Equilibrium and optimal urban systems with heterogeneous land*

– Working title (work in progress) –

Kristian Behrens† Frédéric Robert-Nicoud‡

Revised: September 11, 2014

Abstract

Not all land is equally attractive to develop cities. We modify the canonical urban model to allow for heterogeneous production amenities and a continuum of sites. We find, first, that there are decreasing returns to the system of cities and to individual cities at the optimal allocation. This differs markedly from the canonical model with homogeneous land where cities have the efficient scale and there are constant returns to the system of cities. Second, heterogeneity in developable land prevents land developers from achieving the first best allocation because land heterogeneity relaxes competition forces. Finally, a particularly harmful form of NIMBY-ism arises in the equilibrium with local governments.

Keywords: land developers; heterogeneous land; agglomeration; congestion

JEL Classification: R12; D31

*We are grateful to J. Vernon Henderson for extremely valuable comments and feedback. Behrens and Robert-Nicoud gratefully acknowledge financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada for the funding of the *Canada Research Chair in Regional Impacts of Globalization*.

†Department of Economics, Université du Québec à Montréal, Canada; National Research University – Higher School of Economics, Russia; CIRPÉE, Canada; and CEPR, UK. E-mail: behrens.kristian@uqam.ca

‡Geneva School of Economics and Management, Université de Genève, Switzerland; SERC, UK; and CEPR, UK. E-mail: frederic.robert-nicoud@unige.ch

1 Introduction

Not all land is equally attractive to develop cities and existing city locations are not random. We develop a simple framework to study the equilibrium and optimal allocations in urban systems with heterogeneous land. Following the pioneering work by Henderson (1974), we modify the canonical urban model to allow for heterogeneous sites, as proxied by heterogeneous production amenities. Within that modified framework we then study the decentralized equilibria as well as (locally and globally) optimal urban systems and allocations involving (possibly strategic) land developers.[1]

The canonical urban model analyses what Fujita and Thisse (2002) refer to as *the fundamental trade off in urban economics* between *agglomeration economies* – defined as external scale economies that operate at the city level – and *urban costs* such as congestion and expensive housing. When land is homogeneous and sites are replicable and there are no indivisibilities, the solution to this trade off only requires looking at individual cities in isolation, for one may replicate the (local) optimal solution and build identical cities elsewhere. Also, the location of cities does not matter and they can be treated as 'floating islands'. Yet, existing city locations are not random (Davis and Weinstein, 2002; Bleakley and Lin, 2012) and certainly not homogenous (Roback, 1982; Saiz, 2010; Albouy, 2012; Desmet and Rossi-Hansberg, 2013). In that case, the solution to the fundamental tradeoff is substantially altered: optimal individual city sizes and their location *depend on the distribution of production amenities and, therefore, on the whole urban system.*

As is well know, dealing with heterogeneous sites in a system of cities is not a trivial task since one has to focus on 'large' economies to circumvent the 'lumpiness problem' (i.e., the problem of dividing the population into a discrete number of cities of discrete sizes). As Henderson (1988, p.177) put it: *"The problem [with heterogeneous sites and a 'large' economy] is allowing for the existence of amenity bundles and an infinite number of cities. For example, if amenities are distributed uniformly on the line* $[0,1]$*, that allows for an infinite number of cities and population. However, it is then difficult to describe the discrete allocations to different sites, which involves balancing a limited population across a limited number of sites* [...]*. This remains an unsolved problem."*

We propose to solve the problem by taking a continuous approach.[2] Although the continuity assumption may not be to everyone's taste, we agree with Aumann (1964) when he wrote: *"The*

[1] Albeit important, one aspect of the problem that we do not consider is the timing of city formation in a growth context with heterogeneous sites. See, e.g., Henderson (1986, 1988), Anas (1992), Helsley and Strange (1994) and, more recently, Henderson and Venables (2009) for dynamic aspects of city formation and growth. Anas and Abdel-Rahman (2004) survey also a part of that literature.

[2] Many economic problems that are a priori hard to tackle can be made tractable using a continuous approach. For example, the Ricardian model – which is hard to handle with a large integer number of sectors – is quite tractable in its continuous version (Dornbusch et al., 1979; Eaton and Kortum, 2002). See also Davis and Dingle (2013) who embed what is essentially a continuous Ricardian model into an urban setting to make predictions on the specialization and skill composition of cities.

idea of a continuum of [cities] may seem outlandish to the reader. Actually, it is no stranger than a continuum of prices or of strategies or a continuum of "particles"in fluid mechanics. In all these cases, the continuum can be considered an approximation to the "true" situation in which there is a large but finite number of particles (or [cities] or strategies or possible prices). The purpose of adopting the continuous approximation is to make available the powerful and elegant methods of the branch of mathematics called "analysis", in a situation where treatment by finite methods would be much more difficult or even hopeless (think of trying to do fluid mechanics by solving n-body problems for large n)."

In our model, amenities a are continuously distributed on the line $[0, 1]$, with a continuous density $g(a)$ of sites of quality a. Production exhibits external economies of scale with constant elasticity ϵ, which are balanced by increasing urban costs with constant elasticity γ. Within that (purposefully) simple model, we provide a full characterization of: (i) the decentralized equilibria; (ii) the market outcome with competitive or strategic land developers; (iii) the 'locally optimal' solution as implemented by local governments with exclusionary power; and (iv) the globally optimal solution as implemented by a central planner. Whereas with homogeneous sites the outcomes with urban developers, local governments, and a central planner all coincide (Henderson, 1988), as does the Pareto efficient decentralized equilibrium, this no longer holds with heterogeneous sites.

We find, first, that there are decreasing returns to the urban system at the optimal allocation (there are constant returns with homogeneous land). The reason is that a larger population requires less favorable sites to be put into use (extensive margin), and a population increase in existing cities which operate on the decreasing returns portion of utility (intensive margin). Second, heterogeneity in developable land may prevent land developers from achieving the first-best allocation, because land heterogeneity relaxes competition forces. This becomes important when developers behave strategically, for example because they control a sufficient mass of sites to have an influence on the market outcome. Third, competition among land developers ensures that the best locations host cities in equilibrium. Fourth, the central government can implement the globally optimal allocation with the help of a proportional income tax across cities. The intuition is that the proportional tax serves as a Pigouvian subsidy which fully internalises the net external effect and allows to equalise the net marginal social benefits across all developed sites. Last, we show that the equilibrium allocations with competitive land developers and local governments widely differ when sites are heterogeneous, which has potentially stark allocative efficiency implications.

