Pollution and the efficiency of urban growth

Martin F. Quaas[a,*] and Sjak Smulders[b]

[a] Department of Economics, University of Kiel, Germany

[b] Department of Economics, University of Calgary, Canada and Tilburg University, The Netherlands

June 2008

Abstract. We analyze the efficiency of urbanization patterns in a dynamic model of endogenous urban growth with three sectors of production. Urban production exhibits increasing returns to scale on aggregate. Urban environmental pollution, as a force that discourages agglomeration, is caused by domestic production. We show that cities are too large and too few in number in equilibrium if economic growth implies increasing aggregate emissions. If, however, production becomes cleaner over time ('quality growth') the urbanization path approximates the efficient urbanization path after finite time without need of coordination.

JEL-Classification: Q56, R12, O18

Keywords: cities, urbanisation, pollution, growth, migration, sustainable development

[*]Mail address: Olshausenstr. 40, 24118 Kiel, Germany. Email: quaas@economics.uni-kiel.de. Martin Quaas gratefully acknowledges support from the German Academic Exchange Service (DAAD).

1 Introduction

There is a widely held concern in the urban economics literature that cities tend to be too large in equilibrium. In particular environmental pollution is assumed to be exceedingly high in large cities, expecially in developing countries (Henderson 1977; Henderson 2002a; Henderson 2002b; Trolley 1979; Shah and Nagpal 1997; UNFPA 2001; UN-Habitat 2003). This becomes even more of an issue in a dynamic context where economic activity and urban population grow over time. Yet, unlike other forces limiting city sizes, urban environmental pollution may either increase or decrease over time, even in a growing economy. As the extensive literature on the so-called Environmental Kuznets Curve indicates, the environmental load with some pollutants – like e.g. SO_2 – increases with output at early stages of economic development, but may ultimately decrease again (e.g., Andreoni and Levinson 2001, Egli and Steger 2007, Grossman and Krueger 1995, Lieb 2002; 2003, Plassmann and Khanna 2006, Stern 2004). In light of these observations the crucial question is, if cities are currently too large compared to their optimal size, will this problem become worse or better in the course economic development?

Seminal contributions by Henderson (1974,1988) provide a theoretical foundation of the concern that cities are too large in equilibrium. He showed that in an uncoordinated migration equilibrium, cities are too few in number and too large each, because in the absence of a coordinating mechanism people and firms are reluctant to found new cities and rather stay in inefficiently large existing cities. As a solution to this coordination failure he proposed that powerful land developers should be given the right to found new cities and receive all land rents people and firms pay in the new city. The competition of such land developers for tenants would then lead to efficient city sizes. Recent dynamic theories of urban growth assume that such powerful land developers exist and ensure efficient city sizes (Black and Henderson 1999, Rossi-Hansberg and Wright 2007). In contrast to this view, Helsley and Strange (1997) argue that in practice developers only have limited power and, thus, may fail to implement efficient city sizes. But also the very diagnosis has been challenged that without coordination equilibrium city sizes are inefficient. In a model of two regions with an exogenously growing total population, Anas and Xiong (2005) show that an efficient population distribution over two cities may emerge without developers if there are positive externalities between cities.

In this paper, we re-examine the question of whether and under which conditions equilibrium city sizes are inefficient in a dynamic context. For this sake, we develop a simple and analytically solvable model of endogenous urban growth that is related to the recent theories of urban growth (e.g. Black and Henderson 1999, Rossi-Hansberg and Wright 2007). Endogenous growth is driven by human capital accumulation, and an endogenous number of cities forms as the result of (aggregate) increasing returns to scale in production. In contrast to current theories, we include environmental pollution into the model, as pollution is a problem of major concern in many cities.

Unlike current theories our model can explain a *decreasing* number of cities even in a growing economy – this is the case when environmental pollution decreases, such that equilibrium city sizes increase. We also show that an endogenous number of cities is crucial: considering just two cities, as in much of the literature (e.g., Anas and Xiong 2005 or many models of the so-called New Economic Geography, e.g. Krugman 1991 or Puga 1999) may generate misleading results. In particular, the number and patterns of equilibrium outcomes may be substantially different when there are more cities.

Our main result is that cities may be of efficient sizes or inefficiently large in an equilibrium development path, depending on the nature of economic growth. If economic growth is accompanied by increasing environmental pollution, cities are inefficiently large in the uncoordinated equilibrium. Under conditions of 'quality growth', with decreasing environmental pollution, city sizes are close to optimal, i.e. the uncoordinated market equilibrium becomes efficient. This, however, does not happen immediately, but there is a hysteresis effect. It takes a sufficiently large reduction of environmental pollution, and, correspondingly, a sufficiently long interval in time over which pollution is reduced, until the efficient urbanization pattern is reached in equilibrium.

The next section develops the model and derives the factor prices and outputs for a given distribution of population across regions and sectors, i.e. the short-run equilibrium allocation. Section 3 contains the dynamic analysis of the model, and derives the equilibrium and Pareto-optimal paths of urban growth and urbanization patterns. In the final Section 5 we summarize and discuss our results.

2 The model and first results

We consider a small open economy that trades a primary resource and a final consumption good on world markets at given prices. The price of the resource is normalized to unity, the price of the consumption good is P. At these prices, the primary resource and the consumption good are also traded between cities within the economy. By contrast, intermediate goods that are used to produce the final consumption good are non-tradable and can be used only within the city where they are produced. One may imagine the intermediate goods as specific sub-contracted production services that are provided locally.

The economy is divided into a large number of regions i where a city could possibly exist. A city, in our model, is a region that is actually inhabited by a positive number of (urban) residents. While the number of regions is exogenous, the number of cities is one of the most important endogenous variables of the model. All regions are identical in the first place, and we we will omit the index i when setting up the model in the following, as long as no confusion may arise. In a city, two sectors of production exist: an industrial sector which produces the final consumption good and a small and medium-sized enterprises (SME) sector with a large variety of firms. Each firm in the SME-sector produces a particular variety of the intermediate good, using the resource and sector-specific human capital.

To operating a business in the SME-sector one unit of specific human capital K has to be employed at a rental rate of r. In addition, for each unit of output, $(\epsilon - 1)/\epsilon > 0$ units of the resource are consumed. Hence, a firm's total costs of producing x units of output are:

$$C(x) = r + \frac{\epsilon - 1}{\epsilon} x \tag{1}$$

The industrial sector uses a composite X of intermediate goods and specific human capital H to produce a quantity M of the final consumption good with a technology described by the production function

$$M = H^{\mu} X^{1-\mu}, \tag{2}$$

where H is the firm's employment of human capital and

$$X = \left[\int_0^I x(i)^{\frac{\epsilon-1}{\epsilon}} \, di \right]^{\frac{\epsilon}{\epsilon-1}} \tag{3}$$

is the aggregate of a mass I of different intermediate inputs of quantity x each. The different varieties are assumed to be (imperfect) substitutes in production, i.e. the elasticity of substitution exceeds one, $\epsilon > 1$.

Let w denote the rate of return to human capital specific to the industrial sector, p the domestic price of a variety of the intermediate good and G the price index of the intermediate composite. Profit maximization of competitive firms in the industrial sector yields (see appendix A.1)

$$\begin{aligned} w\,H &= \mu\,P\,H^{\mu} X^{1-\mu} && (4) \\ p\,x &= (1-\mu)\,P\,H^{\mu} X^{1-\mu}\,G^{\epsilon-1}\,p^{1-\epsilon} && (5) \\ G\,X &= (1-\mu)\,P\,H^{\mu} X^{1-\mu}, && (6) \end{aligned}$$

where the price index for the intermediate good is

$$G = \left[\int_0^I p(i)^{1-\epsilon} \, di \right]^{\frac{1}{1-\epsilon}}. \tag{7}$$

Monopolistically competitive profit maximization of firms in the intermediate goods-sector yields $p = 1$ (see appendix A.1). Given free entry of firms, profits are zero. From this condition, we derive the output per firm $x = \epsilon\, r$.

Given a region's endowment of human capital K specific to the SME-sector, the number of businesses in an urban region is $I = K$, as operating a firm in the intermediate sector requires one unit of specific capital. Hence, the price index for intermediates is

$$G = K^{\frac{1}{1-\epsilon}}, \tag{8}$$

and the aggregate input of intermediates in the industrial sector is

$$X = \epsilon\, r\, K^{\frac{\epsilon}{\epsilon-1}}, \tag{9}$$

i.e., $G\,X = \epsilon\, r\, K$. With this, we derive from equations (4) and (6) the rates of return to both types of human capital,

$$r = \frac{1}{\epsilon}\,((1-\mu)\,P)^{\frac{1}{\mu}}\;H\,K^{\psi-1} \tag{10}$$

$$w = \frac{\mu}{1-\mu}\,((1-\mu)\,P)^{\frac{1}{\mu}}\;K^{\psi}, \tag{11}$$

where we used the abbreviation $\psi = \frac{1-\mu}{\mu}\,\frac{1}{\epsilon-1} > 0$. Total output is, using (9), $x = \epsilon\, r$ and (10)

$$M = H^{\mu}\left(\epsilon\, r\, K^{\frac{\epsilon}{\epsilon-1}}\right)^{1-\mu} = ((1-\mu)\,P)^{\frac{1-\mu}{\mu}}\;H\,K^{\psi}. \tag{12}$$

This equation shows that on aggregate the urban production technology exhibits increasing returns to scale. In particular, the parameter cluster ψ may be interpreted as the degree of increasing returns in city production.

