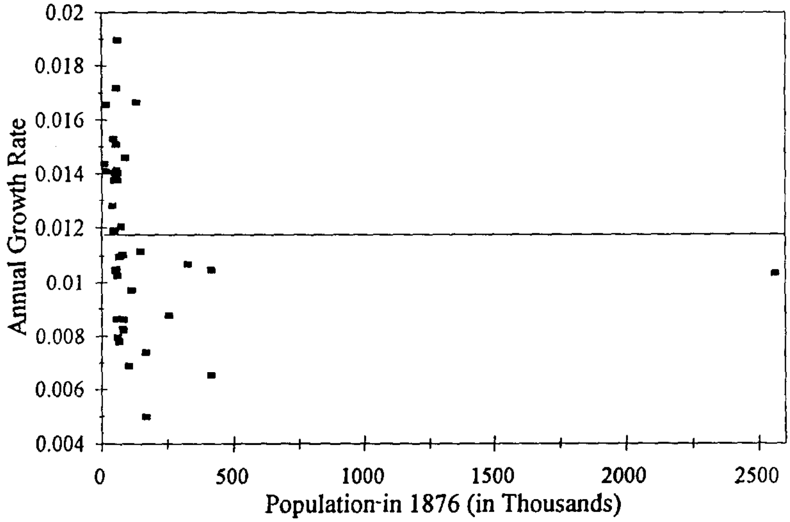


Fig. 2. Lorenz curves for Japanese cities.

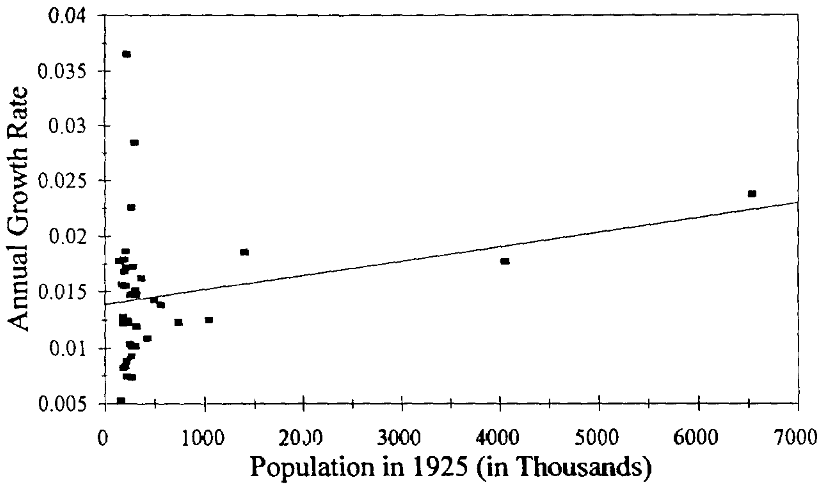
The stability of the Lorenz curves could be the consequence of any number of dynamic processes driving the population growth of individual cities. The most obvious possibility is that all cities on average grow at the same rate starting at different levels ('parallel growth'). Two possibilities are ruled out, however. If there were an upper bound on city size that was attained by any city in our sample then we would expect the initial level of population of the city to be negatively correlated with average growth rate ('convergence of size distribution'). The size distribution of cities would then be getting more equal over time. On the other hand, if the growth rate of a city is positively correlated with its initial size ('divergence of size distribution') then we would expect the Lorenz curves to exhibit increased inequality over time.

Fig. 3 displays the French average annual growth rates of the cities from 1876 to 1990 and the initial level of each city in 1876. The regression line implies that there is no correlation between the initial size of the agglomeration and the growth rate during that period. (The slope coefficient is -1×10^{-6} with a S.E. of 1.56×10^{-6}.) Fig. 4 presents the equivalent picture for the Japanese cities. Here, as well, there is no obvious correlation between the initial level. The slope is



R-Squared	0.0247	Coefficient	-1.28E-06
No. of Observations	39	Std. Error of Coef.	1.317E-06

Fig. 3. French cities: Growth rates and 1876 cities size.



R-Squared	0.0708	Coefficient	1.3E-09
No. of Observations	40	Std. Error of Coef.	8E-10

Fig. 4. Japanese cities: Growth rates and 1925 cities size.

positive but not significant (1.3×10^{-9} with a S.E. of 0.8×10^{-9}). This result is consistent with the stability of the Lorenz curves as well as with parallel growth.[17]

2.2. *The rank-size rule*

Given that the distribution of relative city size was so stable over the period, we now ask whether the populations of these urban areas obeyed the 'rank size rule.' According to this rule city populations among any group of cities at any time are proportional to the inverse of the ranking of their populations in that group. This result would obtain, for example, if the rank of a city of size N has the expectation:

[17] The results in Figs. 3 and 4 are similar to Barro (1991) on the relationship between per-capita income growth and the initial level of income per capita using the Summers and Heston (1991) country data. If Paris is omitted from the French data, then the regression has a significant negative slope. If Tokyo is omitted, then, the regression for Japan has a significant positive slope. Given the central role of these cities we think that it is wrong to drop them from the sample.

$$G(N) = CN^{-\theta}$$

with $\theta = 1$.[18] Hence, the expected size of the largest city is given by C. One simple way to examine the particular distribution is to estimate the equation:

$$\ln r_{it} = \ln C_t - \theta_t \ln N_{it} + u_{it}$$

where r_{it} is city i's rank in period t, N_{it} is its population, and u_{it} is a random deviation from the rank-size rule. Note that in general we allow for different intercepts and slopes in each period.

We estimated this relationship for France and Japan separately. While the estimates of $\ln C_t$ rose substantially from period to period the estimates of θ_t varied only negligibly from the estimate for the most recent period except for one of the 5 earlier periods in France and 3 of the 12 earlier periods in Japan. We report only the regressions in which coefficients are apparently significant in Tables 1 and 2.[19]

The evidence for France supports the hypothesis that $\theta = 1$ for any but the earliest period of observation, for which it is substantially below 1. The estimated value of θ for 1875 is 0.87. The implication is that the distribution of city sizes was more unequal during this year. For Japan, our estimate of θ is slightly less

Table 1
Pooled rank size regression (France)

Variable	Coefficient	STD. error	T-STAT.	2-TAIL SIG.
C	8.6540995	0.1061080	81.559344	0.0000
LX	−1.0310845	0.0179566	−57.420966	0.0000
D1876	−2.0122409	0.2050722	−9.8123538	0.0000
D1911	−0.8458149	0.0460474	−18.368341	0.0000
D1936	−0.6493937	0.0450498	−14.415031	0.0000
D1954	−0.5519536	0.0446744	−12.355031	0.0000
D1962	−0.3865258	0.0441162	−8.7615464	0.0000
LX1876	0.1518833	0.0427954	3.5490549	0.0005
R-squared	0.934963	Mean of dependent var.	2.739796	
Adjusted R-squared	0.933245	S.D. of dependent var.	0.860436	
S.E. of regression	0.222311	Sum of squared resid.	13.09682	
Log likelihood	27.19425	F-statistic	544.2302	
Durbin-Watson stat.	1.533374	Prob. (F-statistic)	0.000000	

Dependent variable is LG = log rank; SMPL range: 1–273; number of observations: 273.
LX = Log Size; LX1876 = dummy for log size in 1876.

[18] See Simon (1955) or Mills (1972) for a discussion.

[19] Since we have no basis for considering the error in this equation to be independent of the population we treat this exercise as curve fitting rather than as hypothesis testing. Standard errors should be interpreted as suggestive of goodness of fit rather than the basis for rigorous hypothesis testing.