On top of providing a simple framework in which we can rigorously analyse the optimal and equilibrium allocations across a system of heterogeneous sites, our results are important for at least two other reasons. First, researchers who use profit maximising land developers as an equilibrium selection device usually justify their choice because of its *normative properties*: the equilibrium allocation with competitive land developers is efficient. In practice, land developers may either not exist, or they may just be too small to develop large cities from scratch or in a significant way.[3] Here, we show that – as with imperfect competition among firms – the interactions between heterogeneous land (i.e., firms with different production costs) and strategic land developers (i.e., firms who can manipulate the market conditions) lead to a breakdown of the equivalence between the equilibrium and the optimal allocations.[4]

Second, the equilibrium allocations with competitive land developers and local governments are the same when land is homogeneous *but widely different when they are not.* We show that the allocation with local governments that maximise the well-being of their constituents (and who take in newcomers as long as it raises the utility of the incumbents) yields a very inefficient equilibrium. The reason is that when all cities have the individually optimal size, but different levels of well-being across locations because of heterogeneous quality of local amenities and the exclusionary power of local governments, residents of cities with high production amenities enjoy rents. Allowing additional residents in the best cities would have welfare effects that are second-order small for incumbents (by definition of the local optimum) but that would increase the well-being of the newcomers in a first-order manner (since these people leave locations that command a strictly lower level of utility). The policy implication is stark: green belt policies and other forms of land use regulations that in effect prevent urban development should not be too decentralised since they may give rise to a particularly harmful form of NIMBY-ism.

The remainder of the paper is organized as follows. Section 2 introduces notation, develops the canonical model, and shows that the optimal allocation, the equilibrium allocation with local governments, and the equilibrium allocation with competitive developers are all equivalent (Proposition 1). Section 3 relaxes the assumption of homogeneous land and establishes that this equivalence result breaks down (Propositions 2 and 5). We also show that a proportional nominal income tax may serve to implement the optimal allocation if the marginal tax rate is chosen appropriately (Propositions 3 and 4). Last, Section 4 concludes.

2 The canonical model: Homogeneous urban land

We build on the canonical model by Henderson (1974) to introduce notation and display its basic results. Consider a city of population size ℓ. We assume that per capita gross output in that city

[3]As noted by Helsley and Strange (2012, p.26): *"So the question becomes: are there real world institutions that correspond to such powerful city developers? Our answer is: no. Ascribing entrepreneurial incentives to city governments seems to be at odds with actual city government behavior. Even if local governments were fully entrepreneurial, the fragmentation of local government means that exclusion of the sort required for efficiency is not possible. Private developers are rarely large enough to be reasonably seen as controlling the development of entire cities."*

[4]See Helsley and Strange (1994) for an approach where city formation is modeled as a dynamic game where city developers move either simultaneously or sequentially and commit to sunk expenditures on public goods. In that setting, since the market for city development is not contestable, the equilibrium with developers is inefficient. In our model, market power originates from the heterogeneity of sites.

is equal to

$$y(\ell, a) = a\ell^{\epsilon}, \tag{1}$$

where $a > 0$ is a productivity parameter and ℓ^{ϵ} is a scale effect external to the representative firm. In equation (1), $\epsilon > 0$ is the elasticity of *agglomeration economies* with respect to city size. City life is also impeded by external congestion effects. We assume that per capita *urban costs*, given by ℓ^{γ}, are increasing in city size at a rate $\gamma > 0$.[5] There is a land market in each city and we assume that urban land is equally owned by the local residents. Individual rents usually vary across locations within the city but all rents that are collected in a city are then evenly redistributed to its residents. Thus, per capita urban costs ℓ^{γ} are net of land rents. Utility, or per capita output net of urban costs, is then equal to

$$u(\ell, a) = a\ell^{\epsilon} - \ell^{\gamma}. \tag{2}$$

We impose $\gamma > \epsilon$, which ensures that more than one city is developed in equilibrium. Note also that a can be viewed as a 'net' amenity that weights urban benefits relative to urban costs.[6]

2.1 Optimal allocation in a system of cities

Consider first the socially optimal allocation of a population of size L across a great many potential sites, all equally amenable to urban development.[7] It follows that all cities will be of the same size at optimum. The maintained hypothesis in the literature and in the current paper is that the planner cannot solve the market failures associated with the external urban economies and costs. This is the most reasonable assumption to make, for too little is still known about the precise microeconomic mechanisms that give rise to these urban (dis)economies. The planner can, however, influence the distribution of population across cities and she can develop new sites. Thus, the central planner simply chooses the number n and the size ℓ of cities. Formally, she solves the following program:

$$\max_{\ell, n, \mu} \ \mathcal{L}(\ell, n, \mu) \equiv n\ell u(\ell, a) + \mu(L - n\ell),$$

where μ is the Lagrange multiplier and $u(\ell, a) = a\ell^{\epsilon} - \ell^{\gamma}$ by (2). The interior solutions for the optimal size and optimal number of cities are

$$\ell_o = \left(\frac{\epsilon}{\gamma} a\right)^{\theta} \quad \text{and} \quad n_o = L\left(\frac{\epsilon}{\gamma} a\right)^{-\theta}, \quad \text{where} \quad \theta \equiv \frac{1}{\gamma - \epsilon} > 0. \tag{3}$$

[5] We are agnostic about the precise microeconomic foundations of both agglomeration economies and urban costs. Duranton and Puga (2004) survey a wide class of models that deliver (1) for the former. The Alonso-Muth-Mills monocentric model is a classic way to deliver the latter (Fujita, 1998; Duranton and Puga, 2014).