Environmental pollution occurs locally as an unwanted by-product of the production of intermediate goods. Pollution load is assumed to be proportional to the aggregate input of the primary resource, $I\,\frac{\epsilon-1}{\epsilon}\,x$ (cf. equation 1), with a proportionality factor $e(t)$,

$$E = e(t)\,I\,\frac{\epsilon-1}{\epsilon}\,x. \tag{13}$$

Given the technology of production, in particular (2), it is possible to abate emissions by substituting human capital for intermediate goods. It is however not possible to 'export' pollution, as intermediate goods cannot be traded between cities. We assume an exogenous 'pollution-saving' technical progress at a rate $\eta > 0$, i.e. the amount of pollution per unit of resource input decreases over time, such that $e(t) = e_0\,\exp(-\eta\, t)$. Thus, even with increasing use of resources, local pollution does not necessarily increase at the same rate. Actually, it may even decrease if η is large enough, i.e. larger than the growth rate of resource use. We call such a situation 'quality growth' or 'smart growth'.

Using the above results $I = K$ and $x = \epsilon\, r$ in equation (13), we obtain aggregate local environmental pollution

$$E = e(t)\, K\, \frac{\epsilon - 1}{\epsilon}\, \epsilon\, r \overset{(10)}{=} e(t)\, \frac{\epsilon - 1}{\epsilon}\, ((1-\mu)\, P)^{\frac{1}{\mu}}\; H\, K^{\psi}. \tag{14}$$

This equation shows that environmental pollution increases with the city's endowments of human capital of both types.

Total urban population $N(t)$ is assumed to be growing (or declining) at a constant rate γ, capturing rural-urban migration as well as natural growth (or decline) of urban population, such that $N(t) = N_0 \exp(\gamma\, t)$. Each individual decides in which city she lives and in which sector she is employed, i.e. there is inter-regional as well as inter-sectoral migration. Each individual can only live at one city and work only in one sector at a given moment in time. We do not consider any migration costs. Concerning inter-regional migration, we denote the population share of region i with $n_i \in [0, 1]$, i.e. $N\, n_i$ people live in region i. The number of cities is endogenous and given by the number actually inhabited regions, i.e. those regions i with a non-zero population, $n_i > 0$. Of course, the number of cities and the population of each city may change over time. Concerning inter-sectoral migration, we denote the share of workers employed in the SME-sector of city i by l_i. Assuming that all individuals are employed, the share of workers in the industrial sector of city i is $1 - l_i$. Hence, total labor force employed in the SME-sector in city i is $N\, n_i\, l_i$, and the total labor force employed in the industrial sector of the same city is $N\, n_i\, (1 - l_i)$. The shares of workers employed in either of the sectors is endogenous, too, and may change over time.

Each individual is initially endowed with k_0 units of human capital specific to the SME-sector and h_0 units of human capital specific to the industrial sector. Both types of human capital can be transferred between cities, and newly born individuals (or rural-urban migrants) are immediately endowed with the average level of human capital. Further human capital can only be accumulated if the individual under consideration actually works in the respective sector, i.e. no 'learning' is possible without 'doing'. Per capita human capital $h(t)$ specific to the industrial sector is accumulated with a linear technology according to the equation

$$\dot{h} = \theta_h\, (1 - u_h)\, h, \tag{15}$$

where the dot superscript denotes a derivative with respect to time, u_h is the share of human capital used in production and $\theta_h > 0$ would be the growth rate of human capital if all of it could be used for accumulation rather than for producing intermediate goods. Per capita human capital k specific to the SME-sector is accumulated with a similar technology according to

$$\dot{k} = \theta_k \,(1 - u_k)\, k, \tag{16}$$

where u_k and θ_k are defined analogously to u_h and θ^h.

Intertemporal utility function of each individual is

$$\int\limits_0^\infty \left(\ln\, c + \phi\, \ln\, \left(\bar{E} - E\right)\right)\, \exp(-\rho\, t)\, dt, \tag{17}$$

where c is the amount of goods consumption, E is environmental pollution in the individual's city of residence. Beyond a maximum tolerable level of pollution $\bar{E}$ life at the place of residence is impossible. The parameter $\phi > 0$ is the weighting factor of environmental pollution in utility; $\rho > 0$ is the rate of time preference.

Income of an individual working in the industrial sector of city i is $y_{h,i} = u_h\, h\, w_i$, i.e. the industrial wage rate in city i times the amount of human capital used in production. Similarly, the income of an individual employed in the SME-sector is $y_{k,i} = u_k\, k\, r_i$. All current income is used for the consumption of goods. Hence, demand for the consumption good is $c_{h,i} = y_{h,i}/P$ for an individual who works in the industrial sector and $c_{k,i} = y_{k,i}/P$ for an individual who is employed in the SME-sector of city i.

In order to derive the time paths of human capital endowments, we determine the optimal decisions concerning the u_h, h, u_k, and k. Suppose the individual under consideration joins the industrial sector of the urban region i at time $\underline{t}$ and leaves it at time $\bar{t}$. Using $c_{h,i} = y_{h,i}/P = u_h\, h\, w_i/P$ in (17), her optimization problem reads

$$\max_{u_h} \int\limits_{\underline{t}}^{\bar{t}} \left[\ln\,(u_h\, h\, w_i/P) + \phi\, \ln\, \left(\bar{E} - E_i\right)\right]\, \exp(-\rho\, t)\, dt, \tag{18}$$

subject to (15) and given $h(\underline{t})$. Since the individual under consideration takes w_i, P, and E_i as given, this is problem is equivalent to $\max\limits_{u_h} \int\limits_0^\infty \ln\,(u_h\, h)\, \exp(-\rho\, t)\, dt$ subject to (15) and given $h(\underline{t})$. Hence, the accumulation of human capital does not depend on the city of residence. This means, the optimization problem (18) is independent of when the individual under consideration leaves or enters the industrial sector of a specific city, i.e. it is independent of $\underline{t}$ and $\bar{t}$. Hence, the optimal accumulation of human capital is determined by the solution of

$$\max_{u_h} \int\limits_0^\infty \ln\,(u_h\, h)\, \exp(-\rho\, t)\, dt \quad \text{s.t. (15)}, \tag{19}$$

given $h(0) = h_0$. Similarly, the corresponding optimization problem for an individual working in the SME-sector only depends on the stock of sector-specific human capital and the time spent in production. The solution of these optimization problems is given by the following lemma.

Lemma 1. *The share of human capital specific to the industrial sector used in production is $u_h = \rho/\theta_h$ =const. Human capital is accumulated according to*

$$h(t) = h_0 \exp((\theta_h - \rho)\, t). \tag{20}$$

The share of human capital specific to the SME-sector used in production is $u_k = \rho/\theta_k$ =const. Human capital is accumulated according to

$$k(t) = k_0 \exp((\theta_k - \rho)\, t). \tag{21}$$

Proof. see appendix A.2. □

Using these results, we can also determine the sectoral distribution of labor. In order to derive the population shares engaged in either of the sectors, we consider the decision of which sector to join at time $t = 0$. By assumption, all individuals at this time are equally endowed with human capital specific to both sectors. It turns out that the share of workers employed in either sector is constant over time and independent of the city of residence.

Lemma 2. *In each city i the share l_i of workers employed in the SME-sector is constant over time and given by*

$$l_i = \frac{\epsilon\, \mu}{1 - \mu + \epsilon\, \mu}. \tag{22}$$

Proof. see appendix A.3. □

Given the population share n_i of city i, lemma 1 with lemma 2 determines the total human capital stocks of both types, i.e. $K_i = N\, n_i\, l_i\, u_k\, k$ and $H_i = N\, n_i\, (1 - l_i)\, u_h\, h$. Moreover, incomes are the same in both sectors, i.e. $y_i = y_{h,i} = y_{k,i}$ with (see appendix A.3)

$$y_i = \frac{\mu}{1 - \mu}\, ((1 - \mu)\, P)^{\frac{1}{\mu}}\, \frac{\rho}{\theta_h} \left(l_i\, \frac{\rho}{\theta_k}\right)^{\psi} N^{\psi}\, k^{\psi}\, h\, n_i^{\psi}. \tag{23}$$

Given human capital stocks and the population share of city i, environmental pollution in that city is

$$E_i = \frac{\epsilon - 1}{\epsilon}\, ((1 - \mu)\, P)^{\frac{1}{\mu}}\, (1 - l_i)\, \frac{\rho}{\theta_h} \left(l_i\, \frac{\rho}{\theta_k}\right)^{\psi} e\, N^{1+\psi}\, k^{\psi}\, h\, n_i^{1+\psi}. \tag{24}$$

Hence, both income and environmental pollution crucially depend on the city's population, in particular in the city's share n_i of total urban population.