Table 2
Pooled rank size regression (Japan)

Variable	Coefficient	STD. Error	T-STAT.	2-TAIL SIG.
C	15.791570	0.1614714	97.797928	0.0000
LX	−0.9649629	0.0118962	−81.115513	0.0000
D1930	−0.7010550	0.0433455	−16.173661	0.0000
D1935	−0.6212362	0.0431740	−14.389113	0.0000
D1940	−0.580487	0.0430958	−13.480403	0.0000
D1955	−0.3533213	0.0427595	−8.2629910	0.0000
D1960	−0.2907617	0.0426989	−6.8095780	0.0000
D1965	−0.2124461	0.0426426	−4.9820113	0.0000
D1970	−0.1345708	0.0426083	−3.1583235	0.0017
LX1925	−0.0621668	0.0034146	−18.206338	0.0000
LX1947	−0.0396425	0.0032989	−12.016757	0.0000
LX1950	−0.0336490	0.0032713	−10.286187	0.0000
R-squared	0.928438	Mean of dependent var.	2.758016	
Adjusted R-squared	0.926889	S.D. of dependent var.	0.863236	
S.E. of regression	0.233412	Sum of squared resid.	27.67633	
Log likelihood	24.79727	F-statistic	599.1590	
Durbin-Watson stat.	0.305530	Prob. (F-statistic)	0.000000	

LS//dependent variable is LG=log rank; SMPL range: 1–520; number of observations: 520.
LX=Log Size; LXYEAR=dummy for the log size of the year.

than 1, and somewhat lower still in 1925, 1947 and 1950. This finding reflects the somewhat greater inequality of city size in Japan relative to France. Nevertheless, for neither country did our estimates of θ appear to be changing systematically over time.[20]

2.3. Size and other features

For France we also have data on the average salary per full-time employee in each agglomeration in 1982 and 1989.[21] A regression of wage income on the 1982 and 1990 populations of the agglomeration yields a positive and significant coefficient for each year separately (see Table 3a). Furthermore, the coefficients of the two years estimated separately turn out to be close.[22] Hence, wages are higher in larger cities following a stable relationship. Table 3b reports the regression of

[20] How well do the two countries' primate cities fit the relationship? The Tokyo metropolitan area fits the rank size rule well, while Paris is significantly larger than what the estimated relationship would predict.

[21] These appear in Table A5 of the working paper version (Eaton and Eckstein, 1994).

[22] Since wages are highly correlated with human capital, we expect that the same cross-section correlation holds between city population and the average level of human capital.

Table 3
Cross-section regressions for French data

(a) Ln(Wage) on Ln(City's Population)

Year	1982	Year	1990
Constant	1.8644	Constant	1.7787
Std. err. of Y est.	0.0219	Std. err. of Y est.	0.0362
R-squared	0.3997	R-squared	0.3789
No. of observations	39	No. of observations	39
Degrees of freedom	37	Degrees of freedom	37
X coefficient(s)	0.0508	X coefficient(s)	0.0794
Std. err. of coeff.	0.0102	Std err. of coeff.	0.0167
T stat.	4.9638	T stat.	4.7506

(b) Ln(Price) on Ln(City's Population)

Constant	81.185
Std. err. of Y est.	7.3627
R-squared	0.5867
No. of observations	20
Degrees of freedom	18
X coefficient(s)	0.0044
Std. err. of coeff.	0.0009
T stat.	5.0551

the price of housing on city size, which indicates that larger cities have a higher cost of housing.[23]

2.4. Evolving size distributions

The Lorenz curves and results on the rank-size relationship establish that the size distribution of cities remained very constant during the period. This evidence does not indicate if cities' individual rankings themselves were stable, as cities relative positions could have been changing substantially within the distribution even though the distribution itself was quite stationary.

Quah (1993) provides a statistical method that we use to make further observations on the cross-section dynamics of relative levels and growth rates of income per capita in terms of evolving distributions. We apply his method to our samples of French and Japanese urban agglomerations. Following Quah's procedure, we group the two samples of cities into 6 cells defined according to a division of population sizes relative to the average population for the respective period.

[23] The data for this regression are from 1982, and are available for only 20 agglomerations. A city's level of education plays a key role in the model that we develop below. Unfortunately we could not find data on education for our units of observation for either country.

The frequency distribution is assumed to follow a first-order Markov transition process. We describe the evolution of this process with a transition matrix each element of which is the probability that a city initially in the cell corresponding to its column will join the cell corresponding to its row in the subsequent period. (Columns sum to 1.) As the diagonal elements of the matrix approach 1 the pattern of growth converges to one of exact parallel growth. We find that these elements are not exactly 1, however. Two question then arise: (1) What is the propensity of cities in each cell to move into other cells? (2) How does the long-run frequency distribution implied by the transition process differ from what it has been historically? Our goal is to estimate the transition probability matrix and the consequent long-run city size distribution.

To examine the sensitivity of the results to the particular cell division, we tried different assignments.[24] Alternative cell divisions provided very similar results.

Our division classifies cities according to whether their population fell in the range: (1) less than 0.30 of the mean, (2) between 0.30 and 0.50 of the mean, (3) between 0.50 and 0.75 of the mean, etc. These tables show the frequency distribution and numbers corresponding to each cell. The stability of these distributions across time and the similarity between France and Japan is, of course, another manifestation of the observations we made about the Lorenz curves.

We define F_t as a 6×1 vector indicating the frequency of cities in each cell at time t. We assume that F evolves according to:

$$F_{t+1} = MF_t \tag{1}$$

where M is a 6×6 transition probability matrix, mapping the assignment from period t into an assignment in the subsequent period. The s-period-ahead predictor for the distribution is thus:

$$F_{t+s} = M^s F_t.$$

Taking s to ∞ we can characterize the long-run (ergodic) distribution of F_t, defined as F_∞ (if it exists and is unique).

Defining $M_{i,i+1}$ as the actual transition matrix from period i to period $i+1$, we have estimated the matrix M by computing the average $M_{i,i+1}$ for all the periods in the sample. The estimated M matrices for France and Japan are given in Table 4 Table 5, respectively. The large values of the diagonal terms and the many low values and zeros of the off-diagonal terms of both matrices indicate high persistence. For France, diagonal terms tend to increase with relative size, indicating more persistence for larger cities. For Japan, the diagonal terms are higher for small and large cities than for medium-sized cities. Overall, the values on the diagonal are higher for Japan than for France and there are no off-diagonal terms for the highest cell in the Japanese matrix. These results suggest that there is

[24] Table 2 and Table 3 of the working paper version of the paper (Eaton and Eckstein, 1994) give the cell divisions for the results here.

Table 4
Average transition matrix for French cities (1876–1990)

	Cell's upper endpoint					
	0.3	0.5	0.75	1	2	20
0.3	0.723	0.154	0	0	0	0
0.5	0.254	0.796	0.118	0	0	0
0.75	0.024	0.05	0.741	0.275	0	0
1	0	0	0.14	0.692	0.083	0
2	0	0	0	0.033	0.792	0.097
20	0	0	0	0	0.125	0.903

more persistence in the Japanese data, with no movement in and out of the cell containing the largest 3 cities, and less movement among the smaller cities than is the case for France.[25]

We obtain the ergodic probability distribution by taking the average of the implied ergodic distribution from each date in the sample using the estimated average period-to-period transition matrix. That is, using the estimated M we calculate F^{∞} by first calculating, for each year i in the data:

$$F_{\infty}^{i} \equiv M^{\infty} F_{i},$$

and then estimate F_{∞} as the simple average of F_{∞}^{i}. These values are reported in Table 6 Table 7 for France and Japan, respectively. Concentration of the frequencies around 1 would imply convergence to the mean. The results show no such convergence for either country. For France, about 90% of the cities will be below the average. Only a few cities will be above the average size. About 59% of the cities will be less than one half of the average size. The results for Japan are similar: 88% of the cities are below average and 75% are below one half of the average. Hence, the dispersion of population in Japan is expected to be less equal

Table 5
Average transition matrix for Japanese cities (1925–1985)

	Cell's upper endpoint					
	0.3	0.5	0.75	1	2	20
0.3	0.888	0.114	0	0	0	0
0.5	0.112	0.855	0.152	0	0	0
0.75	0	0.032	0.776	0.111	0	0
1	0	0	0.072	0.826	0.069	0
2	0	0	0	0.063	0.931	0
20	0	0	0	0	0	1

[25] The choice of time period (e.g., beginning to end of sample) affects the estimated M matrix. The M matrix turned out to be similar under alternative specifications.

