AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

HWWA HAMBURG

DISCUSSION PAPER

# City Size Distribution and Growth

**Andreas Kopp**

HWWA DISCUSSION PAPER

97

Hamburgisches Welt-Wirtschafts-Archiv (HWWA)
Hamburg Institute of International Economics
2000
ISSN 1432-4458

The HWWA is a member of:

- Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL)
- Arbeitsgemeinschaft deutscher wirtschaftswissenschaftlicher Forschungsinstitute (ARGE)
- Association d‘Instituts Européens de Conjoncture Economique (AIECE)

# City Size Distribution and Growth

**Andreas Kopp**

The paper has been presented at the HWWA-workshop „Agglomerations, Centres and Periphery“. Comments by the workshop participants are gratefully acknowledged.

This paper is part of the research programme „European Integration and Spatial Development“.

**HWWA DISCUSSION PAPER**

**Edited by the**
**PRESIDENTIAL DEPARTMENT**

Hamburgisches Welt-Wirtschafts-Archiv (HWWA)
Hamburg Institute of International Economics

Öffentlichkeitsarbeit
Neuer Jungfernstieg 21 – 20347 Hamburg
Telefon: 040/428 34 355
Telefax: 040/428 34 451
e-mail: hwwa@hwwa.de
Internet: http://www.hwwa.de/

Andreas Kopp
Hamburg Institute of Internatinal Economics
Telefon: 040/428 34 434
Telefax: 040/428 34 529
e-mail: kopp@hwwa.de

## Content:


**Abstract** 6

**Summery**

**1 Introduction** 7

**2 Externalities, urbanization and growth** 9

2.1 City size distribution and dynamic urbanization externalities 10

2.1.1 The static model of a city 11

2.1.2 Migration and city size 17

2.2 City size distribution, types of cities and localization externalities 19

2.2.1 The structure of cities 21

2.2.2 Urban Growth 24

2.3 Discussion 25

**3 Empirics of dynamic agglomeration economies** 26

**4 Conclusions** 28

## Abstract

We discuss theoretical approaches to study the relationship between the size distribution of a nation's cities and macroeconomic growth. The discussion is based on the hypothesis of the New Growth Theory that inter-personal spillovers of education and skills determine the long-run growth of the economy. Growth theory treats such externalities as being uniformly effective over national territories and completely internal to nation-state. This suggests a link to urban economics which has a long tradition of considering human capital externalities as driving forces of the growth of urban centers, with productivity increases inducing immigration. From the perspective of the urbanization literature long-run macroeconomic growth is thus determined by the functioning of cities as a catalysts for human capital accumulation. Theoretical avenues to the relationship between the development of cities of different sizes and aggregate growth can be distinguished according to whether the spillovers occur between different production sectors or just within a single sector. Empirical studies of the interrelationship have proven to be inconclusive. In view of the results of the empirical research, the character of inter-personal human capital spillovers depends much on the technological maturity of city industries and calls for taking account of industry-specific innovation cycles. More attention has also to be devoted to the question of the generality or sector-specificity of human capital in peculiar industries.

**Zusammenfassung:**

Das Diskussionspapier präsentiert Ansätze zur theoretischen Analyse der Beziehung zwischen der Größenverteilung der Städte eines Landes und dem volkswirtschaftlichen Wachstum. Die Bedeutung einer solchen Beziehung ist abgeleitet aus der Hypothese der Neuen Wachstumstheorie, daß Externalitäten der Humankapitalbildung der Wirtschaftssubjekte das langfristige Wachstum einer Volkswirtschaft bestimmen. Im Unterschied zur Wachstumstheorie standen diese Externalitäten in der Stadtökonomik seit langem im Mittelpunkt der Forschung, auf der Grundlage ihrer ausschließlich lokalen Wirksamkeit. Die theoretischen Ansätze zur Explizierung des Zusammenhangs zwischen der Stadtstruktur eines Landes und seinem langfristigen Wachstum unterscheiden sich in der Frage, ob die Externalitäten sektorunabhängig oder sektorspezifisch sind. Empirische Studien zu dieser Frage führen zu keiner Ablehnung einer der beiden Ansätze. Die Ergebnisse dieser Studien legen nahe, daß die theoretische Forschung Innovationszyklen einzelner Sektoren berücksichtigen muß. Darüber hinaus ist von Bedeutung, in welchem Verhältnis die in einem Sektor nachgefragten Qualifikationen sektor-spezifisch oder sektor-ungebunden sind.

# 1 Introduction

Why could it be interesting at all to study the relationship between the size distribution of cities of a country and the aggregate growth of a country? Is there reason to believe that it is useful to give up the restriction to the macroeconomic level of analysis?

The study of the relationship between the size distribution of cities is indeed justified by the dissatisfaction with the results of the recent theoretical and empirical work on growth. It appears to be part of the many attempts to go beyond looking at the primitives of stocks of factors of production as the sources of aggregate economic growth. Other 'ancillary variables' of economic growth which have received attention at the macroeconomic level are political stability, the income distribution, or the financial development (Benhabib/Spiegel 1997, Aghion et al. 1999) or, at the microeconomic level, the work organization of plants (Prescott 1997). To look at the influence of urbanization on aggregate growth originates, at least to some extent, from the implausibility of the nation state as the territorial entity of economic study. The problems arising therefrom have been alluded to in the new endogenous growth theory literature at an early stage (above all Lucas 1988). The consequences of turning to cities or the system of cities as the object of study of growth economics have only recently been made explicit and formalized.

With the upsurge of growth theory in the eighties there was also renewed interest in empirical studies of growth. Somewhat strikingly, the empirical literature employed the sources of growth accounting framework which was associated with the neoclassical model. Most prominently the studies of Mankiw et al. (1992), and Alwyn Young (1994, 1995) found that the East Asian growth miracles were due more to increases of labour and capital than to rising productivity. Krugman (1994) popularized this view and claimed that there are no growth miracles, that the rapid growth of the East Asian countries was nothing but the reflection of a rapid but temporary expansion of the physical and human capital employed.