Based on the results derived in this section the analysis in the following section will focus on how the population shares n_i of the different regions, i.e. the urbanization patterns, evolve over time. We shall derive both the equilibrium outcomes and the Pareto-optimum, given the exogenous development of pollution-saving technical progress $e(t)$, growth (or decline) of urban population $N(t)$ and the dynamics of human capital accumulation, equations (20) and (21).

3 Urbanization dynamics

As individuals can freely choose their city of residence, they will move to the region with the highest utility level. Considering the level of utility at region i as a function of income y_i and environmental pollution E_i, the utility level in city i is $\ln\left(\frac{y_i}{P}\right) + \phi\,\ln(\bar{E} - E_i)$. Using equations (23) and (24), we can express utility as a function of time and city i's population share n_i.

$$\ln\left(\frac{y_i}{P}\right) + \phi\,\ln(\bar{E} - E_i) = \ln\left(\frac{\mu}{1-\mu}\,((1-\mu)\,P)^{\frac{1}{\mu}}\,\frac{\rho}{\theta_h}\left(l_i\,\frac{\rho}{\theta_k}\right)^{\psi} N^{\psi}\,k^{\psi}\,h\,n_i^{\psi}\right)$$
$$+\,\phi\,\ln\left(\bar{E} - \frac{\epsilon-1}{\epsilon}\,((1-\mu)\,P)^{\frac{1}{\mu}}\,(1-l_i)\,\frac{\rho}{\theta_h}\left(l_i\,\frac{\rho}{\theta_k}\right)^{\psi} e\,N^{1+\psi}\,k^{\psi}\,h\,n_i^{1+\psi}\right). \quad (25)$$

This expression looks quite complicated. But indeed, it may be simplified to a great extent. Transforming (25) in an appropriate way, we obtain the following result

Lemma 3. *An individual will choose to live in the city such as to maximize the transformed utility function*

$$v(n_i) = n_i^{\psi}\left(1 - z(t)\,n_i^{1+\psi}\right)^{\phi} \quad (26)$$

with

$$z(t) = \frac{1}{\bar{E}}\,\frac{\epsilon-1}{\epsilon}\,((1-\mu)\,P)^{\frac{1}{\mu}}\,(1-l_i)\,\frac{\rho}{\theta_h}\left(l_i\,\frac{\rho}{\theta_k}\right)^{\psi} e(t)\;N(t)^{1+\psi}\,k(t)^{\psi}\,h(t). \quad (27)$$

Utility $v(n_i)$ is a hump-shaped function of n_i that is zero at $n_i = 0$ and $\bar{n} = z(t)^{-\frac{1}{1+\psi}}$. For all $z(t) > \frac{\psi}{\psi+\phi\,(1+\psi)}$, $v(n_i)$ has a unique maximum at $n^{max} = \left[\frac{\phi}{\psi} + 1 + \phi\right]^{-\frac{1}{1+\psi}} z(t)^{-\frac{1}{1+\psi}}$.

Proof. see appendix A.4 □

The quantity $z(t)$ defined in equation (27) may be interpreted as a measure of aggregate environmental pollution. It is equal to total emissions – measured relative to the maximum possible amount of pollution $\bar{E}$ – that would prevail if the whole economy's production was concentrated in a single city. Of course, $z(t)$ can exceed one, i.e. aggregate pollution may be larger than the maximum tolerable pollution in any single city.

By definition $z(t)$ does not depend on n_i, but urbanization dynamics are driven by the dynamics of $z(t)$, as $v(n_i)$ depends on both city i' population share n_i and aggregate pollution $z(t)$.

Lemma 3 states that utility in city i is a hump-shaped function of the city's population share n_i. The reason is that utility is increasing with the population share for a small population, because environmental pollution is still small and a larger population generates a higher income due to increasing returns to scale in production. For a large population, the increasing environmental damage outweighs the benefit from increasing returns to scale such that utility ultimately decreases with the city's population. Utility assumes a maximum at an intermediate level of population at which increasing returns to scale in production and environmental pollution are in an appropriate balance.

To derive the equilibrium and efficient urbanization patterns at a given moment in time we will consider $z(t)$ as given. After that, in order to derive the urbanization dynamics, we will consider how the dynamics of $z(t)$ determine the dynamic evolution of urbanization patterns. For in any migration equilibrium the utility level must be the same in all cities. Otherwise, an incentive to migrate would still exist. Hence, a migration equilibrium requires

$$v(n_i) = v(n_{i'}) \tag{28}$$

for all cities i and i' with $n_i > 0$ and $n_{i'} > 0$. In general, no closed form solution exists for this equation. Many general results can however be derived without such a solution.

Figure 1 illustrates the properties of the utility function $v(n_i)$ that determine the possible equilibrium and efficient urbanization patterns. Of course, utility in any pair of two cities is the same if both cities have the same population share, i.e. $n_i = n_{i'} \Rightarrow v(n_i) = v(n_{i'})$. Second, given the hump-shaped utility function $v(n_i)$ (lemma 3), for any given n_i equation (28) is solved by exactly one $n_{i'} \neq n_i$, as illustrated in figure 1. This result is formally proven in the following lemma.

Lemma 4. *There are at most two different equilibrium sizes of cities.*

Proof. see appendix A.5. □

This means that if in equilibrium there are three (or more) cities, at least two of them are of equal size. As a consequence, either all cities are inhabited by an equal share of population, or there are two types of cities, one with a larger share of population and the other one with a smaller share of population. This result holds for any migration equilibrium, whether it is stable or not.

Of particular interest, however, are the stable equilibria, i.e. those migration equilibria to which the economy returns after a small perturbation of the population distribution. Focusing on stable equilibria rules out the majority of asymmetric migration

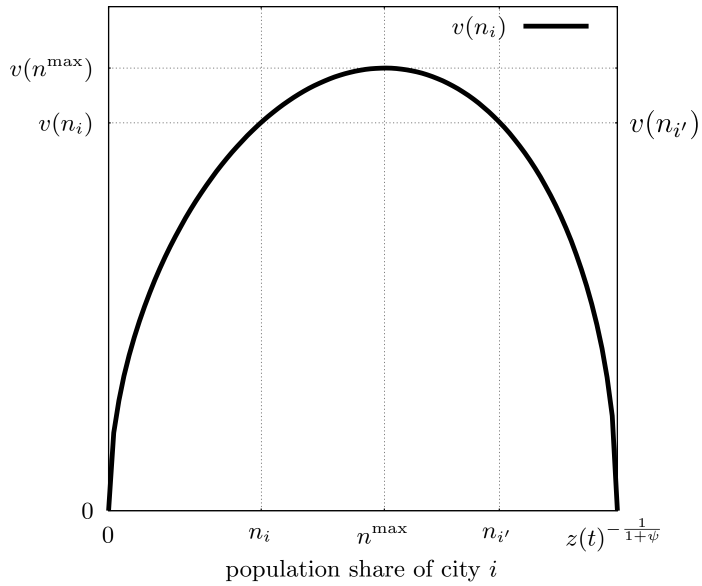

Figure 1: The (transformed) utility $v(n_i)$ of an individual living in city i as a function of this city's population share n_i.

equilibria. In particular, only migration equilibria can be stable where all or all but one cities are 'large', i.e. have a share of population $n_i > n^{\max}$, as shown in the following lemma.

Lemma 5. *Migration equilibria in which more than one city has a population share smaller than n^{max} (i.e. the population share at which utility is maximal) are unstable.*

Proof. By lemma 4 the two (or more) small cities must be of equal size in equilibrium. A small deviation from this equilibrium would lead to a population differential between the two (or more) small cities. This would create an incentive for further migration from the smaller city to the city that has become a little bigger, as utility increases with population size for $n_i < n^{\max}$. Hence, such an equilibrium is unstable. □

Taken together, lemma 4 and lemma 5 imply that only two types of stable urbanization patterns exist in the model: either all cities are of equal size or there is just one small city in addition to one or more large cities. If all cities are of equal size, each city has the same share of population, i.e. $n_i = 1/m$ where m is the number of cities, i.e. regions that are actually inhabited. For such an equal distribution of population across m cities it is easy to determine whether the corresponding urbanization pattern is stable or not. If each city's population share $n_i = 1/m$ is smaller than $n^{\max}$ the pattern is unstable. A small deviation from the equilibrium would imply that there is an incentive for further migration to the city that has become a little larger than the others. For, as illustrated in figure 1, utility is increasing in n_i for all $n_i < n^{\max}$. By contrast, if each city's population share $n_i = 1/m$ is larger than $n^{\max}$ the pattern is stable. Any deviation from the equilibrium would create an incentive to move back to the smaller city, because utility is decreasing in n_i for all $n_i > n^{\max}$ due to high environmental damages. In this sense, environmental pollution is a negative feedback mechanism that prevents unlimited growth and thereby stabilizes city sizes.