Table 6
Average of ergodic probabilities for French cities

	Cell's upper endpoint					
	0.3	0.5	0.75	1	2	20
Frequency	8.229	14.898	7.956	4.095	1.677	2.145
Probability	0.211	0.382	0.204	0.105	0.043	0.055

than that for France (as was also suggested by the movement of the Lorenz curves).

The predicted relative size of Paris will be smaller and there will be fewer cities that are large. That is, compared with the historic relative frequency for France, the ergodic frequency has more weight on the smaller cells. In 1990 the larger than average cities in France are about 15% of the total and their share will drop to about 10%. For Japan, as we explained above, there is almost no change at the top but a small increase in the share of the smaller cities relative to 1985 distribution.

Hence, the evidence rejects a divergence hypothesis and it is consistent with parallel growth with a somewhat less disperse distribution for France. In particular, Paris and the other larger cities would be somewhat closer to the average size. Similarly, Japan displays no evidence of convergence. In summary, the data support the view that a wide range of city sizes will persist.

3. Growth in a system of cities

We now present a model of a system of cities that captures our empirical findings on the size distribution of French and Japanese cities, and correlations between city size and other city characteristics. Our theory has the implication that cities' populations converge to a common growth rate, with different relative populations. A city's relative size depends upon its productivity as a place to acquire human capital. In steady state, wages per worker are higher in larger cities because the level of human capital per worker is higher. City populations adjust to remove any incentive to migrate.

We describe each of K individual cities in terms of a representative resident who lives, works and learns in that city. We begin at the level of the individual,

Table 7
Average of the ergodic probabilities for Japanese cities

	Cell's upper endpoint					
	0.3	0.5	0.75	1	2	20
Frequency	15.08	14.88	3.12	2.04	1.84	3
Probability	0.377	0.372	0.078	0.051	0.046	0.075

examining the optimization problem facing a resident of a particular city. We then describe the typical city itself. We first model the city's static production technology, and derive the equilibrium relationships between wage, population and city area.[26] We then characterize how a city grows as its residents acquire human capital. Finally, we describe growth in a system of interconnected cities. We first examine how productivity growth in different cities interact, treating individuals as fixed in a particular city. Finally, we consider the incentive to migrate and its consequences for relative city size.

3.1. Individual optimization

An individual's utility depends only on lifetime consumption. Hence individuals choose where to live only on the basis of the implications for what they can consume over their lifetimes. As in Yaari (1965); Blanchard (1985), an individual k in city i faces a constant hazard ϕ of dying and maximizes an objective function:

$$V_{ki0} = \int_0^\infty \exp[-(\rho + \phi)t] \ln(c_{kit})\mathrm{d}t, \; \rho,\phi \geq 0, \tag{2}$$

where c_{kit} is period t consumption and ρ is a subjective discount factor.[27] Individuals can annuitize their wealth and so face the intertemporal budget constraint:

$$\int_0^\infty \exp[-\int_0^t (r_s + \phi)\mathrm{d}s]c_{kit}\mathrm{d}t \leq \int_0^\infty \exp[-\int_0^t (r_s + \phi)\mathrm{d}s]w^*_{it}h_{kit}e_{kit}\mathrm{d}t + a_{kio} \tag{3}$$

where r_s is the interest rate (which we treat as common to all cities) at time s, w^*_{it} is the after-tax wage rate per unit of effective labor in city i at time t, h_{kit} is the individual's level of human capital, e_{kit} is work effort and a_{kio} is the individual's financial wealth at period 0. While the individual takes r_s and w_{it} as given, we determine these magnitudes endogenously in the steady state which we consider below.

As in Lucas (1988) workers must take time off from work in order to acquire human capital. Workers have a time endowment of 1 out of which to choose time at work e_{kit}. The rest of the time they learn, adding to their human capital. Human capital accumulates according to the relationship:

[26] Our static model of production takes much from the literature on circular cities. See, for example, Mills (1967); Helpman and Pines (1980); Henderson (1987).

[27] At some notational expense, the analysis generalizes to the case in which period utility exhibits constant elasticity of the marginal utility of consumption. A constraint is that the elasticity must exceed an amount between 0 and 1 (see Lucas (1988)). In our case the elasticity is 1 so that this constraint is satisfied.

$$\dot{h}_{kit} = H^*_{it}(1 - e_{kit}) \tag{4}$$

where H^*_{it} represents the return to learning in city i at time t. We think of H^*_{it} as the knowledge base upon which the residents of city i draw when they study. Hence the larger this base the more effective is the time spent learning. We relate the knowledge base in any city i to the average level of human capital in that and in other cities. Specifically, if there are K cities each with an average level of human capital H_{jt}, $j = 1,...K$, then:

$$H^*_{it} = \sum_{j=1}^{K} \delta_{ij} H_{jt} \tag{5}$$

where $\delta_{ij} \geq 0$ is the contribution of city j's human capital to the knowledge base of city i. If city i has relatively larger values of δ_{ij} then it is a more productive place to learn than other cities, while if city j has relatively larger values of δ_{ij} then its human capital contributes more to learning than other cities'. To the extent that the matrix of δ's is diagonal dominant then growth is generated primarily by local factors. Less diagonal dominance implies more pooling of knowledge across cities. For example, the notion that a primate city serves as the sole source of external knowledge for all other cities can be captured by setting $\delta_{ij} = 0$ for all $i \neq j$, except for $j = 1$, where 1 is the index of the primate city.

We assume that the average levels of human capital in each city, and hence the knowledge bases in each city, grow at a common rate g_{H*T}, which for now we treat as exogenous. Below we show that, under a broad set of conditions, human capital in each city will, in steady state, grow at a common rate. We relate the steady-state growth rate to underlying parameters of the system.[28]

Thus, the problem facing an individual is to choose at each moment t a level of consumption c_{kit} and a work effort e_{kit} to maximize the objective function Eq. (2) subject to the intertemporal budget constraint Eq. (3) and the equation of motion for human capital Eq. (4), given the individual's initial level of human capital h_{ki0}. The first-order conditions for an optimum are the standard one that individual consumption grow according to the relationship:

$$g_{ct} = r_t - \rho \tag{6}$$

We define the value of the individual k's human wealth at time t, i.e., the discounted present value of expected future labor income, as b_{kt}, where

[28] Lucas (1988) treats what we call H^*_{it} as proportional to the individual's own level of human capital. We could allow it to depend on the individual's own level of human capital as well as on the average levels in surrounding cities. As will become clear below, however, an essential aspect of our model is that H^*_{it} depends on human capital elsewhere.

$$b_{kit} = \int_t^{\infty} w_{is}^* h_{kis} e_{kis} \exp[-\int_t^{s} (r_u + \phi) du].$$

Optimal consumption calls upon the individual to consume a constant share $\rho + \phi$ of human and financial wealth, i.e.,

$$c_{it} = (\rho + \phi)(b_{is} + a_{kit})$$

At an interior point at which the individual is both working and learning:

$$r_t + \phi + g_{H*t} = g_{w*t} + \frac{H_{it}^*}{h_{kit}}. \tag{7}$$

(Here the term g_x denotes the growth rate of variable x.) Note that Eq. (7) is independent of e, time spent working. If the left-hand side of this expression exceeds the right then the individual works full time, accumulating no human capital, while if the right-hand side exceeds the left then the individual studies full time.