Only recently have the results of these studies been questioned. Klenow and Rodriguez-Clare (1997) challenged the claim that (physical and human) capital accumulation would explain the bulk of cross-county differences in levels and growth rates of income. Neither Young nor Mankiw et al had looked at the cross country variation in growth rates. Physical and human capital accumulation could not explain these variations.

Klenow and Rodriguez-Clare (1997) calculated that 91 % of the cross country differences in growth rates were due to the growth of Total Factor Productivity. In depth investigations of whether human capital had been responsible for high income levels (or growth rates) or the other way round doubt that we can take the causal direction asserted in the growth accounting exercises for granted (Bils/Klenow 1998, Pritchett n.d.). The same has been found for physical capital (Blomstrom et al. 1996). In light of a theorem of Diamond/McFadden and Rodriguez (1978) it seems impossible that these disputes can be solved econometrically (cf. also Rodrik 1996 and Jones 1997).

One rather immediate way of looking at ancillary variables is to follow the perspective laid out by Lucas (1988) in emphasizing local knowledge spillovers as a possible source of long run growth. As he wrote in sec. 6 of the "Mechanics of Development":

> "My concern to this point has been almost exclusively with the aggregate mechanics of economic development, and I am afraid the discussion in these lectures will not get much beyond these mechanics. But I believe a successful theory of development (or of anything else) has to involve more than aggregative modeling, and I would like both to explain what I mean by this and to indicate where one might look to extend the analysis to a deeper and more productive level." (p.35), and
>
> "The particular aggregate models I have set out utilize the idea of human capital quite centrally, but assign a central role as well to what I have been calling the *external effects* of human capital. This latter force is, it seems to me, on a quite different footing from the idea of human capital generally: The twenty years of research I have referred to earlier is almost exclusively concerned with the *internal* effects of human capital, or with investments in human capital the returns to which accrue to the individual (or his immediate family)" (p.36), and further
>
> "I have been concerned with modeling the economic growth of *nations,* considered either singly, or as linked through trade."..."Most of our data come in the form of national time series, so 'fitting the facts' is taken to mean the nation is again the natural unit, for the most important fiscal and commercial policies are national and affect national economies in a uniform way. But from the viewpoint of a *technology* - like (11) -through which the average skill level of a group of people is assumed to affect the productivity of each individual within the group,

> a *national* economy is a completely arbitrary unit to consider." ... "The external effects that the term $h_a^{\gamma}$ in (11) is intended to capture have to do with the influences people have on the productivity of others, so the *scope* of such effects must have to do with the ways various groups of people interact, which may be affected by political boundaries but are certainly an entirely different matter conceptually." (p. 37)

These ideas had, however, been around for many years in the urbanization literature. Local externalities as the driving force of agglomerations had played a central role in the thinking on the static and dynamic functions of cities. Prior to the new growth theory it had focused on static externalities such as immediate information spillovers over market conditions. Two types of static externalities have been distinguished: localization economies in which a firm benefits from local firms in just the same industry, and urbanization economies, in which a firm benefits from overall local urban scale and diversity. Both types suggest advantages of being in bigger cities. The city sizes are limited by local congestion and commuting costs.

Dynamic externalities, by contrast, deal with the role of prior information accumulation in the local area on current *productivity* and hence employment. Such accumulations are fostered by a history of interactions and cultivated long-term relationships, which lead to a buildup of knowledge ("local trade secrets") available only to firms in a local area. As for their static counterparts, there are two types of dynamic externalities. In the terminology of Glaeser et al. (1992) dynamic extrnalities may be Marshall-Arrow-Romer (MAR) (localization) economies that derive from a buildup of knowledge associated with ongoing communications among local firms in the same industry or Jacobs (1984) (urbanization) economies, which derive from a buildup of knowledge or ideas associated with historical diversity.

## 2 Externalities, urbanization and growth

The theoretical relationship between the development of urban centers and aggregate economic growth goes beyond the parallelism between the sectoral transformation and the urban transition, which is much discussed in development economics (cf. Chenery and Squyrin 1975). It deals with questions of divergent, parallel or convergent sizes of cities in the growth process, or more specifically on theoretical foundations for the often

observed constancy of the city size distribution along the growth path of an economy.

All the stories told appear to have loose ends and rely on rather restrictive assumptions. They can broadly be distinguished according to the nature of the externalities which drive agglomeration processes, and how inter-urban interactions are considered. Multisectoral models emphasize dynamic (own-industry) localization effects (e.g. Black/Henderson 1999) with little inter-urban growth effects, while those models which emphasize general knowledge spillovers, have a stronger focus on the development of the overall city system (Eaton/Eckstein 1997). Both approaches on the relationship between aggregate growth and the city system are based on a combination of elements of the theory of circular cities (Arnott 1979, Henderson 1987, 1988) and the theory of endogenous growth (above all Lucas 1988 and Glaeser 1997). What the role of cities in increasing productivity is concerned, the primary informational role of cities may not be in creating cutting edge technologies, but rather in creating learning opportunities for everyday people. The idea is that dense urban areas ensure a faster rate of contacts between individuals, and each new contact provides an opportunity for learning.[1]

## 2.1 City size distribution and dynamic urbanization externalities

While much of the urban economics' literature has focused on the development of cities that specialize in the production of particular commodities, the emphasis on dynamic urbanization externalities was associated with the assumption that cities are typically not specialized and that their productivity derives from the interaction of individuals with complementary forms of knowledge. In sharp contrast to the literature focusing on dynamic *localization* externalities the formation of new cities is of no importance for the relationship between urbanization and growth. A growth process which is based on the growth of existing cities is called intensive growth, as opposed to an extensive growth process being based on te creation of edge cities. In the next subsection I present a static model of the city following the tradition of the theory of a circular city: Land is a factor of production, and total productivity within a city declines with the distance of production from the city center.