For very low aggregate environmental pollution $z(t) \leq 1$, complete agglomeration of economic activity in one region is the only stable equilibrium. It is easily verified that for an equal distribution of population over two cities, each city would have a population share of less than $n^{\max}$ if $z(t) \leq 1$. Also a pattern with one large city and one small city would not be stable against a deviation from this equilibrium such that the smaller city would grow and the larger city would shrink a little bit.

For larger aggregate environmental pollution, $z(t) > 1$, complete agglomeration is not an equilibrium anymore. Rather all equal distributions of population across cities are stable for which each city has a population share greater than $n^{\max}$. In addition, a pattern with one small and several large cities may be stable. The stable urbanization patterns are derived in the following proposition.

Proposition 1 (stable urbanization patterns)
1. For $z(t) > 1$, stable urbanization patterns consist of $m^\star$ cities with a population share $1/m^\star$ each for all

$$m^\star \in \left\{ m \in I\!N \;\middle|\; z(t)^{\frac{1}{1+\psi}} \leq m \leq \left(\frac{\phi}{\psi} + 1 + \phi\right)^{\frac{1}{1+\psi}} z(t)^{\frac{1}{1+\psi}} \right\}. \tag{29}$$

2. For all

$$m^{\star\star} \in \left\{ m \in I\!N \;\middle|\; \left(\frac{\phi}{\psi} + 1 + \phi\right)^{\frac{1}{1+\psi}} z(t)^{\frac{1}{1+\psi}} - 1 \leq m \leq z(t)^{\frac{1}{1+\psi}} \right\}, \tag{30}$$

an asymmetric distribution of population over $m^{\star\star} - 1$ cities with a population share $n^+ > n^{max}$ and one city with a population share $n^- < n^{max}$ is the only stable equilibrium.

Proof. Part 1. If the population share $1/m$ of each of the m cities exceeds the value $z(t)^{-\frac{1}{1+\psi}}$, transformed utility would be negative (lemma 3). Hence, there would be an incentive to move to any empty region where transformed utility is zero, i.e. an equal distribution of population across m regions with $m^{1+\psi} \geq z(t)$ is unstable. This is the lower inequality in condition (29). Provided this condition holds, an equal distribution over m cities is a stable equilibrium if utility decreases with each city's population share, i.e. if $v'(1/m) < 0$. Using (26), this condition is

$$v'(1/m) = m^{-\psi}\left(1 - z(t)\, m^{-(1+\psi)}\right)^{\phi}\left[\psi\, m - \frac{\phi\,(1+\psi)\, z(t)\, m^{-\psi}}{1 - z(t)\, m^{-(1+\psi)}}\right] < 0,$$

which is the case, if and only if $\psi\, m^{1+\psi} - \psi\, z(t) < \phi\,(1+\psi)\, z(t)$, i.e. if $m^{1+\psi} < (\phi\,(1+\psi)+\psi)/\psi\, z(t)$. Rearranging leads to the upper inequality in condition (29).
Part 2. There is no integer fulfilling condition (29), if

$$m < z(t)^{\frac{1}{1+\psi}} \quad \text{and} \quad m+1 > \left(\frac{\phi}{\psi} + 1 + \phi\right)^{\frac{1}{1+\psi}} z(t)^{\frac{1}{1+\psi}}$$

Putting both inequalities together and rearranging yields condition (30). If this condition is fulfilled, no stable symmetric equilibrium exists. Hence, an asymmetric equilibrium must be stable. Since the only stable asymmetric equilibrium is one with only one 'small' city (lemma 5), this is the only stable equilibrium. □

For any given $z(t)$, condition (29) determines a set of integers. The larger $z(t)$, the larger is the number of subsequent integers that are contained in this set. Hence, proposition 1 shows that the higher $z(t)$ the more stable symmetric equilibria exist at the same time. The economic intuition for this result is as follows. The larger aggregate environmental pollution is, the smaller are the city sizes at which pollution limits city sizes. If there are many cities with small population each, however, environmental pollution in each city would increase only by a small amount if one city's population would distribute over all remaining cities. Such an urbanization pattern with one city less would still be stable. Only if many cities would vanish in this manner each city's population and environmental pollution would ultimately become so large that maximum tolerable level of pollution is exceeded.

As stated in the second part of the proposition, it may also be the case that no stable symmetric equilibrium exists. In that case, a pattern with one small city and several large cities is the only stable migration equilibrium. However, this may only happen for comparatively small values of aggregate pollution: the larger $z(t)$, the smaller is the number of integers that fulfill condition (30). Eventually the set determined by this condition will be empty. In particular, this is the case for all levels of aggregate pollution for which the left hand side of the inequality contained in (30) is greater than the right hand side, i.e. for all levels of aggregate pollution for which

$$z(t) > \left(\left(\frac{\phi}{\psi} + 1 + \phi\right)^{\frac{1}{1+\psi}} - 1\right)^{-(1+\psi)} \equiv \bar{z}. \tag{31}$$

Asymmetric equilibria may be stable for higher values of $z(t)$. In that case there also exist stable symmetric equilibria at the same level $z(t)$ of aggregate pollution. The population shares of the cities in stable equilibria for different $z(t)$ are depicted in figure 2 (for the parameter specification $\psi = 1$ and $\phi = 1$). A horizontal solid line in this figure,

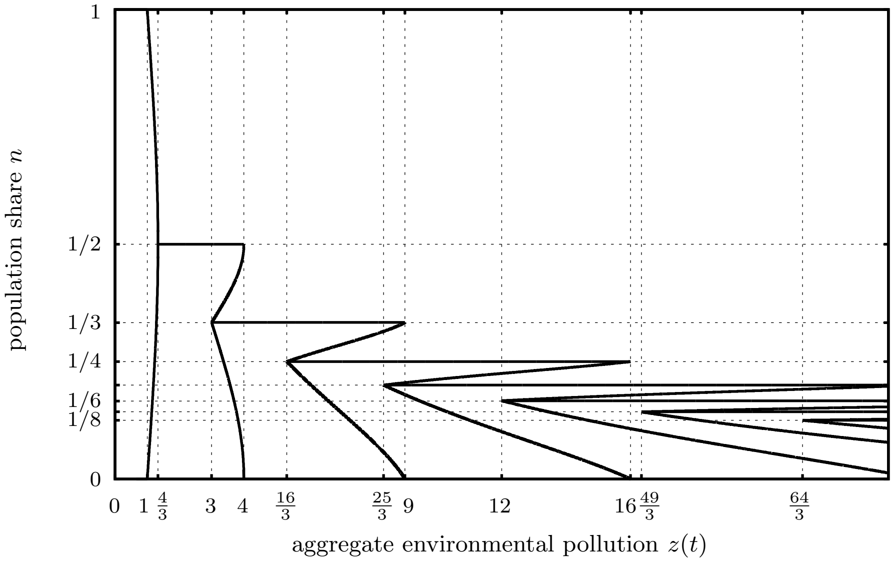

Figure 2: The population shares n of one of the cities in stable equilibria for different $z(t)$, for the specification $\psi = 1$ and $\phi = 1$.

e.g. at $n = 1/4$, depicts a stable urbanization pattern with 4 cities having a population share of 1/4 each. For the specification of parameters used in figure 2, an asymmetric distribution of population over one small and one large city is the only stable equilibrium for $z(t) \in (1, 4/3)$. For larger values of $z(t)$, stable symmetric equilibria exist. For large levels of aggregate pollution, $z(t) > 25/3$, several symmetric equilibria may be stable at the same time. In the example shown in figure 2 asymmetric equilibria are stable also for values of $z(t)$ which are larger than $\bar{z}$. This is the case, e.g. for $z(t) \in (3, 4)$. For these levels of aggregate pollution a stable urbanization pattern exists with one small city, with a population share between zero and 1/3, and two large cities, with a population share between 1/3 and 1/2.

As many stable equilibria may exist at the same time, the questions arise which of these equilibria will actually occur when $z(t)$ is evolving over time and whether or not these equilibria are Pareto efficient. The rest of the analysis is concerned with these questions. Concerning the latter question, it is obvious that the stable symmetric equilibria differ with regard to the level of utility. In particular, among all stable symmetric equilibria those with a larger number of cities, i.e. smaller population shares each, involve a higher level of utility. This is because in a stable symmetric equilibrium utility by definition is decreasing with the share of population living in each city. Hence if there are (at least) two stable symmetric equilibria, the pattern with more cities of smaller size is Pareto-superior to the other one. The following proposition identifies the Pareto-efficient equilibrium, which is unique, given a value of aggregate pollution $z(t)$.