[6] In general, $u(\ell, a) = a\ell^{\epsilon} - b\ell^{\gamma} = b[(a/b)\ell^{\epsilon} - \ell^{\gamma}]$. Clearly, b in front of the square brackets plays no role in the problem as it is just a scaling. Hence, the coefficient a/b in front of the per capita output subsumes urban benefits relative to urban costs. We denote it by a for short.

[7] We impose the technical condition $L > (a\gamma/\epsilon)^{1/(\gamma-\epsilon)}$ for reasons that will become clear shortly.

Empirically, γ and ϵ seem both smaller than 0.1, so that $\theta > 1$ (Behrens et al., 2014; Combes et al., 2014). The optimal city size and utility level are increasing in the scope of agglomeration economies, ϵ, decreasing in urban congestion, γ, and increasing in urban productivity, a. The second order condition holds by virtue of $\gamma > \epsilon$.

Optimal cities operate at the efficient size, namely, there are constant returns to scale in this system of cities. A corollary of this is that per capita utility is invariant to the size of the population and marginal utility from a change in L equals average utility:

$$\mu_o = \frac{\partial \mathcal{L}}{\partial L} = \frac{\mathcal{L}(\ell_o, n_o, \mu_o)}{L} = u(\ell_o, a) = \frac{1}{\theta\epsilon}\left(\frac{\epsilon}{\gamma}a\right)^{\theta\gamma}. \tag{4}$$

The central planner can always create additional optimal cities on identical sites to accommodate a growing population.

2.2 Decentralised equilibria

Consider next a decentralised equilibrium. The central result here is that coordination failures usually arise. Normalising the outside option of workers to zero, any allocation with city sizes ℓ_e such that, for all cities, $u(\ell_e, a) = u_e$ with some $0 \leq u_e \leq u_o$, and $(\partial u/\partial \ell)(\ell_e, a) \leq 0$, is a stable equilibrium. Therefore, there exists a continuum of equilibria with $\ell_e \in [\ell_o, \ell_m]$, where ℓ_o is defined in (3) and where the maximum city size ℓ_m consistent with an equilibrium equals

$$\ell_m = a^\theta > \left(\frac{\epsilon}{\gamma}\right)^\theta a^\theta = \ell_o. \tag{5}$$

Hence, in the decentralised equilibrium cities are oversized by a factor up to $(\gamma/\epsilon)^\theta$ (Henderson, 1974). From $\ell_o \leq \ell_e \leq \ell_m$ and from the total population constraint, the equilibrium number of cities n_e obeys $n_e \in [n_m, n_o]$, where the minimum equilibrium number of cities is equal to $n_m = L/\ell_m = L/a^\theta$. Hence, there are usually too few cities in equilibrium.

2.3 Competitive land developers

Consider next the case of competitive land developers. Each developer owns all the land of a potential city and collects local land rents. In this context, it is worth distinguishing between gross and net urban costs. Land rents are redistributed to city residents in (2) so that net per capita urban costs, equal to ℓ^γ, are lower than gross per capita urban costs, equal to $(1+\delta)\ell^\gamma$ for some $\delta > 0$ which may depend on city size ℓ (but does not in the Alonso-Muth-Mills model).[8] Land developers collect all land rents in the city, $R = \ell \times \delta\ell^\gamma$, and they attract households by setting individual subsidies T (a mnemonic for 'transfers') in order to maximize their profits:

$$\pi \equiv R - T\ell. \tag{6}$$

[8] In our simple model, assuming a monocentric structure with commuting costs td^γ at distance d from the center, we have $\delta = t\gamma/(\gamma+1)$.

Workers flock into a city as long as $u(\ell, a, T) = a\ell^\epsilon - (1+\delta)\ell^\gamma + T \geq u_d$, where u_d is the utility they can obtain in other cities. Developers take u_d as given, i.e., they are 'utility takers'.

Profit maximising land developers set workers at their reservation utility in equilibrium so that $T = u_d + (1+\delta)\ell^\gamma - a\ell^\epsilon$. Plugging this into (6) yields $\pi = \ell(a\ell^\epsilon - \ell^\gamma - u_d)$. The first-order condition to the developer's program, $\max_\ell \pi(\ell, a, u_d) \equiv \ell(a\ell^\epsilon - \ell^\gamma - u_d)$, is

$$\frac{\partial \pi}{\partial \ell}(\ell_d, a, u_d) = a(1+\epsilon)\ell_d^\epsilon - (1+\gamma)\ell_d^\gamma - u_d = 0, \tag{7}$$

where the subscript d stands for 'developers'. Free-entry among developers drives their profits to zero and this yields $a\ell_d^\epsilon - \ell_d^\gamma = u_d$ at the symmetric, unique equilibrium. Together with (7), we finally have

$$\ell_d = \ell_o = \left(\frac{\epsilon}{\gamma}a\right)^\theta \quad \text{and} \quad u_d = \mu_o = \frac{1}{\theta\epsilon}\left(\frac{\epsilon}{\gamma}a\right)^{\theta\gamma}$$

by (3) and (4). Hence, we see that the equilibrium with competitive land developers is unique and socially optimal.[9]

2.4 Local governments

Finally, assume that local governments – which have the power to exclude people from the city by using, e.g., urban growth bounds – maximise the utility of their residents, ignoring the effects on residents of other cities. Voters are happy to expand their city as long as the marginal benefit of doing so outweights the cost. Local authorities then choose ℓ_g (the subscript g is a mnemonic for 'government') so as to maximise average (and median) well-being u in (2). The solution to this problem is $\ell_g = \ell_o$, that is, local governments chose the socially optimal city size provided in (3), despite their disregard for the consequences of their choice on the rest of the economy. This result is expected since we have seen before that the optimal city size that a planner would choose is independent of L. The only margin of adjustment is the number of cities. With developers, the optimal city size is still independent of L, and the margin of adjustment is free entry (to create as many cities as are needed).