1 On the empirical evidence supporting such a view cf. Rauch (1993) and Glaeser/Maré (1994).

In the subsequent subsection the dynamic dimension is added. Total factor productivity in a city is related to its average level of human capital, as modeled in Lucas (1988) for the aggregation level of the nation state and alluded to for the city development. A basic characteristic of a city is the environment that it provides for the acquisition of human capital. 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 national pools of knowledge in the models of international diffusion of technical knowledge (e.g. Eaton/Kortum 1997). The interaction of the human capital stocks of different cities determines how the populations of individual cities will grow in the long run. Migration provides the link between the growth and distribution of human capital across cities and their relative populations. In subsection 2.1.3 I will show how migration occurs between cities of different relative levels of human capital.

### *2.1.1 The static model of a city*

The central feature of city is here that its 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 in larger cities. City populations adjust to remove any incentive to migrate.

An individual city is described in terms of a representative agent who lives, works and learns in that city. There are K such cities. We first characterize the equilibrium relationships between wage, population, and city area implied by the maximization problem of the representative resident.

The representative resident of each city supplies a unit of labor each period regardless of the wage. Residents behave dynastically, i.e. maximize over an infinite time horizon and maximize the following objective function:

$$V_0 = \int_0^\infty e^{-\rho t} \ln(c_t) dt, \ \rho \geq 0 \tag{1}$$

where $c_t$ is the consumption in period t, and $\rho$ the subjective discount rate.

What the production technology is concerned, land and labor are assumed to be the only factors of production. Output is homogeneous. The production function is of a Cobb-

Douglas type with the production elasticity of land being $\beta$. Total factor productivity of city i is $A_i$. The TFP index could reflect differences in technology between cities, different average levels of human capital of its work force, or city infrastructure. Here it reflects the average level of human capital. Land scarcity is captured by the assumption that productivity declines with distance from the city center. Output at a distance d from the city center $Q_d$ is then

$$Q_d = Ad^{-\varepsilon}(2\pi d)^{\beta} L_d^{(1-\beta)}, \ \beta, \varepsilon \in (0,1) \tag{2}$$

$2\pi d$ is land input, $L_d$ is labor supply at the distance d from the city center. $\varepsilon$ is the elasticity of total factor productivity with respect to the distance from the city center. Factors earn their marginal products, such that payments exhaust total production. Worker mobility within the city establishes a common city wage W. For the private marginal product of labor to be equal across the city the amount of labor working at a distance d from the center must be:

$$L_d = 2\pi\left\{Ad^{(\beta-\varepsilon)}\left[(1-\beta)/W\right]\right\}^{1/\beta} \tag{3}$$

The expression for the corresponding land rent is

$$R_d = \beta\left\{Ad^{-\varepsilon}\left[(1-\beta)/W\right]^{1-\beta}\right\}^{1/\beta} \tag{4}$$

The city size is determined by the rental rate of non-urban use of the land $R_a$, and the distance from the center D at which $R_d$ has fallen to $R_a$.

This condition relates the radius of a city D to the wage W and to the average level of productivity A. The derived expression for W shows that a city's wage increases with its total factor productivity but falls with its area D.

$$W = (1-\beta)\left[AD^{-\varepsilon}(\beta/R_a)^{\beta}\right]^{1/(1-\beta)} \tag{5}$$

Substituting (5) into (3) and integrating across all urban locations from the city center to the city limits gives an expression for the urban labor force L in terms of city radius D and total factor productivity A:

$$L = \frac{2\pi\beta}{2\beta-\varepsilon}\left[R_a D^{\Gamma}/(\beta A)\right]^{\frac{1}{1-\beta}}, \tag{6}$$

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

(5) and (6) together provide an expression for the wage in terms of productivity and the effective labor force, with the area determined endogenously:

$$W = (1-\beta)\left\{[2\pi\beta/(2\beta-\varepsilon)]^{\varepsilon}\left(\frac{\beta}{R_a}\right)^{2\beta-\varepsilon} A^2 L^{-\varepsilon}\right\}^{1/\Gamma} \tag{7}$$

That is, the wage rate has expected responses to variations in the productivity and the size of the labor force. The equation relates the wage per unit of effective labor in a city to the city's effective labor force and its total factor productivity. To construct the equilibrium configuration beween cities in addition to the intra-urban equilibrium conditions we need a theory of interurban migration.

***Migration***

The next step is therefore to show how the city level of total factor productivity is determined endogenously. The theory follows Lucas' model of endogenous growth in that total factor productivity growth is generated by investment in human capital. The accumulation of human capital follows from the dynamic optimization decisions of the representative agents of the individual cities.

$N_i$ denotes the physical labor force in city i, $H_i$ the average level of human capital in city i, and $e_i$ the average time worked there. The expression for effective labor then is $L_i = e_i N_i H_i$.

The total factor productivity is assumed to be a function of the average level of human capital in a city: $A_i = H_i^{\gamma}$, with $\gamma$ being a parameter relating city i's average level of human capital to its total factor productivity.

Differentiating expressions (6) and (7) with respect to time, we obtain for the growth of the wage per unit of effective labor:

$$g_w = \frac{(2\gamma-\varepsilon)g_H - \varepsilon g_N}{\Gamma}. \tag{8}$$

In the steady state the growth in the wage per physical worker which is equal to the growth rate in per capital output and consumption is

$$g_c = g_w + g_H = \frac{2(1-\beta+\gamma)g_H - \varepsilon g_N}{\Gamma}. \tag{9}$$

Whether per capita consumption increases or decreases over time depends on whether the effect of human capital accumulation overcomes the congestion effects of population growth.

As in the Lucas (1988) model, the workers have to take time off from work to acquire human capital. Workers split the endowment of one unit of labor into a fraction e for work and (1-e) for learning. The productivity of time spent learning increases with what has been learned already. In contrast to the original Lucas model it is assumed that the learning productivity depends not on the own, individual stock of human capital but on a linear combination of the average levels of human capital in each city, the "knowledge base" which can be accessed by the agent.