An implication for a city in which the average worker (with human capital H_i) both works and learns is assimilation: The human capital of all individuals in the city will eventually converge to the city average. This result follows from the observation that, since Eq. (7) holds for the average individual, then:

$$r_t + \phi + g_{H*t} = g_{w*t} + \frac{H_{it}^*}{H_{it}}. \tag{8}$$

Consider someone with $h_{kit} > H_{it}$. This person will only work, so that the city average level of human capital will eventually catch up. Conversely, anyone for whom $h_{kit} < H_{it}$ will only study, and consequently acquire human capital at a faster rate than the city average. In either case the individual's level of human capital converges to the city average, at which point the individual will behave like everyone else.

3.2. The city

We now relate the parameters facing an individual to characteristics of the city. To focus on the issues at hand, we assume that labor and land are the only factors of production.[29] These factors produce a single kind of output, which has a price one each period. Production is Cobb-Douglas in labor and land, with land having a share β. We introduce the notion of urban land scarcity in a very simple way. Total factor productivity diminishes with distance from the city center. Denoting the

[29] As long as capital is perfectly mobile across cities, its incorporation into the analysis is largely inconsequential, although its presence complicates the analysis.

physical labor force in city i at time t as N_{it}, the average level of human capital in that city as H_{it}, and the average time worked per worker as e_{it}, the effective labor supply in city i at time t is $L_{it} = e_{it}N_{it}H_{it}$.

We follow Lucas (1988) in assuming that the average level of human capital in a community H_{it} influences total factor productivity there, although we treat the community as the city rather than the country.[30] Specifically, output a distance d from the center, Q_{idt} is:

$$Q_{idt} = H_{it}^{\gamma} d^{-\epsilon}(2\pi d)^{\beta} L_{idt}^{1-\beta} \gamma,\ \beta,\ \epsilon \in (0, 1) \tag{9}$$

where γ relates city i's average level of human capital to its total factor productivity, L_{idt} is the effective labor supply a distance d from the center. Hence total factor productivity a distance d from the city center is $H_{it}^{\gamma} d^{-\epsilon}$, where parameter ϵ is the elasticity of total factor productivity with respect to distance from the city center. We restrict the land share β to exceed $\epsilon/2$. Otherwise, the population would concentrate at a spike in the center, as shown by Eq. (12) below.

Factors earn their pretax marginal products, so that payments to land and labor exhaust output. Worker mobility within the city establishes a common pretax city wage w_{it}. For the private marginal product of labor to be equal across the city the amount of labor working a distance d from the center must be:

$$L_{idt} = 2\pi\{H_{it}^{\gamma} d^{(\beta-\epsilon)}[(1-\beta)/w_{it}]\}^{1/\beta}. \tag{10}$$

The corresponding land rent is:

$$r_{idt} = \beta\{H_{it}^{\gamma} d^{-\epsilon}[(1-\beta)/w_{it}]^{1-\beta}\}^{1/\beta},$$

which falls as distance from the center increases.

We assume that it costs r_a to use land for urban purposes (which we treat as constant across time and cities).[31] The distance from the center at which r_{idt} has fallen to r_a establishes the city limits D_{it}. This condition relates the radius of the city D_{it} to the wage w_{it} and to the average level of human capital H_{it}. Solving for w_{it}:

$$w_{it} = (1-\beta)[H_{it}^{\gamma} D_{it}^{-\epsilon}(\beta/r_a)^{\beta}]^{1/(1-\beta)}. \tag{11}$$

Hence a city's wage increases with its average level of human capital H_{it}, but falls with its area D_{it} (or equivalently, as we show next, with its labor force L_{it}).

[30] Again, allowing productivity to depend on the level of human capital elsewhere, as well as on the average city level, would not affect our basic results. What is key is that the city average matters.

[31] We interpret this cost as reflecting such things as lighting, street maintenance, water, sewerage, etc. Introducing this cost has the same implications for nearly all of what follows as the more standard assumption that land has a nonurban alternative use that yields r_a. We choose to make the cost of using land for urban purposes as a direct rather than as an opportunity cost because it is convenient to treat the supply of land for potential urban use as infinite, but we do not wish nonurban land to generate infinite income.

Substituting Eq. (11) into Eq. (10) and integrating across all urban locations from the city center to the city limits gives an expression for the urban labor force L_{it} in terms of city radius D_{it} and the average human capital level H_{it}:

$$L_{it} = \frac{2\pi\beta}{2\beta - \epsilon} [r_a D_{it}^{\Gamma} / (\beta H_{it}^{\gamma})]^{1/(1-\beta)} \tag{12}$$

where $\Gamma \equiv 2(1-\beta)+\epsilon$.

Integrating the land rent r_{idt} between the city center and city limits D_{it} implies that the total return to land in city i at time t (net of the cost of maintaining the land $r_a \pi D_{it}^2$) is:

$$R_{it} = \frac{\epsilon\pi}{2\beta - \epsilon} \{\beta^{2\beta} r_a^{\epsilon - 2\beta} H_{it}^{2\gamma} [(1-\beta)/w_{it}]^{2(1-\beta)}\}^{1/\epsilon} \tag{13}$$

Finally, Eqs. (11) and (12) together provide an expression for the wage in terms of the average level of human capital and the effective labor force, with the area determined endogenously:

$$w_{it} = (1-\beta)\{[2\pi\beta/(2\beta-\epsilon)]^{\epsilon} (\beta/r_a)^{2\beta-\epsilon} H_{it}^{2\gamma} L_{it}^{-\epsilon}\}^{1/\Gamma}. \tag{14}$$

Eq. (14) relates the wage per unit of effective labor in a city to the city's effective labor force and its average level of human capital. As one would expect, the wage in a city is higher the higher the average level of human capital and the smaller the labor force.

Total city income net of maintenance costs consists of wage plus (net) rental income $Y_{it} = w_{it}L_{it} + R_{it}$. Substituting Eq. (14) into Eq. (13) implies that the ratio of land rent to wage income, defined as η, is:

$$\eta = \frac{\epsilon}{2(1-\beta)}$$

We assume that rents are taxed at 100% and redistributed as a proportional wage subsidy. Hence $w_{it}^* = w_{it}(1+\eta)$.[32] Since the ratio of pretax land rents to labor income is constant over time so is the percentage rate of the subsidy.

3.3. City dynamics

We now characterize the steady-state dynamics of an individual city. An individual city is in steady state when: (i) all residents' human capital levels have

[32] Making the more natural assumption that land is untaxed and constitutes annuitized financial wealth complicates the analysis without affecting our basic conclusions. The problem is that introducing nonzero net nonhuman wealth introduces an additional nonlinear relationship between the interest rate and the value of land (see Blanchard, 1985). Hence, we assume that the tax system converts land income into human wealth.

converged to the citywide average, (ii) individuals are working a constant amount of time e_i, (iii) total consumption and total income grow at a constant rate g_Y, and (iv) wages grow at a constant rate g_w, and (v) human capital grows at constant rate g_H.