Proposition 2 (Pareto-optimal urbanization patterns)
1. An equal population distribution over $\hat{m}$ cities, where

$$\hat{m} = \min \left\{ m \in I\!N \left| \frac{m^{-\frac{\psi}{\phi}} - (m+1)^{-\frac{\psi}{\phi}}}{m^{-\left(\frac{\psi}{\phi}+\psi+1\right)} - (m+1)^{-\left(\frac{\psi}{\phi}+\psi+1\right)}} \geq z(t) \right. \right\} \tag{32}$$

is a Pareto optimum.
2. Other equilibria are not Pareto-optimal. In particular,

a) equal population distribution over $m \neq \hat{m}$ cities is not Pareto-optimal and

b) equilibria with cities of different size are not Pareto-optimal.

Proof. We proceed by proving the following two steps: (i) given a number m of cities, no Pareto improvement is possible by changing relative city sizes, compared to the equal distribution of population. (ii) when $m = \hat{m}$ as defined in the proposition, $v(1/(\hat{m}-1)) < v(1/\hat{m})$ and also $v(1/(\hat{m}+1)) < v(1/\hat{m})$, i.e., the equal distribution of population over $\hat{m}$ cities is an optimum. This will also prove part 2.a of the proposition.

Ad (i). Consider $m \geq 1/n^{\max}$ ($m < 1/n^{\max}$). For any deviation from $n_i = 1/m$, there must be some city with a population $n_i < 1/m$ ($n_i > 1/m$). In that city, $v(n_i) < v(1/m)$, as is easily confirmed. Hence, a deviation from the symmetric distribution $n_i = 1/m$ for all m is not a Pareto improvement. Ad (ii). An equal distribution over m cities is superior to an equal distribution over $m+1$ cities, if $v(1/m) \geq v(1/(m+1))$, i.e. if

$$m^{-\psi} \left(1 - z(t)\, m^{-(1+\psi)}\right)^{\phi} \geq (m+1)^{-\psi} \left(1 - z(t)\,(m+1)^{-(1+\psi)}\right)^{\phi}. \tag{33}$$

Clearly, as long as $1/m < n^{\max}$, i.e., if $m > 1/n^{\max}$, this inequality will be always fulfilled. If $m+1 < 1/n^{\max}$, on the other hand, this equation is not fulfilled. Thus, the smallest integer that fulfills inequality (33) is the closest integer to $1/n^{\max}$, i.e., the optimal number of cities. With little rearrangement, we arrive at condition (32).

Finally, to prove part 2.b of the proposition, consider an equilibrium with (at least) one city with a population share $n_1 < n^{\max}$ and (at least) one city with a population share $n_2 > n^{\max}$. If one individual would move from city 2 to city 1, utility would increase in both cities, i.e. a Pareto-improvement is obtained. □

Thus, a Pareto-optimal equilibrium exists, but just one of the different equilibria is Pareto-optimal and moreover the optimum is not necessarily a stable equilibrium.

In order to analyze under which conditions the equilibrium outcome is optimal, we further investigate how the economy dynamically evolves over time. Since both the population distribution in the different equilibria and in the Pareto optimum depend on aggregate emissions $z(t)$, we shall concentrate on how $z(t)$ evolves over time. According to equation (27), the growth rate of $z(t)$ is

$$\frac{\dot{z}(t)}{z(t)} = \frac{\dot{e}(t)}{e(t)} + (1+\psi)\,\frac{\dot{N}(t)}{N(t)} + \psi\,\frac{\dot{k}(t)}{k(t)} + \frac{\dot{h}(t)}{h(t)} = -\eta + (1+\psi)\,\gamma + \psi\,(\theta_k - \rho) + \theta_h - \rho \quad (34)$$

Whether $z(t)$ increases or decreases over time depends on the relative sizes of the rate η of pollution-saving technical progress, the growth rate γ of population, and on the growth rates of both types of human capital. In particular if there is no pollution-saving technical progress, $z(t)$ will increase over time unless population is declining at an extreme pace. Only if the rate of pollution-saving technical progress η is high enough to outweigh the increase in total production output (given by the other growth rates), $z(t)$ will decrease over time. We discuss the two possibilities of growing or declining aggregate pollution in turn.

In order to understand how the equilibrium urbanization pattern changes over time when aggregate pollution increases, consider condition (29) in proposition 1. If $z(t)$ increases, the upper bound on the equilibrium number of cities becomes less restrictive while the lower bound becomes more restrictive. Hence, for a given stable symmetric equilibrium with m_+ cities at some moment in time the lower bound becomes binding. If aggregate pollution increases further, the pattern with m_+ cities is not an equilibrium anymore. Rather, a first individual will move to a formerly empty region and set up a business. A second individual will move to the same region rather than to another one, because due to increasing returns to scale the already inhabited region offers a higher level of utility. Hence, just one new city will form and the resulting new equilibrium will consist of $m_+ + 1$ cities.

So, if people move one by one (Anas and Xiong 2005 call this "laissez-faire"), with growing aggregate pollution the equilibrium number of cities will be as small as possible. The lack of coordination in the laissez-faire outcome involves too few cities which are too large each, compared to the Pareto-optimum, as stated in the following proposition.

Proposition 3 (inefficiency of urbanization with increasing aggregate emissions)
When aggregate pollution $z(t)$ increases from a value $z(0) = \bar{z}$ (eq. 31), a $z^\star$ exists such that for all $z(t) \geq z^\star$, the equilibrium outcome is not Pareto optimal. The difference between the optimal and equilibrium number of cities is increasing over time and unbounded from above.

Proof. The larger $z(t)$ the larger the number of stable symmetric equilibria (cf. proposition 1). We choose $z^\star$ such that there are at least two stable symmetric equilibria. Since the number of cities in with growing $z(t)$ is the minimum number of cities among all stable symmetric equilibria (part 1 of the proposition), another stable symmetric equilibrium exists with $m^\star_+ + 1$ cities. Because for stable symmetric equilibria utility is decreasing in each city's population share, all individuals enjoy a higher level of utility in the equilibrium with $m^\star_+ + 1$ cities compared to the equilibrium with $m^\star_+$ cities. Hence the latter is not Pareto optimal. Since the number of stable symmetric equilibria is unbounded from above, also the difference between the equilibrium number of cities and the optimal number (as given by proposition 2) is unbounded. □

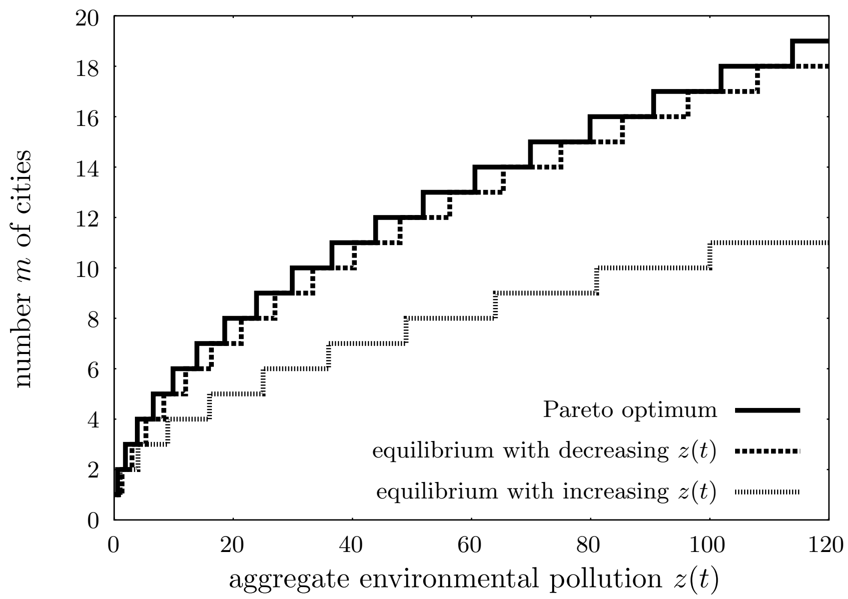

Figure 3: The number of cities for varying $z(t)$ in the Pareto optimum (solid line), in the equilibrium with decreasing aggregate pollution $z(t)$ (broken line) and in the equilibrium with increasing aggregate pollution $z(t)$ (dotted line). The parameters are $\psi = 1$ and $\phi = 1$.

The optimal number of cities and the equilibrium number of cities are plotted as functions of $z(t)$ in figure 3. In this figure the solid line depicts the Pareto-optimal number of cities and the dotted line depicts the equilibrium number of cities when aggregate pollution is increasing. For $z(t)$ sufficiently large (i.e. $z(t) > 4$) these two numbers clearly deviate from each other. The difference between the optimal and equilibrium number of cities is increasing with aggregate pollution $z(t)$, because in the optimum the number of cities grows much faster with $z(t)$ than the equilibrium number of cities with increasing aggregate pollution. Indeed, the difference in the number of cities grows without bound when aggregate environmental pollution continues to grow (proposition 3).