The extent to which an agent can access the knowledge which is seated in other cities thus defines how broad the basis for learning possibilities is. The individual human capital develops according to

$$\dot{h}_{kit} = \bar{H}_{it}(1-e_{kit}) \tag{10}$$

where $\bar{H}_{it} = \sum_j \delta_{ij} H_{jt}$ denotes the knowledge base of city i and $\delta_{ij}$ reflect the contributions of cities j to the learning possibilities in city i. That is, if a city has very good access to the human capital in other cities, it might be a very productive place to learn, even though its own average human capital may be small.

To determine the dynamics of a single city we assume that cities are connected by a common capital market, such that the individuals in all cities have the opportunity to lend and borrow at the same, constant interest rate r. Ignoring variable labor supply and risk, optimal learning decisions just maximize the discounted present value of wage income. Hence, what is of interest here is just the allocation of time between work and learning, which is independent of the decisions on consumption and saving. Consequently, in each period individual k in city i with human capital $h_{kit}$ chooses $e_{kit}$ to maximize

$$\int_0^\infty e^{-\rho\tau} w_{i\tau} h_{ki\tau} e_{kit} d\tau , \tag{11}$$

under the restriction of the equation of motion (10).

Solving the problem for individual k in city i implies that, in case of an interior solution where the individual both learns and works, we have

$$r + g_{\bar{H}_i} - g_{w_i} = \frac{\bar{H}_{it}}{h_{kit}} \tag{12}$$

The condition is independent of the time spent working. The reason is that the returns to working and learning are each linear in time spent in that activity. Hence, at an interior solution, an individual is indifferent between the two activities. If the individual is both learning and working, both activities must yield the same marginal, and hence average, return.

If the right hand side of (12) exceeds the left, then the average and marginal returns to learning exceed the returns to working so that the individual only learns, while if the left-hand side exceeds the right, the individual only works.

For the representative resident of a city $h_{kit} = H_{it}$. Therefore, for the representative agent in city i to be learning and working the above condition must be

$$r + g_{\bar{H}_{it}} - g_{w_i} = \frac{\bar{H}_i}{H_i} . \tag{12'}$$

A consequence of (12) and (12') is that an immigrant arriving in a city in which the representative resident is both working and learning will eventually become assimilated. An immigrant arriving with a level of human capital below the city average will find the reward to learning strictly greater than the reward to working, so will only lean. The immigrant's human capital will consequently grow faster than the city average until it gets equal to the city average. Conversely, an immigrant arriving with a level of human capital above the city average will find the reward to working strictly greater than the return to learning. As the immigrant's human capital stops growing, the city average will eventually catch up with that of the immigrant. Either way there is a process of complete assimilation.

***Steady state growth***

The steady state is defined as follows:

a) Each city's average level of human capital and population grow at the same rate.
b) Each individual's human capital corresponds to the city-wide average.
c) Individuals have no reason to migrate between cities.

If, as noted, within a city the average level of human capital will move toward the average for that city, criterion (a) implies that (12') holds for each city, with the same left-hand side applying for all cities. Hence the right hand side must be equal for all cities as well.

The nationwide interest rate r is determined by consumption decisions. With Bernoulli preferences the steady state interest rate equals the pure rate of time reference plus the growth in per captain consumption, $g_c$, which is equal to the growth in human capital plus the growth n the wage per unit of effective labor. Hence

$$r = \rho + g_c = \rho + g_H + g_w \tag{13}$$

Substituting (13) into (12') implies that, in steady state, for each city i, we have

$$\rho + 2g_H = \frac{\bar{H}_i}{H_i} \tag{14}$$

Since the left hand side holds for all cities, the system of equations for all cities can be written in matrix form as

$$\lambda H = \Delta H \tag{15}$$

where $\lambda = \rho + 2\ g_H$. H is the vector of city - level human capital and $\Delta$ is the matrix of city interaction effects. Steady state growth requires that this system has an eigenvalue (Froebenius root) that exceeds $\rho$ whose corresponding eigenvector is non-negative. If the interaction matrix is indecomposable, and the elements are sufficiently large, there exists a unique steady state growth rate of human capital $g_H = (\lambda^F - \rho)/2$, where $\lambda^F$ denotes the Froebenius root. The corresponding eigenvector $H^F$ gives the relative steady-state levels of human capital. If the elements of the interaction matrix are so small that $\lambda^F < \rho$ then no learning occurs in steady state and no growth. The fact that the

Froebenius root and the vector of eigenvalues depend only on the parameters of the knowledge spillover matrix implies that the levels and the growth rates of human capital depend in steady state on these parameters as well. The growth rate of human capital depends on these parameters and the discount factor $\rho$.

The conditions for the steady state growth imply that the cities necessarily grow in parallel, in terms of human capital. If there were also a parallel growth of the city population we would have the result of a parallel growth of wages and consumption. The city rank size distributions with respect to human capital, population, wages and consumption would be constant.

### *2.1.2 Migration and city size*

In this subsection we deal with the third criterion for the steady state, i.e. the representative resident of each city should have no incentive to migrate. This criterion determines relative city populations in steady state.

A central issue of the migration decisions is the degree to which the human capital which is individually acquired in one city contributes to the productivity of that individual in another city. $\varphi$ be a parameter which deflates human capital acquired elsewhere in the city where it is used. A value of $\varphi = 1$ implies that human capital is fully general, while $\varphi = 0$ means that it is completely city specific.

Migration decisions will depend on whether a worker moves to a city with a higher level of human capital or a lower level of human capital than in the city of origin. The decisions of moving up to a more human capital rich city or moving down is studied for an economy that is in steady state.