2.5 Summary

So far, we have shown the following:

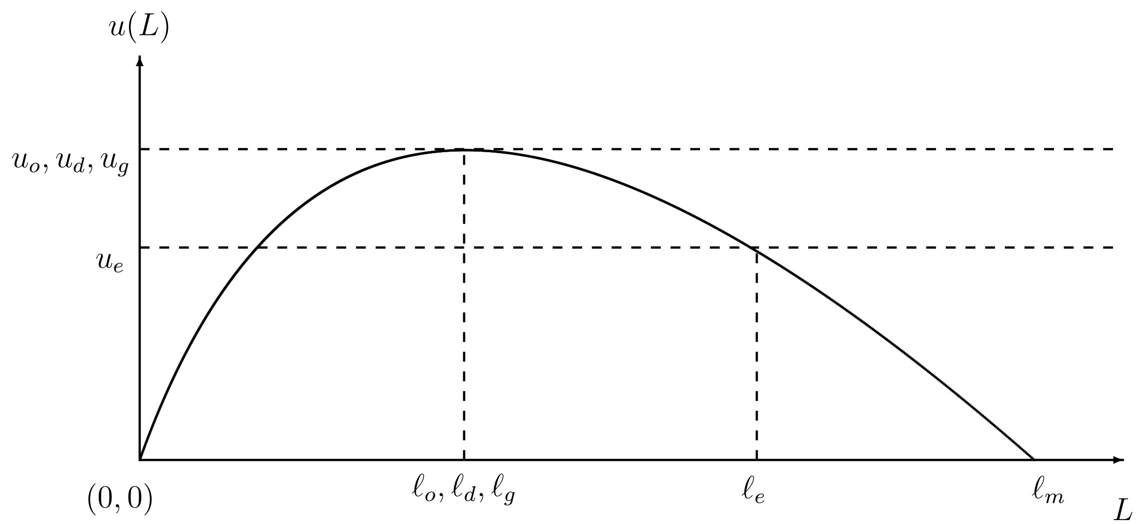
[9]This result can be viewed as a special instance of the Henry George Theorem, i.e., the 'golden rule' of local public finance (Flatter et al., 1974; Arnott and Stiglitz, 1979; Kanemoto, 1980). A non-distortionary 100% tax on local land rents raises enough revenue for developers to subsidize cities at their efficient size. We will see later that the same holds true irrespective of whether sites are heterogeneous or not. We will also show that efficiency can be implemented globally in our setting through a federal income tax that acts as a Pigouvian subsidy to internalize the externalities generated by agglomeration and congestion effects.

Proposition 1 (Allocations with homogeneous sites) *Consider a perfectly elastic supply of homogeneous sites, all equally amenable to urban development. Then: (i) the individually optimal city size, the globally optimal city size, and the city size at the equilibrium with competitive land developers all coincide. City sizes and individual utility are given by* (3) *and* (4)*. (ii) In the decentralised equilibriums, cities are weakly oversized by* (5) *and equilibrium utility is weakly lower than u_o, namely, $u_e \in [0, u_o]$.*

Proof. In the text. ∎

Figure 1 illustrates the foregoing results. Note, in particular, that the presence of competitive land developers solves the coordination failure that arises in a decentralised equilibrium. This is the reason why land developers have been widely used in the literature as an 'equilibrium selection device': not only is the equilibrium with developers unique, but it is also socially optimal. This result fundamentally changes once site heterogeneity is taken into consideration.

Figure 1: City sizes with homogeneous sites.



3 The extended model: Heterogeneous urban land

In the previous section all land was equally amenable to urban development. This assumption is convenient, unrealistic and, as we now show, far from innocuous. Assume that per capita consumption of land is exogenous and normalised to unity. To avoid a typology of uninteresting cases, we assume that the mass of developable urban land is large enough to accommodate the whole population.[10] There is a measure one of potential sites that can host cities. Land is of heterogeneous quality and we let a be distributed according to the cumulative distribution function $G(\cdot)$ over

[10] We do not deal with nonreplicability or indivisibility in the number of sites. See, for example, Papageorgiou and Pines (2000) for a discussion and treatment of these cases.

the unit interval by choice of units. For simplicity, we assume that $G(\cdot)$ is twice continuously differentiable.

In what follows, we again look at the globally optimal, the individually optimal, and the decentralised allocations. Whereas we used the subscripts o, g, end e with homogeneous sites, we now subscript the corresponding allocations with heterogeneous sites by 0, G, and E.

3.1 Optimal allocation in a system of cities

The central planner's problem is to maximise total output net of total urban costs by choosing which sites to develop as cities and how many individuals to locate there. Both choices are summarized by a function $\ell : a \to \ell(a)$ that allocates people to sites. We may thus write the central planer's problem as follows:

$$\max_{\ell(a),\mu} \mathcal{L} \equiv \int_0^1 \ell(a)u(\ell(a),a)\mathrm{d}G(a) + \mu \left[L - \int_0^1 \ell(a)\mathrm{d}G(a)\right], \tag{8}$$

where $u(\ell(a),a) = a\ell(a)^\epsilon - \ell(a)^\gamma$ by (2). The first order conditions with respect to $\ell(a)$, evaluated at the optimal allocation are:

$$a(1+\epsilon)\ell_0(a)^\epsilon - (1+\gamma)\ell_0(a)^\gamma - \mu_0 \leq 0, \quad \ell_0(a) \geq 0, \tag{9}$$

with complementary slackness. Equation (9) is the standard condition that states that the net marginal social product of labour must be equalized across all occupied sites (Flatters et al., 1974; Arnott and Stiglitz, 1979).