Since the utility level in each city depends on the population share of each city, it depends, in a symmetric equilibrium, on the number of cities. Utility is (considerably) higher in the Pareto-optimum than in the equilibrium with increasing pollution. The difference in the Pareto-optimal utility level and the equilibrium level of utility is increasing with aggregate pollution. That means, with increasing aggregate pollution $z(t)$, the equilibrium outcome becomes worse compared to the efficient urbanization pattern. This result is due to a coordination failure: if the inhabitants of the few large cities that exist in equilibrium could commit themselves to jointly move to empty places such that new cities will form a Pareto-improvement would be obtained. Henderson (1974; 1988) suggests that powerful land developers could assume a coordinating role in order to achieve such Pareto improvements. In the case of increasing aggregate emissions, also our model suggests that some sort of coordination mechanism, as e.g. endowing land developers with a sufficient amount of power, is needed.

However, an increase in aggregate emissions $z(t)$ is not the only possibility. If pollution-saving technical progress occurs at a sufficiently high rate (and/or urban population is shrinking), aggregate emissions may also decrease over time. This is the case, if the rate of pollution-saving technical progress η fulfills the following condition

$$\eta > (1-\psi)\,\gamma + \psi\,(\theta_k - \rho) + \theta_h - \rho. \tag{35}$$

We call such a development of the economy "smart growth" or "quality growth", because aggregate environmental quality increases in spite of an overall increasing economy. In the following we show that the coordination failure (almost) vanishes under conditions of smart growth.

In order to understand how the equilibrium urbanization pattern changes over time when aggregate pollution *decreases*, consider again condition (29) in proposition 1. If aggregate pollution $z(t)$ decreases from some large level, the lower constraint on the number of cities becomes less restrictive while the upper constraint becomes more restrictive. Any equilibrium number m_- of cities remains stable until the upper constraint is binding. If aggregate pollution declines further a symmetric distribution of population across m_- cities is not stable anymore: the stabilizing effect of environmental pollution has become weaker than the destabilizing effect of increasing returns to scale. Hence, one city will vanish, giving rise to a new stable equilibrium with $m_- - 1$ cities of equal size. This equilibrium will remain stable until the upper constraint becomes binding again. The upper constraint on the equilibrium number of cities in condition (29) is derived from the condition that for a larger number of cities a symmetric equilibrium would be unstable. That means, for a larger number of cities, utility would be increasing in the population share of each city (for an illustration see figure 1). Let $\overline{m}$ be the maximum number of cities such that for a given $z(t)$ the symmetric population distribution across $\overline{m}$ cities is stable, i.e. a symmetric equilibrium with $\overline{m}+1$ cities is unstable. Thus, we have $1/\overline{m} > n^{\max}$ and $1/(\overline{m}+1) < n^{\max}$. In words, in the symmetric equilibrium with $\overline{m}$ cities each cities population share is close to the population share that maximizes utility. This is proven formally in proposition 4.

Of course, the number of cities may not be optimal in the beginning. As $z(t)$ decreases, however, at some moment in time the upper constraint in condition (29) will become binding. From that moment onwards, the equilibrium allocation will be close to optimal. As shown in the following proposition, from any initially stable, but inefficient, urbanization pattern it will take finite time until a stable equilibrium close to the optimal one is reached.

Proposition 4 (efficiency of urbanization with decreasing aggregate emissions)
When aggregate pollution $z(t)$ decreases from a large level, i.e. condition (35) holds, after finite time the equilibrium outcome is almost Pareto optimal in the following sense: the difference between the optimal number $\hat{m}(t)$ and the equilibrium number $m_{-}^{\star}(t)$ of cities is either zero or one, i.e. $\hat{m}(t) - m_{-}^{\star}(t) \leq 1$.

Proof. Consider a stable population distribution in the beginning ($t = t_1$) that involves the worst possible utility level $v(n_m) = 0$. That means, the number m of cities in a symmetric equilibrium is such that the lower constraint on m in condition (29) is binding, i.e. $m = z(t_1)^{\frac{1}{1+\psi}}$. When $z(t)$ decreases, per capita utility increases, until the upper constraint in condition (29) becomes binding, i.e. until $m = \left(\frac{\phi}{\psi} + 1 + \phi\right)^{\frac{1}{1+\psi}} z(t_2)^{\frac{1}{1+\psi}}$. At this moment in time (or earlier), the population distribution is Pareto optimal. At the same time, it becomes unstable, and one city vanishes, as soon as $z(t)$ has become slightly smaller. A new symmetric equilibrium emerges with $m-1$ cities. At this moment in time, the equal distribution of population across m cities would be optimal, but is not stable. If $z(t)$ continues to fall, utility would decrease for an urbanization pattern with m cities and increases for the pattern with $m-1$ cities, until utility in the pattern with $m-1$ cities is larger than in the pattern with m cities, i.e. the former becomes the Pareto optimum. Hence, the difference between the equilibrium number $m_{-}(t)$ of cities and optimal number of cities is either one or zero, i.e. $\hat{m}(t) - m_{-}(t) < 1$.

The maximum time to reach the Pareto optimum is needed if the initial urbanization pattern is the worst possible. It is given by the time $t_2 - t_1$ until $z(t)$ decreases from

$z(t_1) = m^{1+\psi}$ to $z(t_2) = m^{1+\psi} / \left(\frac{\phi}{\psi} + 1 + \phi \right)$, i.e.

$$\frac{z(t_1)}{z(t_2)} = \frac{\phi}{\psi} + 1 + \phi \tag{36}$$

$$\exp\Big(-\left(-\eta + (1-\psi)\,\gamma + \psi\,(\theta_k - \rho) + \theta_h - \rho\right)\,(t_2 - t_1)\Big) = \frac{\phi}{\psi} + 1 + \phi. \tag{37}$$

By taking the logarithm on both sides leads the interval of time is calculated from which on the pattern of urbanization is almost efficient. □

While in the case of increasing aggregate pollution $z(t)$ the equilibrium number of of cities increases very slowly compared to the maximal possible increase in the number of cities, in the case of decreasing aggregate pollution $z(t)$, the equilibrium number of cities *decreases* very slowly compared to the maximal possible decrease. The difference between the two cases is that the equilibrium outcome is far from optimal in the case of increasing aggregate pollution, but close to optimal when $z(t)$ decreases over time. Indeed, with decreasing aggregate pollution the equilibrium outcome ultimately is a constrained Pareto optimum, if we consider only stable equilibria.

4 The model including the internal spatial structure of cities

So far, we have neglected the internal spatial structure of the cities. In this section, we will include this in form of a simple urban economics model.

Consider a linear city (a circular city would involve a little more notation, but no additional results) with a central industrial district (CID) in the centre, where all production takes place. Every individual consumes, in addition to the consumption of goods, one unit of living space, but production does not need any space. Workers must commute from their place of residence to the CID; one unit of distance from the CID requires ν units of commuting time.

An individual who lives at location s in city m and works in the industrial sector will gain a wage income of only $y = (1 - k\,s)\,u_h\,h\,w_m$, since some working time is lost for commuting. Similarly, if she works in the intermediate goods-sector, her wage in come will be $y = (1 - k\,s)\,u_k\,k\,r_m$. As in the basic model, intersectoral migration will continue, until incomes are the same in both sectors, i.e.

$$u_h\,h\,w_m = u_k\,k\,r_m. \tag{38}$$

The individual living furthest away from the CID lives at location

$$\bar{s} = N\,n_m\ . \tag{39}$$

Land rents outside the city are assumed to be zero, such that this individual has a wage income of $(1-\nu\,N\,n_m)\;u_h\,h\,w_m = (1-\nu\,N\,n_m)\;u_k\,k\,r_m$. A residential equilibrium requires that no individual has an incentive to choose another place of residence within the city. Rents for living space adjust accordingly, such that wage-income net of rents paid for living space is $(1-\nu\,N\,n_m)\;u_h\,h\,w_m = (1-\nu\,N\,n_m)\;u_k\,k\,r_m$ for every worker in the industrial as well as in the intermediate goods sector. Thus, the rent $g(s)$ for living space is

$$g(s) = u_h\,h\,w_m\,\nu\;(N\,n_m - s) = u_k\,k\,r_m\,\nu\;(N\,n_m - s)\;\;. \tag{40}$$

We assume that landlords do not have to work and commute to the CID, i.e. they live at the countryside, neither increasing pollution nor land rents in the cities. Thus, each worker's income spent for the consumption of goods is

$$y_m = (1-\nu\,N\,n_m)\;u_h\,h\,w_m = (1-\nu\,N\,n_m)\;u_k\,k\,r_m\;\;. \tag{41}$$

The maximum population share $\overline{n}$ of a single city is determined by the condition that the worker living at the border of the city just needs all of his time to commute, i.e. $1 = \nu\,N\,\overline{n}$, which corresponds to an income spent for consumption which is zero.