The migration decision of a worker is determined by where the present value of income, measured by a value function $V(H_i, H_i)$ is maximized. The index i orders cities according to their average levels of human capital. Consider an individual contemplating to move up from city i-1 to city i, i. e. $H_i > \varphi H_{i-1}$. An incentive to move up exists if

$$V(\varphi H_{i-1}, H_i) > V(H_{i-1}, H_{i-1}).$$

For the wage differential to remove that incentive, the radius of city i relative that of i-1

must be:

$$\frac{D_i}{D_{i-1}} = \left(\frac{H_i}{H_{i-1}}\right)^{(1-\beta+\gamma)/\varepsilon} \left[\left(\lambda^F - g_H\varphi\right) / \left(\lambda^F - g_H\varphi / \left(\frac{H_i}{H_{i-1}}\right)\right)\right]^{\rho(1-\beta)/g_H\varepsilon}$$

For ratios smaller than the one indicated by this equation the incentive to move up is removed. In larger cities there is a higher wage. Before enjoying this higher wage the migrant has to use the entire time budget to catch up with the average level of human capital in the city he or she migrated to. With the city size differences indicated by the above equation the trade-off is such that the cost of moving in the form of learning costs just compensates the expected future wage advantage.

We have similarly an incentive to move down if we have
$V(\varphi H_i, H_{i-1}) > V(H_i, H_i)$.

The incentive to move down is removed if the following condition holds:

$$\frac{D_i}{D_{i-1}} = \left(\frac{H_i}{H_{i-1}}\right)^{(1-\beta+\gamma)/\varepsilon} \left[(1-\theta)\left(\varphi\left(\frac{H_i}{H_{i-1}}\right)^{-\rho/g_H} + \theta\rho\left(\frac{H_i}{H_{i-1}}\right)\right)\right]^{(\beta-1)/\varepsilon},$$

where $\theta \equiv \frac{\rho}{(\rho - g_H)e^*}$ and e* denotes the steady state work effort. The wage level is smaller after moving but the migrant can work full-time and earn more than in the city of origin where he had to spend some time learning to keep up with the average level of human capital there.

The conditions for removing the incentive of moving up and moving down define a lower bound and an upper bound for the size differences, respectively. Sustainability in steady state requires that the lower bound is not higher than the upper bound. This implies that the range of admissible city sizes declines with an increase of φ. That is, if knowledge is completely general, the steady state city size distribution moves down to the degenerate case of a single optimal city size.

## 2.2 City size distribution, types of cities and localization externalities

The above model is incomplete insofar as it gives no indication of how the initial city size distribution which is constant under certain conditions of inter-urban knowledge spillovers has come about in the first place. An answer to this question gives a theoretical approach pioneered by Henderson (1987, 1988) which considers the city size distribution as the result of a continuous entrance of new cities growing to an optimal city size. An end to the entry of new cities would imply a convergence of all cities of an identical type to a unique city size. Different sizes of mature specialized cities are interpreted as reflecting simply the different industrial bases of these cities. The different sizes of cities which are dominated by different kinds of industries are explained by the nature and the extent of urban scale economies. Growth of the city system is argued for by reference to knowledge spillovers in an endogenous growth context, as suggested in Lucas (1988). Human capital accumulation then fuels urban growth, bearing directly on the question of which city size distribution is conducive to growth.

The fact that different technologies are used in the industrial sectors of different types of cities also implies that private returns to human capital investment and the magnitude of knowledge spillovers will vary naturally across city types. As a consequence, the equilibrium levels of per person human capital and contemporaneous real incomes will vary across different types of cities.

Evolving regional imbalance is then a consequence of growth: Even an integrated economy grows at least with measured income and educational inequality across different types of cities for otherwise identical people. Peculiar assumptions are needed to model migration in this context which ensure that per capita consumption is equal across cities.

*The spatial organization of production and population*

To confine the idea of the correspondence of the city size and sectoral specialization to the simplest case it is assumed that there are only two types of cities, performing different functions and having different equilibrium sizes, per worker human capital levels and incomes. Type 1 cities in the economy produce the numéraire good, an intermediate input, which is purchased by firms in type 2 cities. Firms in type 2 cities

are specialized in producing a consumption good, with price P in units of the intermediate good. There are no costs of inter-urban transactions. Finally there is a city formation process in national land markets involving either land developers or local governments.

*Family migration and human capital investment decisions*

Workers are members of dynastic families. For dynasties, each family starts with the same per person human capital and each family's size grows at the same rate g. Each family discounts the future at a rate ρ. The discount rate is assumed to be larger than the population growth rate to ensure well-behaved solutions. At each instant, dynasties choose how much total family income should be allocated to per member consumption c and how much to increasing the family's human capital stock. Families must allocate also their members across city types and decide on the per person human capital investments for members by the type of city in which they live. For existing family members, current own human capital endowments are nontransferable, except to newborns. These family decisions govern human capital accumulation as there are no formal markets for human capital.

Using the usual form of the utility function of per person consumption, the optimization problem for any dynasty can be formulated as

$$\max_{c,h_1,h_2,x} \int_0^\infty \left( \frac{c^{1-\sigma}-1}{1-\sigma} \right) e^{-(\rho-g)t} dt, \sigma > 0, \rho > g \text{, such that} \tag{16}$$

$$P\dot{H} = ze^{gt}I_1 + (1-z)e^{gt}I_2 - Pce^{gt} \tag{a}$$

$$H = ze^{gt}h_1 + (1-z)e^{gt}h_2 \tag{b}$$

$$\dot{H} \geq 0; \dot{h}_1/h_1 + g \geq 0; \dot{h}_2/h_2 + g \geq 0 \tag{c}$$

In the initial situation the family size is normalized to 1. H is the family's human capital stock, z the proportion of family members assigned to type 1 cities and (1-z) the proportion of those assigned to type 2 cities. $I_1$ and $I_2$ represent net per worker incomes earned by workers living in type 1 and type 2 cities respectively, with $h_1$ and $h_2$ representing human capital levels. Constraint (a) states that the value of family human

capital growth is the difference between total family income and the value of consumption. Constraint (b) states that total family human capital is the sum of individual human capitals ($h_1$ and $h_2$) of members in type 1 and type 2 cities. The constraints in (c) tell us that families can neither borrow nor consume their human capital. Conversion of the consumption good to human capital is irreversible. The second constraint in (c) implies that human capital is only transferable as an endowment to newborns in the same city type. The maximal percentage drop in per member human capital in a city type is the growth rate of its offspring. Neither of these constraints is binding in equilibrium.