The first order condition with respect to μ yields

$$L = \int_0^1 \ell_0(a)\mathrm{d}G(a).$$

In addition, observe that

$$\mu_0 = \frac{\partial}{\partial L}\mathcal{L}\left(\ell_0(a),\mu\right), \tag{10}$$

at the optimal allocation from the envelope theorem. We are in position to show our first fundamental result:

Proposition 2 (Optimal allocation with heterogeneous sites) *There exists an* $a_0 \in [0,1)$ *such that* ***(i)*** $L = \int_{a_0}^1 \ell_0(a)\mathrm{d}G(a)$,

$$\ell_0(a) > 0 \textit{ for all } a \geq a_0 \quad \textit{and} \quad \ell_0(a) = 0 \textit{ otherwise;}$$

(ii) *for all* $a > a_0$*, the optimal city size* $\ell_0(a)$ *in the system of cities with heterogeneous land is increasing in* a*,*

$$\frac{\partial \ell_0(a)}{\partial a} > 0;$$

(iii) *is strictly larger than the individually optimal city size* $\ell_o(a)$ *in* (3) *and strictly lower than the maximum equilibrium size in* (5),

$$\ell_o(a) < \ell_0(a) < \ell_m(a); \tag{11}$$

(iv) *the smallest city has the individually optimal size:*

$$\ell_0(a_0) = \ell_o(a_0);$$

(v) *(gross) individual utility is increasing in* a,

$$\frac{\partial u\left(\ell_0(a), a\right)}{\partial a} = \frac{\gamma - \epsilon}{1+\epsilon}\ell_0(a)^{-1+\gamma}\frac{\partial \ell_0(a)}{\partial a} > 0;$$

(vi) *the nationwide urban system features decreasing returns to scale,*

$$\frac{\mathcal{L}_0}{L} > \frac{\partial \mathcal{L}_0}{\partial L};$$

and ***(vii)*** *the decreasing returns happen at both the intensive and the extensive margins of the urban system:*

$$\frac{\mathrm{d}a_0}{\mathrm{d}L} < 0 \quad \text{and} \quad \frac{\mathrm{d}\ell_0(a)}{\mathrm{d}L} > 0.$$

Proof. See Appendix A. ■

Part (i) of Proposition 2 says that not all sites are developed at the optimal allocation. The better sites are developed in priority, and part (ii) establishes that the better of the developed sites host a larger urban population. This is consistent with the evidence that available developable land and other geographical features are key determinants of local urban development (Saiz, 2010). Part (iv) says that the least productive developed site has the individually optimal size. The intuition is that, given a continuum of sites, the planner could always create additional cities with a just below a_0 (which are not developed yet) to absorb the excess population of the smallest city. If the smallest cities are oversized, reducing their sizes increases the utility of all agents there, whereas the agents that are relocated to a marginally worse city (in terms of smaller a) see their utility barely change (or even increase). Note that the continuum assumption is key here. For every location productivity a, $n(a) = g(a)/\ell(a)$ cities are being developed, and thus the planner can always redistribute the excess population of the smallest cities to any density of cities on marginally less favorable sites.

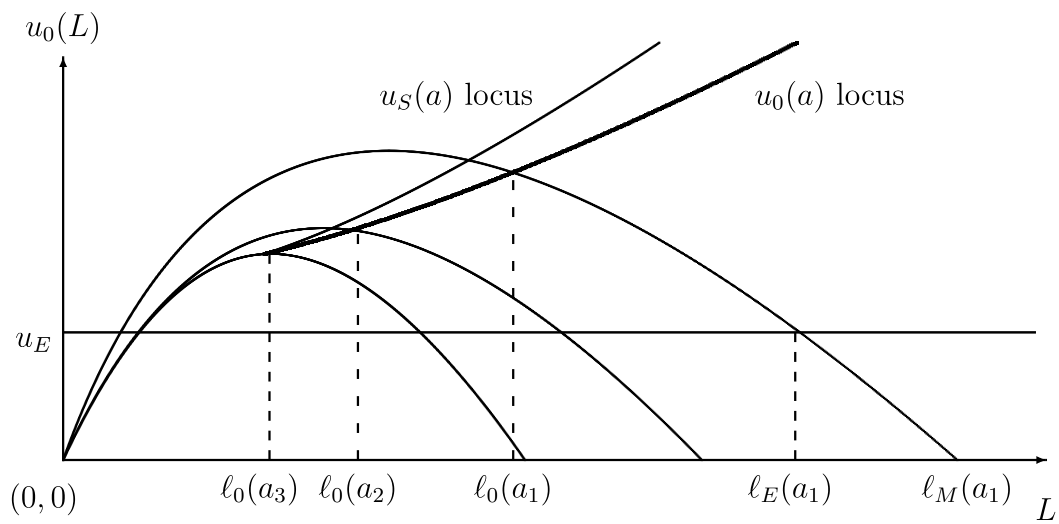
Part (iii) says that all other developed sites above the smallest city are larger than their individually optimal size. This result is fundamental because it shows that achieving an optimal city size locally has a negative effect on other cities: residents who are denied room in a city increase congestion elsewhere. Consider the largest city to fix ideas and assume that its size is locally optimal. Its residents enjoy a strictly higher utility than residents anywhere else, because this site has the best underlying fundamentals a. This allocation cannot be optimal from the point of view of the system as a whole. Thus, the largest city must be 'individually oversized' at the optimal allocation. A similar reasoning holds for all other cities, too, but the one at the bottom of the distribution. This city has, as we have argued, the individually optimal size.

Finally, part (v) says that individual utility is increasing in a and, together with (ii), this implies that individuals that are allocated to large cities are better off than citizens of smaller ones. Note that this property of the efficient allocation makes it inconsistent with our definition of a spatial equilibrium in the absence of mobility frictions. The reason for this result is that efficiency requires equalising the net social costs of an addition urban dweller in each city; this marginal net social cost differs from the average net social cost because the nationwide urban system features decreasing returns to scale by (vi). To establish this, let us evaluate aggregate welfare at the optimum. Integrating (9) over the support of a and plugging the outcome into (8) yields

$$\mathcal{L}_0 = \mu_0 L + \int_0^1 \ell_0(a) \left[\gamma \ell_0(a)^\gamma - a\epsilon \ell_0(a)^\epsilon\right] \mathrm{d}G(a). \tag{12}$$

Together with (3) and (11) and by the second order condition of the optimization problem, the terms in the square parenthesis above are negative, which implies that average utility $\mathcal{L}_0/L$ is larger than marginal utility μ_0, i.e., *there are decreasing returns to scale in the system of cities.* This is in contrast to Section 2, where returns to scale are constant by (4). The reason for this contrasting result is that, by part (vi), land is of heterogeneous quality here and accommodating a growing population requires adding newcomers to existing cities (intensive margin) as well as developing new cities (extensive margin). Both lower average utility: the intensive margin, because cities operate on the decreasing returns portion, and the extensive margin, because the best locations are developed first.[11]

Figure 2: City sizes with heterogeneous sites.