New individuals are born in the city at a constant rate $g_N + \phi$. Since individuals die at rate ϕ the natural population growth rate is g_N. We assume that individuals inherit the human capital of their parents, but not any financial wealth. The assumption that individuals are entering and leaving the economy is convenient for modelling the migration decision below.

Since there is no net financial wealth, aggregate consumption per worker grows at the same rate as individual consumption, and is equal to the growth in the wage per effective worker plus the growth in human capital per worker.[33] Hence:

$$g_c = g_w + g_H.$$

(Since our focus in this section is what goes on in an individual city, to reduce clutter we suppress the city subscript i.) Combining this steady state relationship with Eqs. (6) and (8) we get:

$$\rho + \phi + g_H + g_{H*} = \frac{H_t^*}{H_t}. \tag{15}$$

Incorporating the expression for effective labor into expressions Eqs. (12) and (14), and differentiating with respect to time (with ϵ held constant) gives an expression for the steady-state growth in the wage per unit of effective labor:

$$g_w = \frac{(2\gamma - \epsilon)g_H - \epsilon g_N}{\Gamma}. \tag{16}$$

An implication of Eqs. (7) and (16) is that the growth rate of per capita output and consumption, is:

$$g_c = g_w + g_H = \frac{2(1 - \beta + \gamma)g_H - \epsilon g_N}{\Gamma}.$$

While more rapid accumulation of human capital can lower growth in the wage per unit of effective labor, it always means higher consumption growth. Whether per capita consumption grows or falls over time depends upon whether the effect of human capital accumulation overcomes the congestion effects of population growth.

[33] Here is where our assumption that land rents are taxed to subsidize wages simplifies things. If, instead, land constitutes annuitized nonhuman wealth the relationship between the interest rate and aggregate per capita consumption growth is: $r = [\rho + 2g_c + g_N + \sqrt{(\rho - g_N)^2 + 4(\rho + \phi)(g_N + \phi)\eta/(1 + \eta)}]/2$, which reduces to $r = \rho + g_c$ when $\eta = 0$ or when land rents are taxed to subsidize wages, as we assume here.

Together Eqs. (16) and (8) give us:

$$r + \phi + g_H = \frac{(2\gamma - \epsilon)g_H - \epsilon g_N}{\Gamma} + \frac{H_t^*}{H_t}$$

so that a steady state requires $g_H = g_{H*}$. (While here we treat g_{H*} as an exogenous constant, in the next section we relate the steady-state value of g_{H*} to growth in the entire system of cities.) The steady state relationship between H_{it} and H_{it}^* is thus:

$$\frac{H_t^*}{H_t} = r + \phi + \frac{2(1 - \beta + \epsilon - \gamma)g_{H*} + \epsilon g_N}{\Gamma}$$

The level of human capital is lower relative to the knowledge pool the higher the interest rate, the probability of death, the growth rate of the knowledge pool and the population growth rate. Eq. (15) reduces to:

$$\rho + \phi + 2g_{H*} = \frac{H_t^*}{H_t} \tag{17}$$

Given the knowledge base H_t^*, the average level of human capital is higher the lower the discount factor and the greater life expectancy.

3.4. Growth in a system of cities

We consider the system of K cities to be in steady state when: (i) each city is in steady state, (ii) the cities' levels of human capital grow at a common rate g_H, and (iii) no one has an incentive to move. The second criterion implies that the left-hand side of Eq. (17) is the same for all cities in the system. We can then incorporate Eq. (5) into Eq. (17) and express the relationship between the level of human capital in each city and the common growth rate of human capital in terms of the system of linear differential equations:

$$\lambda H = \Delta H \tag{18}$$

where $\lambda \equiv \rho + \phi + 2g_H$, H is the vector of city-level human capital $\{H_1,...,H_K\}'$, and

$$\Delta \equiv \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1K} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ \delta_{K1} & \delta_{K2} & \cdots & \delta_{KK} \end{bmatrix}$$

the matrix of city interaction effects.

Steady-state growth requires that this system have an eigenvalue that exceeds $\rho + \phi$ whose corresponding eigenvector is nonnegative. Since Δ is nonnegative, Frobenius' theorem ensures that if Δ is also indecomposable then it has a real eigenvalue λ^F (the Frobenius root) where:

(i) λ^F is real and strictly positive;
(ii) associated with λ^F is an eigenvector $H^F > 0$ which is unique up to a scalar multiple;
(iii) λ^F is the only eigenvalue of Δ that has an associated eigenvector that is nonnegative;
(iv) λ^F is the largest eigenvalue in absolute value;
(v) λ^F is increasing in each element of Δ.[34]

These properties ensure that if Δ is indecomposable and its elements are sufficiently large then there exists a unique steady-state growth rate of human capital $g_H = [\lambda^F - (\rho + \phi)]/2$. The corresponding eigenvector H^F gives the relative steady-state levels of human capital. More knowledge spillovers imply higher growth. If, however, the elements of Δ are so small that $\lambda^F < \rho + \phi$ then no learning or growth occurs in steady state.[35]

Note that λ^F and H^F depend only on the parameters of the knowledge spillover matrix. An implication is that the relative levels of human capital depend only on these parameters as well, and that the growth rate of human capital depends only on these parameters and on the discount factor $\rho + \phi$.

In the special case of two cities ($K = 2$):

$$\lambda^F = \frac{\delta_{11} + \delta_{22} + [(\delta_{11} - \delta_{22})^2 + 4\delta_{12}\delta_{21}]^{1/2}}{2}$$

$$\frac{H_1^F}{H_2^F} = \frac{\delta_{11} - \delta_{22} + [(\delta_{11} - \delta_{22})^2 + 4\delta_{12}\delta_{21}]^{1/2}}{2\delta_{21}}.$$

Other things equal, the city in which the contribution of existing knowledge to learning is greater has a relatively higher level of human capital.

So far we have established conditions for the parallel growth of human capital

[34] See McKenzie, 1960, Lemma 1, or Takayama, 1974, Theorem 4.B.1.

[35] The matrix is indecomposable if there is no ordering of its elements that allows it to be partitioned as:

$$\Delta = \begin{bmatrix} \Delta_{11} \Delta_{12} \\ 0 \, \Delta_{22} \end{bmatrix}.$$

If Δ is decomposable then the cities corresponding to the elements of Δ_{22}, what we call the isolated set of cities, do not receive spillovers from any city corresponding to the elements of Δ_{11}. Even if Δ is decomposable, however, it still has a Frobenius root λ^F if Δ_{12} has at least one strictly positive element (so that the set of cities corresponding to the elements of Δ_{11} are nonisolated) and Δ_{11} and Δ_{22} have Frobenius roots λ_1^F and λ_2^F such that $\lambda_2^F \geq \lambda_1^F$. In this case $\lambda^F = \lambda_2^F$: The levels of human capital in the entire set of cities grow at the same rate as in the isolated set of cities left on their own. In this particular case the isolated cities are 'leaders' in that these cities drive the growth of the others. If, however, $\lambda_2^F < \lambda_1^F$ and $\lambda_1^F > \rho + \phi$ then the isolated cities eventually grow more slowly than the nonisolated set, and get left behind.

across cities. We do not, however, have observations on the human capital levels of the cities but on their populations. Our theory implies that if the population of each city is also growing at a common rate, then so are wages and consumption, with parameters of technology linking the growth of human capital and population, on one hand, to the growth of wages and consumption, on the other.