Given that family members will generally earn different incomes $I_1$ and $I_2$ by city type there must generally be a intra-family redistribution across cities to maintain equality of per capita consumption.

### *2.2.1 The structure of cities*

It is assumed that production in a city occurs under "localization" economies of scale - own industry local external economies of scale. Instantaneous efficiency of each firm is enhanced by having more firms in the same industry in a city, with whom the firm communicates about what inputs to buy from whom what product lines to emphasize etc. Over time firm efficiency and the benefits of larger cities will be enhanced by local human capital accumulation. In achieving contemporaneous equilibrium city size, scale benefits are traded off against the higher per person internal commuting and potentially congestion and pollution costs of supporting larger cities.

*Type 1 cities*

Each firm in the city is composed of one worker. In each period that worker decides how much to produce and how much to invest in private human capital accumulation. Having single worker firms is a convenience, so that human capital spillovers only exist across firms, not within firms. Output of firms of the numéraire intermediate input $X_1$ is given by

$$X_{1i} = D_1 \left[ n_1^{\delta_1} h_1^{\psi_1} \right] h_{1i}^{\theta_1} \tag{17}$$

The output accrues entirely as income of the workers. $n_1$ is the employment in industry 1 in this type of city, $h_1$ is the average level of human capital of workers there and $h_{1i}$ is

the human capital of the worker in firm i. $\delta_1$ represents scale economies arising from the total volume of local communications which are proportional to $n_1$. $\psi_1$ is the elasticity of firm i's output with respect to the average level of human capital in the city which represents the spillover benefits of local levels of human capital or knowledge accumulation.

Given that all workers are assumed to be identical the city output is

$$X_1 = D_1 n_1^{1+\delta_1} h_1^{\theta_1+\psi_1} \tag{18}$$

This specification of the production technology assumes that human capital spillovers and scale externalities are purely localized. The presence of another industry at the locality would be of no importance. As agglomerating people into cities is costly on the commuting side, land developers are assumed to form specialized cities.

Using a simple standard version of the internal spatial structure of cities equilibrium in the land market is characterized by a rent gradient, declining linearly from the center of the city, the Central Business District to the city edge where rents are down to zero relative to alternative, non-urban uses. The total land rents are then

$$\frac{1}{2} b n_1^{3/2}, \text{ with } b \equiv \frac{2}{3} \pi^{-1/2} * \max \text{ radius}.$$

Gross rents are income of the city developer.

Type 1 cities are formed by land developers in the competitive context of a large economy with many type 1 cities on the national land market. Each city is run by a developer who collects urban land rents and offers inducements to firms to locate in the city, to achieve the rent maximizing city size. The developer's profits are residential land rents minus transfer payments $T_1$, to each worker/firm. The developer faces a free migration constraint that each worker's net income (gross income minus commuting costs and rents) equals the prevailing net income available in national labor markets.

The decision problem of the developer is then

$$\max_{n_1, T_1} \Pi_1 = \frac{1}{2} b n_1^{3/2} - T_1 n_1, \text{ such that} \tag{19}$$

$$W_1 + T_1 - \frac{3}{2} b n_1^{1/2} = I_1 .$$

Forming the maximum function of (19) and using the zero profit condition for land developers we obtain the "Henry George result" that total transfers to firms ($T_1 n_1$) are equal to total land rents. The per worker transfer closes the gap between private marginal product and social marginal product due to enhanced scale benefits when a worker/firm enters a city.

The city size rises with the scale elasticity $\delta_1$ up to the level of the commuting cost elasticity. Beyond that value, all production of the intermediate good will occur in one city.

*Type 2 cities*

Type 2 cities specialize in production of the economy's consumption good, sold in competitive national markets at a price P. A single worker firm's output is

$$X_{2j} = D_2 \left[ n_2^{\delta_2} h_2^{\psi_2} \right] h_{2j}^{\theta_2} x_{1j}^{1-\alpha} \tag{20}$$

Notation is completely analogous to the production of type 1. The profits for a firm of type 2 are $PX_{2j}$ - $x_{1j}$, where $x_{1j}$ are the intermediate inputs used in the production of the consumption good. From maximizing with respect to the input of the intermediate good and substituting into the profit function we obtain the residual return to the worker

$$W_2 = \alpha (1-\alpha)^{\frac{1-\alpha}{\alpha}} D_2^{1/\alpha} P^{1/\alpha} \left( n_2^{\delta_2} h_2^{\psi_2} \right)^{1/\alpha} h_{2j}^{\theta_2/\alpha} \tag{21}$$

On land rent collection and the grant of transfers analogous conditions of the individual optimum of the land developer in city 2 type cities are obtained.

*Investment and Migration decisions*

Forming a Hamiltonian on the basis of (16) and inserting the wage and transfer expressions families allocate human capital across city types to equalize private returns on investment. The solution of the dynamic optimization program shows that the individual human capital levels in both types of cities will differ and also the personal incomes. The differences are time invariant.

*National market equilibrium*

The equilibrium in national output markets requires a balance of trade among cities. That is, the national market equilibrium involves not only the determination of the relative price of the consumption good P but also the number of cities of type 1, $m_1$, and the number of type 2 cities $m_2$. The trade balance requires that

$$m_1 X_1 = m_2 n_2 x_1 + m_1 \left(b n_1^{3/2}\right) + m_2 \left(b n_2^{3/2}\right).$$

By imposing symmetry across dynasties nationally, each dynasty will send the same proportion of workers to each type of city. From the symmetry we can determine the fraction z of people going to type 1 cities in equilibrium. Due to the optimization over an infinite time horizon and the perfect foresight of households the workers once assigned to a city type never need to change that type. Migration only involves assignment of young workers starting their careers, especially to new cities. From the determination of z follows the allocation of the families' total stock of human capital to the production of intermediate goods $h_1$ and the production of the final good $h_2$.