Aggregate human capital employed in the industrial sector is

$$H_m = N\,n_m\,L_m^h\;\left(1-\nu\,\frac{N\,n_m}{2}\right)\;u_h\,h\;\;, \tag{42}$$

which is the integral of the human capital of workers living at different places in the city times their working hours. (In equilibrium, the places of residence of workers in both sectors will be distributed equally in space.) Aggregate human capital employed in the intermediate goods sector is

$$K_m = N\,n_m\,L_m^k\;\left(1-\nu\,\frac{N\,n_m}{2}\right)\;u_k\,k\;\;. \tag{43}$$

From the condition that inter-sectoral distribution of labour will adjust until incomes are the same, we derive that the inter-sectoral distribution of population is still as given by lemma 2 (with a similar proof). Hence, income used for consuming goods is the same in both sectors,

$$y_m = \Gamma\,N^\psi\,k^\psi\,h\,n_m^\psi\;\left(1-\nu\,\frac{N\,n_m}{2}\right)^\psi\;(1-\nu\,N\,n_m)\;\;. \tag{44}$$

An individual will choose to live in the city m where

$$v(n_m) = n_m^\psi\;\left(1-\frac{n_m}{2\,\overline{n}}\right)^\psi\;\left(1-\frac{n_m}{\overline{n}}\right)\;\left(1-z\,n_m^{1+\psi}\;\left(1-\frac{n_m}{2\,\overline{n}}\right)^{1+\psi}\right)\;\;, \tag{45}$$

is maximal. Here, $\overline{n} = 1/(\nu\,N)$ (as discussed above) and z is aggregate environmental pollution as defined in equation (27). This expression looks more complicated than the expression we derived before for the simpler model (equation 25). It is, however, closely related.

Consider a situation, in which z is small (e.g., due to a low level of development, i.e., low values of h and k). Then, $v(n_m)$ reaches zero for $n_m = \bar{n}$, and it is hump-shaped in the domain $n_m \in [0, \bar{n}]$, while the last factor in (45) remains positive. In other words, the population share a city can host is limited by the maximum commuting distance which still would leave some working time for an individual living at the edge of the city. If total population $N(t)$ increases, this limiting factor becomes relatively more tight and inefficient city sizes would result, similar to the outcome described by proposition 3. One could, however, also think of an increasing $\bar{n}$, e.g., in the case of improvements in commuting infrastructure, such that ν decreases. In that case, maximum possible commuting distance would increase and the maximum possible, as well as the optimal, city sizes would increase. Ultimately, efficient city sizes would result, similar as described by proposition 4.

If $\bar{n}$ is large, and $z(t)$ is large, environmental pollution rather than commuting costs limits city sizes. If this is the case, qualitatively the same story obtains as in section 3: increasing aggregate pollution z leads to inefficient city sizes while decreasing ultimately leads to efficient city sizes. The quantiative results are modified though, as the shape of $v(n_m)$ as given by (45) also depends on $\bar{n}$.

5 Conclusion and discussion

In this paper we have studied the dynamic development of urbanization patterns in a model of endogenous urban growth where urban environmental pollution discourages the formation of very large cities. We considered two urban sectors of production: a small and medium-sized enterprises sector with a large endogenous variety of businesses producing intermediate goods from a primary resource, and an industrial sector, producing the final consumption good from the intermediates. The accumulation of sector-specific human capital, and an endogenous variety of intermediate goods lead to increasing returns to scale on aggregate. Equilibrium and optimal city sizes are determined by the balance between increasing returns to scale as a positive feedback mechanism that favors agglomeration and environmental pollution as a negative feedback mechanism that discourages agglomeration.

Our results also indicate that an endogenous number of cities is crucial. The set of potential urbanization outcomes is much richer than with a limited number of regions, especially when aggregate production output and pollution are large. Hence, considering just two regions, as in much of the literature may generate misleading results. In particular asymmetric equilibria are more likely for a small economy and when only a small number of cities exist. However, for the more realistic setting of a large economy associated also with a large level of aggregate pollution, the model predicts a large number of cities of equal and equal internal structure. This result is due to the assumption that all regions are identical ex ante. Outcomes with heterogenous city sizes could easily be described by our model if we impose differences in the absorption capacity of the urban environments or different factor productivities. Such differences could be, among other causes, due to differences in climate or physical geography. As a result, equilibrium city sizes would differ, allowing for a more realistic outcome. An empirical test of the theory developed in this paper should account for these exogenous differences between regions.

Our main result is that cities may be of efficient sizes or inefficiently large in an equilibrium development path, depending on the nature of economic growth. If economic growth is accompanied by increasing environmental pollution, cities are inefficiently large and too few in number in the uncoordinated equilibrium. Thus, the equilibrium growth path is not efficient. This result confirms the concern that urbanization patterns involve inefficiently large cities. This is different in the case of decreasing pollution, i.e. if there is 'quality growth' and production becomes cleaner over time. First of all, in contrast to current theories our model can explain a *decreasing* number of cities even in a growing economy when environmental pollution decreases. Then, the equilibrium city sizes increase and the economy's population will distribute over a smaller number of cities. In such a situation of quality growth the equilibrium city sizes and number are very close to the efficient urbanization pattern, i.e. the uncoordinated market equilibrium becomes efficient. However, this does not happen immediately, but it takes a sufficiently large reduction of environmental pollution, and, correspondingly, a sufficiently long interval in time over which pollution is reduced, until the efficient urbanization pattern is reached in equilibrium.

The policy conclusion is that a continuous reduction of aggregate emissions will ultimately lead to an efficient urbanization pattern. The point in time when the efficient urbanization pattern is reached can be observed, as it is the moment when the number of cities starts to decline. Between the moment when urban pollution starts to decline and the moment when the efficient urbanization pattern is reached there is a time lag, however, that can be long if the rate of pollution decrease and the degree of increasing returns are low and the preferences for environmental quality are strong. Accordingly it may require some degree of patience until under such an environmental policy also an efficient urbanization pattern is reached.

References

Anas, A. and K. Xiong (2005). The formation and growth of specialized cities: efficiency without developers or Malthusian traps. *Regional Science and Urban Economics*, **35**: 445–470.

Andreoni, J. and A. Levinson (2001). The simple analytics of the environmental Kuznets curve. *Journal of Public Economics*, **80**: 269–286.

Black, D. and J. V. Henderson (1999). A Theory of Urban Growth. *Journal of Political Economy*, **107**(2): 252–284.

Egli, H. and T. M. Steger (2007). A Dynamic Model of the Environmental Kuznets Curve: Turning Point and Public Policy. *Environmental and Resource Economics*, **36**(1): 15 – 34.

Grossman, G. and A. Krueger (1995). Economic growth and the environment. *Quarterly Journal of Economics*, **110**(2): 353–377.

Helsley, R. W. and W. C. Strange (1997). Limited developers. *Canadian Journal of Economics*, **30**(3): 329–348.

Henderson, J. V. (1974). The Sizes and Types of Cities. *American Economic Review*, **64**(4): 640–656.

Henderson, J. V. (1977). Externalities in a Spatial Context: The Case of Air Pollution. *Journal of Public Economics*, **7**: 89–110.

Henderson, J. V. (1988). *Urban Development: Theory, Fact, and Illusion*. Oxford University Press, Oxford.

Henderson, J. V. (2002a). Urban primacy, external costs, and quality of life. *Resource and Energy Economics*, **24**: 95–106.

Henderson, J. V. (2002b). Urbanization in Developing Countries. *The World Bank Research Observer*, **17**(1): 89–112.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, **99**: 483–499.

Lieb, C. M. (2002). The Environmental Kuznets Curve and Satiation: A Simple Static Model. *Environment and Development Economics*, **7**(3): 429–448.

Lieb, C. M. (2003). The Environmental Kuznets Curve: a survey of the empirical literature and of possible causes. University of Heidelberg, Department of Economics, discussion paper 391.

Plassmann, F. and N. Khanna (2006). A Note on 'The Simple Analytics of the Environmental Kuznets Curve'. *Environment and Development Economics*, **11**(6): p697 – 707.

Puga, D. (1999). The rise and fall of regional inequalities. *European Economic Review*, **43**: 303–334.

Rossi-Hansberg, E. A. and M. L. Wright (2007). Urban Structure and Growth. *Review of Economic Studies*, **74**: 597–624.

Shah, J. J. and T. Nagpal (1997). Urban Air Quality Management Strategy in Asia – Greater Mumbai Report. World Bank Technical Paper 381, World Bank, Washington.

Stern, D. (2004). The rise and fall of the Environmental Kuznets Curve. *World Development*, **32**: 1419–1439.

Trolley, G. (1979). *Urban Growth in a Market Economy*. Academic Press, New York.

UN-Habitat (2003). *The challenge of slums: Global Report on Human Settlements*. United Nations Human Settlements Programme, Nairobi.

UNFPA (2001). *Footprints and Milestones – Population and Environmental Change*. United Nations Population Fund.