Individual migration decisions link the distribution and growth of human capital to the size distribution of cities. We now turn to the migration decision of the representative resident of each city, and consider conditions under which the parallel growth of human capital levels across cities implies parallel growth of populations.

3.5. Migration and city size

We now turn to the third criterion for a system of cities to be in steady state, that relative populations remove any incentive to migrate. To capture the notion that moving between cities imposes a large sunk cost, we assume that individuals can choose where among the K cities to take up residence only at the beginning of their lives. Since new individuals enter the labor force at rate $g_N + \phi$, this is the rate at which individuals in the population have the choice of migrating. Since we assume that these new individuals inherit their parents' levels of human capital, in steady state new entrants have the average level of human capital in the city of origin. In steady state relative city populations adjust to remove the incentive for individuals to leave their native cities.

We denote the expected steady-state lifetime earnings of an individual with human capital h arriving at a city where the average level of human capital is H as $V(h,H)$. The expected discounted lifetime earnings at time 0 of a representative resident of a city with an average level of human capital $H_{i,0}$, is thus simply:

$$V(H_{i,0}, H_{i,0}) = e^* \frac{w_{i,0}H_{i,0}}{r + \phi - g_w - g_H} = e^* \frac{w_{i,0}H_{i,0}}{\rho + \phi} \tag{19}$$

where $e^* = (\lambda^{\mathrm{F}} + \rho + \phi)/2\lambda^{\mathrm{F}}$ is the steady-state work effort. To examine the incentive to migrate we compare this amount with expected earnings from moving elsewhere.

3.5.1. The city-specificity of human capital

An issue is the degree to which human capital acquired in one city raises an individual's productivity working elsewhere. At one extreme human capital may be perfectly general, and augment labor productivity by the same wherever the worker goes. At the other extreme, human capital may be totally city-specific, so that acquiring human capital in a city is worthwhile to an individual only if the individual uses it there.

We introduce the parameter φ to deflate human capital acquired elsewhere than in the city where it is used. A migrant from city i to j with an amount of human capital h acquired in the city of origin will arrive in city j with an amount of human capital φh. Hence a value of $\varphi=1$ implies that human capital is fully general, while $\varphi=0$ implies that it is completely city-specific.

In analysing a decision to migrate we need to distinguish between a move to a city where the migrant's human capital, upon arrival, will be lower than the average of the city of destination and a move to a city where the migrant's level of human capital will exceed the average. We undertake this analysis for an individual contemplating a move in an economy that is in steady state.

3.5.2. Moving up

Consider first an individual's decision about whether or not to migrate from city $i-1$ to city i, where $H_i>\varphi H_{i-1}$. The immigrant would arrive in city i with a lower level of human capital than the destination average. As we showed above, upon arrival the immigrant would learn full time until becoming assimilated with the representative resident in the destination city. Wage income during this training period would be zero, but once it was over the immigrant would resume working the steady-state amount.

The value to the potential migrant moving is:

$$V(\varphi H_{i-1,0}, H_{i,0}) = e^*\left[\frac{\lambda^F - g_H(\varphi/\eta_i)}{\lambda^F - g_H}\right]^{-(\rho+\phi)/g_H} \frac{w_{i,0}H_{i,0}}{\rho+\phi}$$

where $\eta_i=H_i/H_{i-1}$ is the ratio of human capital city at rank i to that of city at rank $i-1$ given by the solution to Eq. (18) and $g_H=[\lambda^F-(\rho+\phi)]/2$, the steady-state growth rate of human capital. Hence there is an incentive to migrate from city $i-1$ to city i if this expression exceeds Eq. (19). For the wage differential to remove this incentive, the physical labor force of city i must exceed that of $i-1$ by at least the ratio:

$$\frac{N_i}{N_{i-1}} = \eta_i^{2(1-\beta+\gamma)/\epsilon}[(\lambda^F - g_H)/(\lambda^F - g_H\varphi/\eta_i)]^{(\rho+\phi)\Gamma/g_H\epsilon}. \tag{20}$$

This expression is a lower bound on how much the size of a city with high human capital must exceed that of a city with low human capital to remove the incentive to migrate. If this bound is violated, workers in the lower human capital city have an incentive to migrate to the higher human capital city.

3.5.3. Moving down

Eliminating the corresponding incentive to move from a high to a low human capital city places an upper bound on how much the size of the high human capital city can exceed that of the low human capital city. To establish this bound consider now an individual in city i contemplating a move to city $i-1$, where $\varphi H_i>H_{i-1}$.

Since the immigrant arrives with a level of human capital that exceeds the average in the destination city, the immigrant's incentive is to work full time, not learning at all, until the human capital level in the city has caught up to the immigrant's level. The potential migrant now compares the value of staying home, given by Eq. (19), with that moving to city $i-1$. The value of making this move permanently is:

$$V(\varphi H_{i-1,0}, H_{i,0}) = [1 - (\varphi\eta_i)^{-r'/g_H}] \frac{w_{i-1,0}\varphi H_{i,0}}{r'} + e^*(\varphi\eta_i)^{-r''/g_H} \frac{w_{i-1,0}H_{i-1,0}}{r''}$$

where again $\eta_i = H_i/H_{i-1}$ and $r' = r + \phi - g_w$ and $r'' = \rho + \phi$.

To remove the incentive to migrate from city i to city $i-1$, the maximum amount by which city i's physical labor force can exceed that of city $i-1$ is:

$$\frac{N_i}{N_{i-1}} = \eta_i^{2(1-\beta+\gamma)/\epsilon}[(1-\theta)(\varphi\eta_i)^{(\rho+\phi)/g_H} + \theta\varphi\eta_i]^{-\Gamma/\epsilon}. \tag{21}$$

where:

$$\theta \equiv \frac{\rho + \phi}{(\rho + \phi + g_H)e^*}.$$

A steady-state requires that the lower bound not exceed the upper bound. The interval of permissible city sizes falls as φ rises. In our simulation results the interval became negative with strictly positive discounting ($\rho + \phi > 0$) and values of φ very close to 1. The implication is that, with discounting, a steady-state outcome with multiple cities requires some element of city specificity of human capital (or some other form of moving cost).[36]

The rest of our discussion here posits that values of φ and ρ guarantee a steady state with no equilibrium migration. We discuss two particular cases in which a steady state is guaranteed.

3.5.4. Fully city-specific human capital

Say that $\varphi = 0$, so that knowledge acquired in one city has no value elsewhere. In this case there is no possibility of 'moving down,' since one arrives anywhere new with no human capital that is appropriate for the city of destination. Eq. (20)

[36] We are currently examining a model in which human capital is fully general. In steady state two-way migration occurs. Even without direct spillovers in human capital between cities, so that in isolation cities would grow at different rates, migration leads to parallel growth at typically different levels. The analysis is more complicated. For example, even in steady state cities have heterogenous populations.

evaluated at $\varphi=0$ is thus a lower bounds on the extent to which the area, effective labor force, and population of city i can exceed those of city $i-1$. The upper bound is given by the condition that $V(0,H_{i-1})<V(H_i,H_i)$, i.e., that it is not worthwhile moving to the lower human capital city to start over there. This upper bound necessarily exceeds the lower bound established by the condition that $V(H_{i-1},H_{i-1})>V(0,H_i)$.

3.5.5. Zero discounting

In the limiting case in which there is no time discounting ($\rho+\phi=0$) only steady-state human capital levels matter in determining comparisons across cities. The lower and upper bound on city labor forces converge to:

$$\frac{N_i}{N_{i-1}}=\eta_i^{2(1-\beta+\gamma)/\epsilon}.$$

In this case, then, relative steady-state city size leaves the residents of each city indifferent between staying put or moving to any other larger or smaller city. The city-specificity of human capital makes no difference since the destination city's level of human capital is always acquired in finite time. The larger size (or, equivalently, higher land rents) of cities where learning is more productive make them equally attractive as cities where learning is less productive.