If there is human capital growth, a condition that is generally satisfied in equilibrium, human capital grows in parallel at the same rate in the two types of cities. The only capital transfers required from workers of each type are to their own children. That is, in equilibrium human capital can be nontransferable across people and specific to the technology of producing intermediate goods or consumption goods.

### *2.2.2 Urban Growth*

Differentiating the expressions for optimal city sizes and using the equilibrium price value of the consumption good we have

$$\frac{\dot{n}_2}{n_2} = \frac{\dot{n}_1}{n_1} = 2\varepsilon_1 \frac{\dot{h}_1}{h_1} \tag{22}$$

$\varepsilon_1$ is the elasticity of income with respect to human capital levels in cities of type 1. In steady state this value is close to one. Therefore the number of firms grows at double the rate of the human capital.

From the equations defining the number of cities follows the equation indicating the growth of the cities of different types

$$\frac{\dot{m}_1}{m_1} = \frac{\dot{m}_2}{m_2} = g - 2\varepsilon_1 \frac{\dot{h}_1}{h_1}$$

That is, urban growth across city types is parallel, maintaining a constant relative size distribution of cities. Relative sizes and numbers of the two types of cities are time invariant, with all cities growing in size at the same rate.

*Aggregate Economic Growth*

Under the conditions of endogenous growth ($\varepsilon = \varepsilon_1 - (\varepsilon_1 - \varepsilon_2)(\alpha - 2\delta_2) = 1$) the economy achieves a steady state growth where consumption and human capital grow at a common rate and city sizes grow at $2\varepsilon_1$ of that rate. If $\varepsilon$ is smaller than one, the economy converges to (constant) steady state levels of consumption and human capital and cities achieve a stationary size. In any case, the steady state growth path or the steady state levels are too small compared to the social optimum due to the lack of internalization of the spillovers of human capital investment.

## 2.3 Discussion

The theoretical approaches give clearly distinguishable insights into the theoretical forces which might determine the relationship between the city size distribution and aggregate growth. The complexity of the issues involved seems to require that they are incomplete:

- The Eaton/Eckstein model does not offer a theory of the genesis of the city size distribution. It rather shows what might be responsible for the persistence of the city size distribution of mature, industrialized countries like France and Japan. Being a one sector model it does not allow to distinguish different sectoral contributions. Knowledge must be completely general as it is usable in any city and any economic activity. In this sense the alleged driving force of urban growth is dynamic urbanization externalities.
- The Black/Henderson model accounts for the emergence of new cities and can therefore be used to study the genesis of the city size distribution. It emphasizes the own industry effects, the own industry localization economies. Cities are considered to be completely specialised. Including a model of a local government sector, it offers

a direct access to study the impact of local economic policies. The major limitations result from the absence of a theory of migration, reducing the assignment of people to cities to a restrictive set of assumptions about dynastic decisions on the residential location of newborns. Against the backdrop of the empirical work of Henderson et al. it appears as a severe problem that there is no possibility that cities change their type. In the empirical work, intra - urban sectoral changes in the sense of a "product cycle of cities" have been identified to be important.

For both approaches the microeconomic characteristics of the informational spillovers must be important. So far there is no theoretical foundation for the assumed mechanisms: Why should competing agents share information that bears upon the microeconomic productivity and therefore on the competitive position of the agent? If there are reasons that there are trades of information which have an economic value why are these trades localised, given the dramatic decreases in communication costs?

As the above theories on the relationship between the city size distribution and growth are clearly distinguishable with respect to the nature of the dynamic economies the question arises whether the empirical literature on the dynamic agglomeration economies gives any orientation as to which of the approaches seems to be more promising.

## 3 Empirics of dynamic agglomeration economies

Maybe the best known contribution to the empirical importance of dynamic urbanization economies versus localization economies is the article of Glaeser et al. (1992). The results of the empirical investigations are:

- Measured in terms of employment, those industries grew slower which were located in cities where they were more heavily overrepresented.
- City - industries grew faster when the rest of the city was less specialized.
- Industries grew faster in cities where the average size of firms of that industry is smaller than the overall average.

That is, their findings clearly seem to support the view that urban diversification promotes urban growth. The intersectoral spillovers of technical and organizational

information are more important than the intrasectoral spillovers. In contrast to some of the literature on localization economies the intensity of competition seems to be positively related to urban growth. Consequently, the pressure to adopt new technologies and to imitate are held to be more important than inventive activities.

These results are not confirmed by the empirical study of Henderson et al. (1995). Distinguishing between "young" and "old" industries, they find out that for mature industries only localization economies are relevant. For very young industries urbanization economies are found to be important, with own industry spillovers becoming more and more important as the industries grow older. They conclude that to study the role of the city size distribution for growth, it is important to look at city specialization and diversification. What is more, for each industry the whole spectrum of city sizes is relevant. It is these empirical findings they base the claim on that an urban product cycle is of major importance for the relationship between urbanization, city size distribution and growth. As we have seen this view is not consistient with the Black/Henderson model.

There is some support for the Eaton/Eckstein model from the empirical findings of Glaeser and Maré on "learning in cities". They find out that there is a wage premium of 34 per cent for metropolitan areas surrounding cities of more than one million inhabitants, compared to rural areas, and still a premium of 10 per cent for metropolitan areas not surrounding large cities. That learning is the main function of the cities was inferred from the following results:

- Rural workers who move to cities don't have an immediate wage gain.
- Workers who leave cities do not experience a wage decline.
- The relative advantage in wage payments of workers in metropolitan areas surrounding big cities increases over time.

All of these findings are compatible with the Eaton/Eckstein model of migration and discredit the Black/Henderson model of dynastic decision making.

Recent empirical work of Henderson (1999) questions the relevance of Glaeser et al. (1992) and Henderson et al. (1995) for assessing the existence of dynamic externalities. These studies assert that, if the level of employment in an industry today is correlated with local own industry employment 15 or 30 years ago, that is evidence of dynamic

externalities. The first problem with this interpretation is conceptual: The typical estimating equation contains two key measures - base period own industry employment to control for "mean reversion", induced by allocative shocks, and a base period own industry concentration measure to represent localization externalities. The mean reversion control and the concentration measure are so closely related, that it is hard to distinguish effects. The mean reversion process and how externalities inhibit mean reversion or dampen allocative shocks has never been explicitly modeled. Therefore, the interpretation of the employment growth equations can only be tentative.