[11] In Henderson (1988, p.176), sites of different quality are assumed to be 'replicable', so that the urban system with heterogeneous cities still displays constant returns in the aggregate provided the number of cities is large. The key difference with our result is that we consider a fixed supply of sites $g(a)$ of any given type a.

Figure 2 illustrates the optimal allocation for three cities, where the city with production amenity a_3 is the smallest city being developed at the optimal allocation (and it has the individually optimal size). The figure displays the optimal city sizes for cities with production amenities a_1 and a_2, with $a_1 > a_2 > a_3$ (these are also the equilibrium city sizes with competitive land developers; more on this below). City a_1 is the largest city in the optimal urban system.

3.2 Decentralised equilibria

Consider next a decentralised equilibrium. As in Section 2, coordination failure generally arises. Any allocation such that, for all sites that develop cities, $u(\ell, a) = u_E$, for some $0 \leq u_E \leq u_0$, and $\partial u(\ell, a)/\partial \ell \leq 0$, is a stable equilibrium. Therefore, there exists a continuum of symmetric equilibria with $\ell_E(a) \in [\ell_0(a), \ell_M(a)]$, where $\ell_0(a)$ is implicitly defined in (9) and the maximum city size $\ell_M(a)$ consistent with an equilibrium is equal to $\ell_M(a) = a^\theta > \ell_0(a)$. Again, cities are oversized and there are too few of them in the decentralised equilibrium, as in the case of homogeneous sites in Section 2.

Figure 2 illustrates the maximum city size that may arise in a decentralised equilibrium with extreme coordination failure $\ell_M(a_1)$.

3.2.1 Equilibrium allocation and size distribution

Noting that the equilibrium allocation is implicitly defined by

$$\ell_E(a) = \left[a - \frac{u_E}{\ell_E(a)^\epsilon}\right]^{\frac{1}{\gamma-\epsilon}}, \tag{13}$$

we can derive results on the equilibrium size distribution of cities (see Behrens and Robert-Nicoud, 2014). Indeed, equation (13) shows that $\ell_E(a)$ is smaller than, but gets closer to $a^{1/(\gamma-\epsilon)}$, when $\ell_E(a)$ grows large (to see this, observe that $\lim_{\ell_E(a)\to\infty} u_E/\ell_E(a)^\epsilon = 0$). Therefore, *the upper tail* of the equilibrium city size distribution $\ell_E(a)$ inherits the properties of distribution of site characteristics in the same way as $\ell_0(a)$ and $\ell_G(a)$ do. In other words, the distribution of a is crucial for determining the distribution of equilibrium sizes of large cities. If very good sites are very scarce, large cities will be very scarce and the city size distribution may display properties well approximated by Zipf's law. To summarize, there are a continuum of equilibria, but in any equilibrium the upper tail of the equilibrium size distribution of cities inherits the properties of the distribution of the site parameters a.

3.2.2 Pareto efficiency and taxation

As argued before, there is a continuum of equilibria with asymmetric sites. To derive sharper results, we now focus on one of those equilibria: the Pareto efficient one. We can then prove the following results.

Proposition 3 (Pareto efficient decentralised equilibrium with heterogeneous sites) *Consider a decentralised equilibrium with the following properties:*

1. $\ell_E(a) = 0$ *for all* $a < a_E$*, for some* $a_E > 0$*;*

2. $\ell_E(a_E) = \ell_o(a_E) = (a_E\epsilon/\gamma)^{1/\theta}$ *(i.e. the smallest city has the individually optimal size).*

Then: **(i)** $a_E > a_0$ *(too few sites are developed in equilibrium);* **(ii)** *this decentralised equilibrium Pareto dominates all others;* **(iii)** *there exists a unique equilibrium with these properties; and* **(iv)** *there exists a unique threshold* $\widehat{a} \in (a_E, 1)$ *such that* $\ell_E(a) \gtreqless \ell_0(a)$ *for* $a \gtreqless \widehat{a}$*, and the oversize* $\ell_E(a)/\ell_0(a)$ *is increasing in* a*.*

Proof. See Appendix B. ■

Proposition 3 shows that the market allocation is skewed towards the more productive sites. The reason is that more productive sites are bigger because they attract more mobile workers, but more mobile workers in turn impose a negative congestion externality on a larger mass of other urban dwellers in big cities. The result is that more productive sites are more severely oversized than less productive sites. One may then ask whether a simple policy tool that imposes larger costs for agents in more productive sites can serve to correct the market failure. We can prove that the answer to this question is clearly 'yes' in our simple setup.

Proposition 4 (Taxation and optimum allocations) *Consider a decentralized equilibrium and assume that the central government imposes a uniform income tax at rate* t*. Then the Pareto efficient decentralized equilibrium and the socially optimal allocation coincide (and* $U_t = U_0$*) if and only if*

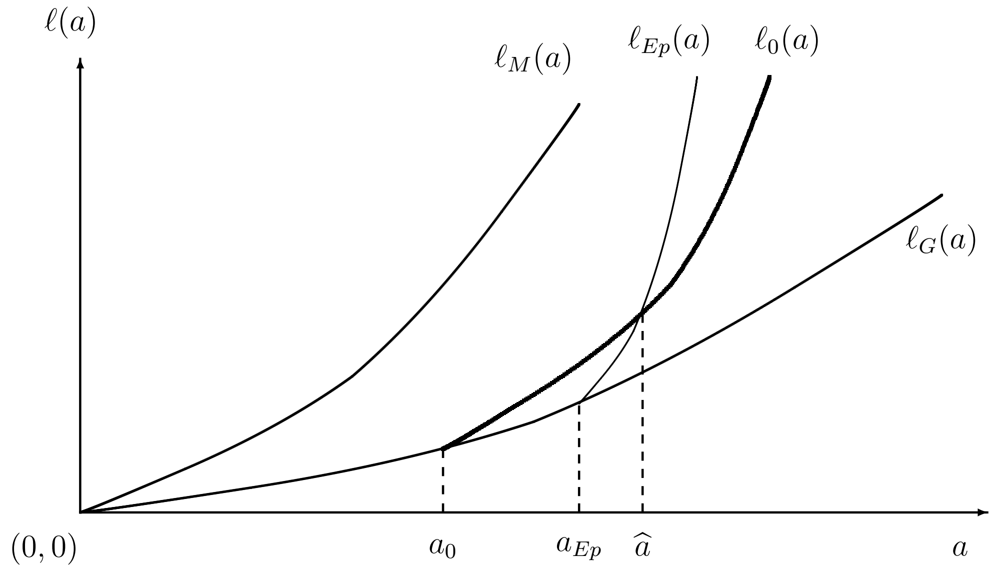
$$t = \frac{\gamma - \epsilon}{1 + \gamma}.$$

Hence, a proportional income tax works like a Pigouvian subsidy and is enough to implement an efficient outcome.