The model thus provides an explanation for the observed stability of the relative populations of cities. The spillover of knowledge between and within cities determines the common rate of growth of total factor productivity of all cities, and their ranking. The migration decision then implies a restriction on the distribution of population among cities. Without discounting, the relative human capital level of a city uniquely determines its relative size. With discounting the growth rate of human capital, which is equal across cities, has an effect on the bounds of the city size as well. Simulation results indicate that very small differences in steady-state human capital levels (around 5%) can imply large differences in city populations and areas. Note that even though workers do have the option of migrating, differences in wages and in levels of human capital between cities persist in steady state. Hence the observation of such differences are not necessarily indicative of a lack of migration opportunities.

Higher costs of congestion are captured by higher values of ϵ. Raising ϵ does not change the distribution of human capital among cities, but the implied distribution of city sizes become flatter, so that the ratio N_i/N_{i-1} is smaller. Hence the shape of the Lorenz curves is affected by both the process of human capital accumulation and the parameters of the production function and preferences. However, if these parameters are stable, the dynamic structure of human capital accumulation alone can explain the stability of the Lorenz curve.

4. A numerical example

To illustrate the model we fit it to the Lorenz curve for the most recent year of data for France. Using only one cross section of data to fit the model we can at most solve for as many parameters as the number of cities. In the example here we set the parameters of preferences and technology to the following values a priori: $\rho=0$, $\phi=0$, $\beta=0.3$, $\gamma=0.1$, $\epsilon=0.4$, $\varphi=0$. We then seek city-specific values of the diagonal elements of Δ and a common value of the off-diagonal elements of Δ that best fit the Lorenz curve and a steady-state growth rate of human capital of 1.5%. The diagonal elements in each case were set at 0.0005. The Frobenius root implied by our estimates is 0.03150. The implied growth rates of human capital is 0.01575.

Table 8 reports the results. The first panel gives the actual relative population and the third the population implied by our calibration. The fourth panel reports the implied diagonal element corresponding to the city in question. The second gives the implied relative level of human capital. Note that relatively small

Table 8
Simulations: French cities 1990

Actual Lorenz curve						
0.00490	0.01040	0.01620	0.02230	0.02920	0.03670	0.04500
0.05350	0.06190	0.07040	0.07920	0.08810	0.09720	0.10630
0.11620	0.12630	0.13700	0.14780	0.15880	0.16990	0.18110
0.19240	0.20480	0.21860	0.23280	0.24730	0.26230	0.27900
0.29610	0.31390	0.33320	0.35510	0.37780	0.40650	0.43730
0.47950	0.53380	0.58940	1.00000			
Human capital: H vector:						
0.11679	0.12021	0.12181	0.12336	0.12721	0.12989	0.13632
0.13632	0.13402	0.13322	0.13376	0.13389	0.13519	0.13557
0.13922	0.13992	0.14228	0.14294	0.14390	0.14358	0.14326
0.14195	0.14728	0.15127	0.15235	0.15315	0.15446	0.15866
0.15960	0.16120	0.16450	0.16978	0.17132	0.18167	0.18497
0.20019	0.21347	0.21643	0.35376			
The Lorenz curve implied by the estimated parameters:						
0.00488	0.01035	0.01613	0.02220	0.02907	0.03653	0.04479
0.05319	0.06161	0.07007	0.07883	0.08769	0.09674	0.10580
0.11565	0.12570	0.13635	0.14710	0.15804	0.16909	0.18023
0.19148	0.20382	0.21755	0.23168	0.24610	0.26103	0.27764
0.29465	0.31236	0.33156	0.35335	0.37594	0.40450	0.43520
0.47732	0.53178	0.58931	1.00000			
Solution for the diagonal of delta (the off-diagonals are 0.0005):						
0.00610	0.00684	0.00717	0.00748	0.00822	0.00871	0.00981
0.00981	0.00943	0.00929	0.00938	0.00941	0.00962	0.00969
0.01027	0.01038	0.01074	0.01084	0.01098	0.01093	0.01088
0.01069	0.01146	0.01200	0.01214	0.01225	0.01241	0.01293
0.01305	0.01323	0.01361	0.01418	0.01434	0.01535	0.01565
0.01689	0.01783	0.01802	0.02345			

differences in the diagonal elements of the matrix and in steady-state human capital levels (varying across cities by at most a factor of about 3) are consistent with much larger differences (up to a factor of 25) in population. That is to say, a relatively small degree of heterogeneity in city characteristics can give rise to large differences in population.

The example illustrates how the model can fit the observed stability of the Lorenz curves and the rank-size rule during a long period when cities were growing substantially in population.

5. Conclusion

The French and Japanese experiences provide striking evidence that, while the process of development is associated with a significant increase in urban population, it has had little effect on the distribution of population among different urban areas. Nor does it appear to give rise to the creation of new cities. This finding suggests that the forces driving the process of industrialization are present in individual cities roughly in proportion to their initial populations.

In economies with labor mobility, per capita output or wages are a poor indicator of total factor productivity across regions.[37] We have developed an equilibrium model in which relative populations reflect total factor productivity differentials across cities. Hence, the observed parallel growth of urban populations in France and in Japan can be interpreted as evidence for parallel growth in total factor productivity across cities. The model also suggests how cities can serve as a fundamental force in the process of industrialization.

The structural link between the dynamic process of human capital growth and the growth and distribution of population among cities is the central aspect of the theory. This link implies that we can use the data on the populations of cities to measure the pattern of growth of total factor productivity. In this paper we use this implication to interpret the stability of the Lorenz curves.

Our theory, taken literally, would also imply that the matrix describing the transition of cities across relative sizes M in Eq. (1) is diagonal, at least in steady state. We found in Section 2, however, that the matrix M includes off-diagonal terms. A topic for future research is the development of a stochastic version of the model that could account for these transitions.

Our theory is compatible with many possible relationships among cities in a growing country. One possibility is that the matrix Δ of knowledge spillovers is

[37] Land rents would provide a much better measure. Unfortunately, data on land rents are rarely available comprehensively. An exception is Japanese data on land values by prefecture. Dekle and Eaton (1994) explore their implications for measuring agglomeration effects in manufacturing and in financial services. Roback (1982) estimates the effect of urban amenities on land rents by using residential rents, controlling for dwelling characteristics.

nearly diagonal, which would suggest that cities are largely 'self reliant' in terms of their growth. In this case the process of economic development is best thought of in terms of the city. Another possibility is that off-diagonal elements are large, in which case the growth of individual cities is highly interconnected. In this case the nation or some broader unit might provide the more appropriate unit of analysis. An important topic for future research, which would require much more detailed data on the growth of a system of cities, is to identify the extent to which factors affecting city growth are local or national in character.[38]

It would also be useful to know the extent to which our findings on the evolution of city size extend to other countries. Observers have found Zipf's law to apply to other countries (Mills, 1972), but the dynamics of the distribution have not been explored. In particular, cities in countries of recent settlement may have a greater propensity to switch rankings.

Our analysis also has implications for the relationship between migration and development. It suggests, for example, how changes in the way knowledge flows across locations, as well as changes in land shares and urban transportation technology and infrastructure, would affect migration patterns and relative city size. Extensions of the analysis could also have implications for the relationship between migration and human capital accumulation. The analysis above suggests that migrants moving toward more populated cities would tend to be less educated than average upon arrival, but would acquire human capital more quickly once they arrive.