The second problem concerns the question whether a partial correlation between present employment levels and past concentration really implies externalities. Rather the correlation can arise from a "fixed effect" in estimation, representing unmeasured time invariant locational attributes such as immobile resource endowments, local culture affecting the legal, tax and institutional environment, and access to national and international markets. In short, the current industrial location pattern may be related to historical ones, not because of dynamic externalities, but because of persistent local comparative advantage.

To avoid the mean reversion and fixed effects problems Henderson has directly examined the effects of changes in past local industrial environments on changes in plant productivity. He found that scale externalities derive from own industry externalities, and were very local. They derived from the numbers of own industry plants in the home county, as opposed to a measure of the industry's total employment, or from activities in surrounding counties in the same Metropolitan Statistical Area. The fact that single plant firms seemed to benefit more from these externalities supports the view that the externalities were dynamic in nature. Externalities also derived from past births of plants, as opposed to numbers of pre-existing plants, suggesting that they were dynamic in nature.

## 4 Conclusions

The clarification of some of the methodological problems of the early empirical studies suggests that major progress in identifying dynamic agglomeration economies calls for the extension of the above models by an analysis of what role innovative activities play for the emergence of new cities. For these edge cities cooperation amoung innovating

firms seems to be the most important determinant of the clusters. With increasing demand, and the maturing of the city, the sectoral diversification might result from "love of variety" type of preferences, with the resulting allocation effects which are emphasized by the New Economic Geography. A *growth* effect might derive from an increase in the division of labor in producing the intermediate inputs, given a constant elasticity of substitution technology with respect to intermediate inputs and decreasing average costs of producing the intermediate goods.

The literature on the link between the urban structure and growth also suggests that a theory of migration which can do without restrictive assumptions on family decision making, must distinguish between human capital which is acquired in schooling and general training on the one hand and technical and organizational knowledge which is completely sector-specific. Past investment decisions with regard to the proportions of general and specific human capital investments will determine moving costs and therefore inter-urban migration flows.

## References


*Aghion, P., E. Caroli, and C. Garcia-Penalosa* (1999)
Inequality and economic growth. Working Paper 9908. CEPREMAP, Paris.

*Arnott, R.* (1979)
Optimum city size in a spatial economy. Journal of Urban Economics 6: 65-89.

*Benhabib, J., and M. M. Spiegel* (1997)
Growth and investment across countries: Are primitives all that matter? Working Paper . C. V. Starr Center for Applied Economics at New York University, New York.

*Bils, M., and P. Klenow* (1998)
Does schooling cause growth or the other way around? Working Paper 6393. NBER, Cambridge, Mass.

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

*Blomstrom, M., R. E. Lipsey, and M. Zejan* (1996)
Is fixed investment the key to economic growth? Quarterly Journal of Economics 111: 269-276.

*Chenery, H. B., and M. Squyrin* (1975)
Patterns of Development *1950 -1970*. London.

*Diamond, P., D. McFadden, and M. Rodriguez* (1978)
Measurement of the elasticity of factor substitution and bias of technical change. In M. Fuss, and D. McFadden (ed.), Production Economics: A Dual Approach to Theory and Applications., Vol. II. Amsterdam.

*Eaton, J., and Z. Eckstein* (1997)
Cities and growth: Theory and evidence from France and Japan. Regional Science and Urban Economics 27: S. 443-474.

*Eaton, J., and S. Kortum* (1997)
Engines of growth: Domestic and foreign sources of innovation. Japan and the World Economy, 9: S. 235-259.

*Glaeser, E., and D. Maré* (1994)
Cities and skills. Working Paper E-94-11. Hoover Institution .

*Glaeser, E. L.* (1997)
Learning in cities. Working Paper 6271. NBER, Cambridge, Mass.

*Henderson, J. V.* (1987)
General equilibrium modeling of systems of cities. In E. Mills (ed.), Handbook of Regional and Urban Economics, Vol. II. Amsterdam.

1988)
Urban Development: Theory, Fact, and Illusion. New York.

*Jones, C. I.* (1997)
Comment (on Klenow and Rodriguez-Clare). In B. Bernanke, and Rothemberg J. (ed.), NBER Macroeconomics Annual *1997*. Cambridge, Mass.

*Klenow, P. J., and A. Rodriguez-Clare* (1997)
The neoclassical revival in growth economics: Has it gone too far? In B. S. Bernanke, and J. J. Rothemberg (ed.), NBER Macroeconomics Annual . Cambridge, Mass.

*Krugman, P. R.* (1994)
The myth of Asia's miracle. Foreign Affairs 73: 62-78.

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

*Mankiw, N. G., D. Romer, and D. N. Weil* (1992)
A contribution to the empirics of economic growth. Quarterly Journal of Economics, 107 (2): 407-437.

*Prescott, E. C.* (1997)
Needed: A Theory of Total Factor Productivity. Research Department Staff Report. Federal Reserve Bank of Minneapolis, Minneapolis.

*Pritchett, L.* (n.d.)
Where has all the education gone? mimeo. World Bank, Washington, D. C.

*Rauch, J.* (1993)
Productivity gains from geographic concentration of human capital: evidence from cities. Journal of Urban Economics 34: 3-33.

*Rodrik, D.* (1996)
Comments (on Collins and Bosworth). In W. C. Brainard, and G. Perry (ed.), Brookings Papers on Economic Activity, Vol. II.

*Young, A.* (1994)
Lessons from the East Asian NICs: A contrarian view. European Economic Review 38: 964-973.

(1995)
The tyranny of numbers. Confronting the statistical realities of the East Asian growth experience. Quarterly Journal of Economics 100: 641-680.