A Appendix

A.1 The industrial firm's profit maximization

The problem is to maximize

$$P\,H^{\mu}\,X^{1-\mu} - w\,H - \int_0^I p\,x, \tag{A.1}$$

where X is given by (3), such that

$$\frac{d}{dx}X = X^{\frac{1}{\epsilon}}\,x^{-\frac{1}{\epsilon}}. \tag{A.2}$$

Hence, the first order condition with respect to x is

$$(1-\mu)\,P\,H^{\mu}\,X^{-\mu}\,X^{\frac{1}{\epsilon}}\,x^{-\frac{1}{\epsilon}} = p \tag{A.3}$$

$$((1-\mu)\,P\,M)^{\epsilon-1}\;p^{1-\epsilon} = x^{\frac{\epsilon-1}{\epsilon}}\;X^{\frac{(\epsilon-1)^2}{\epsilon}} \tag{A.4}$$

$$((1-\mu)\,P\,M)^{\epsilon-1}\;G^{1-\epsilon} = X^{\frac{\epsilon-1}{\epsilon}}\;X^{\frac{(\epsilon-1)^2}{\epsilon}} = X^{\epsilon-1}. \tag{A.5}$$

Rearranging leads to (6). From (A.3), we have

$$p\,x = ((1-\mu)\,P\,M)^{\epsilon}\;X^{1-\epsilon}\,p^{1-\epsilon} = (1-\mu)\,P\,M\;((1-\mu)\,P\,M)^{\epsilon-1}\;X^{1-\epsilon}\,p^{1-\epsilon}.$$

Using (A.5) and rearranging leads to (5).
Firms in the SME-sector choose p such as to maximize profits $p\,x - r - \frac{\epsilon-1}{\epsilon}\,x$ subject to (5), i.e. they solve

$$\max_p \left(p - \frac{\epsilon-1}{\epsilon}\right)\,(1-\mu)\,P\,H^{\mu}X^{1-\mu}\,G^{\epsilon-1}\,p^{-\epsilon} - r. \tag{A.6}$$

The first order condition for this problem is

$$(1-\epsilon)\,p^{-\epsilon} + (\epsilon-1)\,p^{-\epsilon-1} = 0. \tag{A.7}$$

Rearranging leads to $p = 1$.

A.2 Proof of lemma 1

The current-value Hamiltonian for the optimization problem (19) is (suppressing time arguments)

$$\mathcal{H} = \ln\left(u_h\,h\right) + \lambda_h\,\theta_h\,(1-u_h)\,h \tag{A.8}$$

The first order conditions are as follows

$$\frac{1}{u_h} = \lambda_h\,\theta_h\,h \tag{A.9}$$

$$\frac{1}{h} + \lambda_h\,\theta_h\,(1-u_h) = \rho\,\lambda_h - \dot{\lambda}_h \tag{A.10}$$

and the transversality condition requires

$$\lim_{t\to\infty} \lambda_h\,h = 0. \tag{A.11}$$

Using condition (A.9) in (A.10), we obtain

$$\theta_h\,u_h - \rho = \frac{\dot{u}_h}{u_h}. \tag{A.12}$$

This differential equation for u_h is solved by

$$u_h(t) = \frac{\rho}{\theta_h + C\,\exp(\rho\,t)}. \tag{A.13}$$

Using the transversality condition (A.11), rule out $C > 0$ (since then $\lim_{t\to\infty} \lambda_h\, h \to \infty$).[1] Thus, the solution to problem (19) is

$$u_h(t) = \frac{\rho}{\theta_h} \tag{A.14}$$

$$h(t) = h_0\, \exp((\theta_h - \rho)\, t). \tag{A.15}$$

Similarly, we derive for the human capital in the intermediate goods-sector

$$u_k(t) = \frac{\rho}{\theta_k} \tag{A.16}$$

$$k(t) = k_0\, \exp((\theta_k - \rho)\, t). \tag{A.17}$$

A.3 Proof of lemma 2

We start with the conjecture that an individual who once has chosen to join a particular sector will stay in that sector forever, and show that indeed there is no incentive to deviate from the decision once made. The human capital of individuals engaged in the intermediate goods and industrial sector is respectively given by equations (21) and (20). At time $t = 0$, an individual is indifferent between joining either of the sectors, if

$$\int_{t=0}^{\infty} \ln\left(\frac{r_i\, u_k\, k}{P}\right) \exp(-\rho\, t)\, dt = \int_{t=0}^{\infty} \ln\left(\frac{w_i\, u_h\, h}{P}\right) \exp(-\rho\, t)\, dt. \tag{A.18}$$

Since all individuals who are engaged in the same sector in the same city are identical, we get from equations (10) and (11)

$$r_i = \frac{1}{\epsilon}\, ((1-\mu)\, P)^{\frac{1}{\mu}}\; N\, n_i\, (1 - l_i)\, u_h\, h\; (N\, n_i\, l_i\, u_k\, k)^{\psi - 1} \tag{A.19}$$

$$w_i = \frac{\mu}{1-\mu}\, ((1-\mu)\, P)^{\frac{1}{\mu}}\; (N\, n_i\, l_i\, u_k\, k)^{\psi} \tag{A.20}$$

Plugging this into (A.18), and canceling common constants from both sides of the resulting equation, we derive the condition

$$\int_{t=0}^{\infty} \ln\left(\frac{1}{\epsilon}\, N\, n_i\, (1 - l_i)\, u_h\, h\; (N\, n_i\, l_i\, u_k\, k)^{\psi - 1}\; u_k\, k\right) \exp(-\rho\, t)\, dt$$

$$= \int_{t=0}^{\infty} \ln\left(\frac{\mu}{1-\mu}\, (N\, n_i\, l_i\, u_k\, k)^{\psi}\; u_h\, h\right) \exp(-\rho\, t)\, dt, \tag{A.21}$$

If condition (22) is fulfilled at each moment in time, both sides of this condition are equal. Thus, individuals have no incentive to move to a sector other than that in which they engage at $t = 0$.

[1] $C < 0$ is impossible, since then u_h eventually would become negative.

A.4 Proof of lemma 3

Re-arranging (25), we obtain

$$\ln\left(\frac{y_i}{P}\right) + \phi\,\ln(\bar{E} - E_i)$$
$$= \ln\left(\frac{\mu}{1-\mu}\,((1-\mu)\,P)^{\frac{1}{\mu}}\,\frac{\rho}{\theta_h}\,\left(l_i\,\frac{\rho}{\theta_k}\right)^{\psi}\,N^{\psi}\,k^{\psi}\,h\right) + \phi\,\ln\left(\bar{E}\right)$$
$$+ \ln\left(n_i^{\psi}\right) + \phi\,\ln\left(1 - \frac{\frac{\epsilon-1}{\epsilon}\,((1-\mu)\,P)^{\frac{1}{\mu}}\,(1-l_i)\,\frac{\rho}{\theta_h}\,\left(l_i\,\frac{\rho}{\theta_k}\right)^{\psi}}{\bar{E}}\,e\,N^{1+\psi}\,k^{\psi}\,h\,n_i^{1+\psi}\right). \tag{A.22}$$

The first two terms on the right hand side of this equation do not depend on n_i. Hence, they do not alter the migration decisions. Using the definition of $z(t)$ (equation 27), the terms of the utility function relevant for the migration decision are

$$\ln\left(n_i^{\psi}\right) + \phi\,\ln\left(1 - z(t)\,n_i^{1+\psi}\right). \tag{A.23}$$

A monotonic transformation (applying the exponential function) leads to the transformed utility function (26).
It is seen by inspection that the transformed utility function $v(n_i)$ is zero at $n_i = 0$ and $\bar{n} = z(t)^{-\frac{1}{1+\psi}}$. Differentiating $v(n)$ with respect to n yields

$$v'(n) = v(n)\,\left[\frac{\psi}{n} - \frac{z(t)\,(1+\psi)\,n^{\psi}\,\phi}{1 - z(t)\,n^{1+\psi}}\right] \tag{A.24}$$

$v'(n) \geq 0$, if and only if

$$\psi - z(t)\,n^{1+\psi}\,\left[\psi + \phi\,(1+\psi)\right] \geq 0. \tag{A.25}$$

Thus, $v(n)$ is hump-shaped with a unique maximum at $n^{\max} = \left[\frac{\psi}{z(t)\,[\psi+\phi\,(1+\psi)]}\right]^{\frac{1}{1+\psi}}$.

A.5 Proof of lemma 4

Assume the the equilibrium size of the first city is n_1. Then, the equilibrium sizes of all other cities $i \geq 2$ with $n_i > 0$ are determined by the condition

$$v(n_i) = v(n_1) \quad \text{or} \quad n_i = 0. \tag{A.26}$$

The equation $v(n_i) = v(n_1)$ has (at most) two positive solutions, since $v(n_i)$ is zero at $n_i = 0$ and $\bar{n} = z(t)^{-\frac{1}{1+\psi}}$ and has a unique interior extremum, which is a maximum, at $n_i = n^{\max}$.