Our model does not provide a theory of the formation of cities. While such a theory would be useful, a point of our empirical analysis is that over the long period we considered we did not observe the formation or elimination of cities in the countries we examined. This finding suggests that once an area is settled, expansion of existing cities dominates the creation of new cities as a means of accommodating expansion.[39] Economic growth by itself does not spawn the creation of new cities as some models suggest. It appears that new cities are created only when new territories open up. An implication is that, now that most of the world is settled, the set of existing cities will remain the major cities of the world for the foreseeable future.

[38] Coulson and Rushen (1993) provide a promising approach in this direction. They decompose the recent employment growth in the Boston area into factors associated (i) with the national economy, (ii) with defense spending, and (iii) because of its similar industrial composition, with the San Jose economy. An important topic for future research would be to integrate this type of analysis into a general system of city growth such as the one we present here.

[39] Becker and Henderson (1995) have recently modeled the determinants of city size and the number of cities, showing how the number of cities might expand or contract during economic development. Their model differs from ours in its implication that all cities have the same size, even though this size and the number of cities might change over time.

Acknowledgments

We thank Richard Arnott, Paul Beaudry, David Canning, Gerald Carlino, Sam Kortum, Ed Mills, Ed Prescott, Victor Rios-Rull and participants at presentations at Universitat Pompeu Fabra, Harvard University, the Federal Reserve Bank of Philadelphia, and Northeastern University for constructive comments, M. Minni of INSEE for his help in providing data on French agglomerations, and Aditya Bhattacharjea, Yuval Nachtom and Akiko Tamura for their excellent research assistance. Akiko Tamura constructed the Japanese urban agglomeration data from data on the population of Japanese cities.

References

Arrow, K.J., 1962, The economic implications of learning by doing, Review of Economic Studies 29, 155–173.

Ades, A.F. and E.J. Glaeser, 1995, Trade and circuses: Explaining urban giants, Quarterly Journal of Economics 110, 195–228.

Arnott, R., 1979, Optimum city size in a spatial economy, Journal of Urban Economics 6, 65–89.

Barro, R.J., 1991, Economic growth in a cross section of countries, Quarterly Journal of Economics 106, 407–444.

Becker, R. and J.V. Henderson, 1995, City formation, manuscript (Brown University).

Benhabib J. and M.M. Spiegel, 1994, The role of human capital in economic development: Evidence from aggregate cross-country data, Journal of Monetary Economics 34, 143–173.

Blanchard, O.J., 1985, Debts, deficits, and finite horizons, Journal of Political Economy 93, 223–247.

Beckmann, M.J., 1958, City hierarchies and distribution of city size, Economic Development and Cultural Change 6, 243–248.

Chenery, H.B. and M. Syrquin, 1975, Patterns of development: 1950–1970 (Oxford University Press, London).

Coulson, N.E. and S.F. Rushen, 1993, Sources of fluctuations in the Boston economy, manuscript (Pennsylvania State University).

Dekle, R. and J. Eaton, 1994, Agglomeration and the price of land: Evidence from the prefectures, NBER Working Paper No. 4781.

Eaton, J. and Z. Eckstein, 1994, Cities and growth: Theory and evidence from France and Japan, NBER Working Paper No. 4612.

Eaton, J. and S. Kortum, 1994, International patenting and technology diffusion, NBER Working Paper No. 4931.

Glaeser, E.L., H.D. Kallal, J.A. Scheinkman and A. Shleifer, 1992, Growth in cities, Journal of Political Economy 100, 1126–1152.

Gell-Mann, M., 1994, The quark and the jaguar (W.H. Freeman, New York).

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

Henderson, J.V., 1987, General equilibrium modeling of systems of cities, in: E.S. Mills, ed., Handbook of regional and urban economics, Vol. II (Elsevier Science Publishers, Amsterdam).

Henderson, J.V., 1988, Urban development: Theory, fact, and illusion (Oxford University Press, New York).

Helpman, E. and D. Pines, 1980, Optimal public investment and dispersion policy in a system of open cities, American Economic Review 70, 507–514.

Inada, K., 1969, Endogenous technical progress and steady-state growth, Review of Economic Studies 36, 99–108.

INSEE, Annuaire statistique de la France, various issues.

Ioannides, Y., 1994, Product differentiation and economic growth in a system of cities, Regional Science and Urban Economics 24, 461–484.

Jacobs, J., 1969, The economy of cities (Vintage Books, New York).

Jacobs, J., 1984, Cities and the wealth of nations (Vintage Books, New York).

Kanketsu Showa-kokusei Soran.

Kuznets, S., 1966, Modern economic growth (Yale University Press, New Haven).

Lucas, R.E., 1988, On the mechanics of economic development, Journal of Monetary Economics 22, 3–42.

Lucas, R.E., 1993, Making a miracle, Econometrica 61.

McKenzie, L., 1960, Matrices with dominant diagonals and economic theory, in: K.J. Arrow, S. Karlin and P. Suppes, eds., Mathematical methods in the social sciences (Stanford University Press).

Mills, E.S., 1972, Urban economics (Scott, Foresman, London).

Mills E.S., 1967, An aggregative model of resource allocation in a metropolitan area, American Economic Review 57, 197–210.

Miyao, T., 1987, Dynamic urban models, in: E.S. Mills, ed., Handbook of regional and urban economics, Vol. II, (Elsevier Science Publishers, Amsterdam).

Palivos, T. and P. Wang, 1994, Spatial agglomeration and endogenous growth, manuscript (Louisiana State University and Pennsylvania State University).

Parente, S.L. and E.C. Prescott, 1994, Barriers to technology adoption and development, Journal of Political Economy 102, 298–321.

Quah, D., 1993, Empirical cross-section dynamics in economic growth, European Economic Review 37, 426–434.

Rauch, J., 1993, Productivity gains from geographic concentration of human capital: Evidence from the cities, Journal of Urban Economics 34, 380–400.

Roback, J., 1982, Wages, rents, and the quality of life, Journal of Political Economy 90, 1257–1278.

Rodwin, L., 1970, Nations and cities (Houghton-Mifflin, Boston).

Romer, P., 1986, Increasing returns and long-run growth, Journal of Political Economy 94, 1002–1037.

Rosen, K.T. and M. Resnick, 1980, The size distribution of cities: An examination of the pareto law and primacy, Journal of Urban Economics 8, 165–186.

Simon, H., 1955, On a class of skew distribution functions, Biometrika 42, 425–440; reprinted as Chap. 9 in H. Simon, 1957, Models of man (John Wiley and Sons, New York).

Solow, R.M., 1956, A contribution to the theory of economic growth, Quarterly Journal of Economics 70, 65–94.

Summers, R. and A. Heston, 1991, The Penn World Table (Mark V): An expanded set of international comparisons, Quarterly Journal of Economics 106, 327–368.

Takayama, A., 1974, Mathematical economics (Dryden, Hinsdale, IL).

Uzawa, H., 1965, Optimum technical change in an aggregative model of economic growth, International Economic Review 6, 18–31.

Wheaton, W.C. and H. Shishido, 1981, Urban concentration, agglomeration economies, and the level of economic development, Economic Development and Cultural Change 30, 17–30.

Yaari, M.E., 1965, Uncertain lifetime, life insurance, and the theory of the consumer, Review of Economic Studies 32, 137–150.

Young, A., 1991, Learning by doing and the dynamic effects of international trade, Quarterly Journal of Economics 106, 369–406.

Zipf, G.K., 1949, Human behavior and the principle of least effort (Hafner Publishing Company, New York and London); reprinted 1965.