Proof. See Appendix C. ■

Several comments are in order. First, the foregoing result shows that, by imposing a uniform tax, the planner imposes an increasing effective tax on agents in bigger cities. The reason is that large cities are expensive places, so that a dollar is worth less there. Imposing the same nominal tax rate then amount to having a higher effective tax rate. This is in essence the point emphasized by Albouy (2009). Second, to implement the optimal allocation, the planner increases the equilibrium number of sites (recall that $a_0 < a_E$) and redistributes population from large cities to smaller places (see Figure 3, where we depict both the Pareto efficient equilibrium allocation ℓ_{Ep} and the socially optimal allocation ℓ_0). This is due to the fact that the oversize increases in a, which requires generally more and smaller cities – except for the smallest ones, which are too small in equilibrium and need to increase their size. In other words, the equilibrium allocation is over-skewed.[12]

Figure 3: City sizes as a function of site characteristics a.



Note that, unlike in Albouy (2009), large cities are oversized and the oversize is itself increasing in city size via a. Thus, a centralised (federal) income tax proportional to nominal income may achieve the optimal allocation if the rate is chosen optimally. Using $\gamma = 0.08$ and $\epsilon = 0.03$ (as in Behrens et al., 2014) yields $t = 4.63\%$, which is well below current OECD income tax rates. The optimal tax rate would be smaller still with more reasonable values for γ and ϵ (Combes, Duranton, and Gobillon, 2014). In sum, in our setting, *Albouy's (2009) result is qualitatively overturned but it remains quantitatively correct.*

3.3 Land developers

We now turn to studying the equilibrium allocation in the presence of local land developers. Contrary to the case of homogeneous land, a developer who owns site a now owns a rare asset: this land is different from all others and this is a source of Ricardian rent. As we shall see, site heterogeneity relaxes competitive forces and the decentralised equilibrium need no longer be efficient, unlike in the model of Section 2.

Like in Section 2, local land developers may attract urban dwellers only if they set (possibly negative) taxes such that the net utility is weakly larger than the economy-wide utility level, denoted by u_D. By the same token as in Section 2, local land developers set agents at their reservation

[12] In terms of the size distribution of cities, the optimal allocation would have a smaller Zipf coefficient than the equilibrium allocation.

utility and maximise

$$\max_{\ell(a)} \pi(\ell(a), a, u_D) \equiv \ell(a)[a\ell(a)^{\epsilon} - \ell(a)^{\gamma} - u_D]. \tag{14}$$

Denote with the subscript 'D' the 'Developer' equilibrium. The first-order conditions for this program are given by

$$a(1+\epsilon)\ell_D(a)^{\epsilon} - (1+\gamma)\ell_D(a)^{\gamma} - u_D\left[1 + \varepsilon(u_D, \ell_D(a), \cdot)\right] = 0, \tag{15}$$

if $\ell_D(a)$ is positive and the term on the left-hand side is negative otherwise. In (15), ε denotes the elasticity of u_D with respect to $\ell(a)$ and the dot stands for the equilibrium choices of other developers. Observe that this formulation allows for both non-strategic, atomistic developers who disregard the impact of their individual decision on the aggregate variable u_D (like monopolistically competitive firms do), as well as strategic developers who take account of the consequences of their behaviour on the outside option of workers u_D.[13]

Utility in any city is decreasing in its own size in any stable equilibrium. Thus, when a land developer takes in more people from other cities, this raises their reservation utility, namely: $\partial u_D(\ell(a), \cdot)/\partial \ell(a) > 0$ (here we abuse notation because, formally, developers choose taxes and subsidies, not city size directly, and the latter add up to a constant).

We can now show our second set of fundamental results:

Proposition 5 (Land developers with heterogeneous sites) *Let land be heterogeneous and land developers maximise profits. Then:* ***(i)*** *the equilibrium and optimal solutions coincide if and only if land developers are utility takers:*

$$\ell_D(a) = \ell_0(a) \quad \Leftrightarrow \quad \varepsilon(u_D, \ell(a), \cdot) \equiv \frac{\partial u_D(\ell(a), \cdot)}{\partial \ell(a)} \frac{\ell_D(a)}{u_D(\ell(a), \cdot)} = 0.$$

The equilibrium profits of land developers are strictly positive for all $a > a_0$. ***(ii)*** *If land developers are strategic (utility makers), then the equilibrium cities are smaller and more numerous than at the optimal allocation.*

Proof. See Appendix D. ■

Figure 2 illustrates the result. As shown above, $u_D(a) = u_0(a)$ and $\ell_D(a) = \ell_0(a)$ if developers are non-strategic. By contrast, $u_S(a) > u_0(a)$ and $\ell_S(a) < \ell_0(a)$ when developers are strategic, where $u_S(a)$ is the utility gross of the transfer T (recall that the utility after transfer must be the same across cities).

[13] Strictly speaking, our continuum formulation implies that developers are negligible. Hence, they have no impact on the economy-wide utility u_D. For developers to behave strategically, they need to be atoms, i.e., they need to control a measurable subset of sites. In that case, developers are utility makers and they take into account the way their actions affect the market (including potential 'cannibalisation effects' among sites they own). This way of viewing the problem is formally analogous to the models of multiproduct firms or 'mixed market' models of imperfect competition.

Three comments are in order. First, in this paper we are working with a continuum of locations and, therefore, a continuum of land developers. The only rigorous assumption to make in this setting is that land developers are atomistic. In this case, the equilibrium allocation with land developers is optimal. However, assuming a continuum of locations and agents is a technical assumption made for convenience, not because it is a realistic one. Assume instead that the set of potential sites (and the set of land developers that goes with it) is discrete and consider the following game. First, a non binding number of potential land developers decide to enter (develop a city on their site) or not. Second, the entrants compete for urban dwellers as in the rest of the paper. The equilibrium concept to this game is subgame perfect equilibrium. When land is homogeneous, the equilibrium profits of land developers are zero and the equilibrium number and size of cities are optimal. Yet, in the case of heterogeneous land, only the developers owning land with a high a put their land to urban use and all but the marginal developer earn positive profit in equilibrium by Proposition 5. Thus, entry limits the inefficiency associated with the strategic behaviour of land developers without eliminating it. *The equilibrium with land developer is generically suboptimal when land is heterogeneous and when developers have some market power.*

Second, auctioning land to competitive land developers would not solve the problem. Doing so only ensures that profits are redistributed to the central planner rather than kept in the hands of private land developers.

Last, it is useful to make the analogy between our problem at hand and that of imperfect competition among firms. In our model, developers differ by 'productivity' (i.e., the quality of the site they own), but varieties (i.e., sites) are viewed as perfect substitutes by mobile workers (consumers). The developer with the lowest cost (i.e., the best site) cannot capture the whole market because her production cost is convex in city size (because of increasing urban costs). This limits the size of any city. Since all agents eventually have to end up in some location (they cannot 'opt out' of the urban system), this implies that less efficient developers also end up developing sites. The outcome with developers is optimal when developers cannot strategically manipulate u_D, because for any site there always exists another one that is arbitrary similar which constrains developers since sites are viewed as perfect substitutes by mobile agents.

3.4 Local governments and local optimum

We finally assume, as in Section 2, that local governments maximise the utility of their residents, ignoring the effects of their local policy choices on residents of other cities. These local authorities then choose the locally optimal city size in (3):

$$\ell_G(a) = \left(a\frac{\epsilon}{\gamma}\right)^{\theta}, \quad \text{which yields} \quad u_G(a) = \frac{1}{\theta\epsilon}\left(\frac{\epsilon}{\gamma}a\right)^{\theta\gamma}.$$

Both are naturally increasing in a. As shown in Proposition 2, these locally optimal cities are too small relative to the globally optimal ones. This is because the most productive cities keep out potential newcomers to maintain the rents of their incumbents. This lowers welfare for everybody else. This form of NIMBY-ism is ubiquitous and especially prevalent in the United Kingdom (e.g. Cheshire and Sheppard, 2002). It usually takes the form of restrictive land use regulations. This result is also consistent with the estimations in Hilber and Robert-Nicoud's (2013): they find that US locations endowed with desirable amenities host a larger urban population and, as a consequence of endogenous political economy forces, more regulated. The normative implication of the model is then that green belts and other forms of land use regulations that heavily restrict urban development should not be in the hands of local authorities. This is also the conclusion of Vermeulen (2011), although cities in his model are usually too small and too many, whereas cities are too large and too few in our setting. Regardless of the specifics of the model, the bottom line is that decentralised land use regulations are likely to impose significant welfare costs since the intercity externalities are not internalised.

4 Conclusions

The characterization of urban systems in the presence of heterogenous land and of their differences under various institutional settings – central government, local government, land developers, laisser faire – is a surprisingly complex task. As evidence of this, the current literature focuses either on some specific institutions or on special cases. A recurrent finding is that actual urban systems are likely suboptimal since the congestion cost net of agglomeration economies almost certainly differ across cities at the margin since few, if any, of these institutions fully internalise these external costs and benefits.

In order to make progress, this paper takes a bold route: we assume a continuum of sites and a continuous population, which eliminates the integer constraints, and we assume specific functional forms for preferences, agglomeration economies, and congestion costs. These assumptions are in line with canonical model developed by Henderson (1974) and widely used in the structural empirical literature.

Against this loss of generality, we are able to fully characterize the equilibrium urban systems under the aforementioned institutional settings and to compare the equilibrium outcomes among themselves and against the socially optimal allocation. Beyond these purely theoretical considerations, this exercise allows us to make sense of the use and abuse of land use regulations, of the risk of NIMBY-ism when land use regulations are in the hands of local authorities, and of the inefficiencies that arise with strategic land developers.

In ongoing work, we intend to delve further into the issue of strategic land developers in the presence of heterogenous land. Indeed, with a continuum of locations, homogenous population, costless mobility, and the strategy set of developers consisting in transfers, the outcome is necessarily efficient because land developers are monopolistically competitive and the equilibrium is socially efficient. We need to relax the assumption of atomistic land developers – so that they become strategic – and to allow for mobility frictions such as heterogenous preferences for locations in the population – so as to avoid extreme Bertrand competition among them land developers.

References

[1] Abdel-Rahman, Hesham M., and Alex Anas (2004). "Theories of Systems of Cities." In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, Elsevier: North-Holland, pp. 2293–2339.

[2] Abdel-Rahman, Hesham M., and Masahisa Fujita (1993). "Specialization and Diversification in a System of Cities." *Journal of Urban Economics*, 3**(2)**, 189–222.

[3] Albouy, David Y. (2012). "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas." Processed, University of Illinois.

[4] Albouy, David Y. (2009). "The Unequal Geographic Burden of Federal Taxation." *Journal of Political Economy* 117**(4)**, 635–667.

[5] Anas, Alex (1992). "On the Birth and Growth of Cities: Laissez-faire and Planning Compared." *Regional Science and Urban Economics* 22**(2)**, 243–258.

[6] Arnott, R. and J.E. Stiglitz (1979). "Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size." *Quarterly Journal of Economics* 93**(4)**, 471–500.

[7] Aumann, Robert J. (1964). "Markets with a Continuum of Traders." *Econometrica* 32**(1)**, 39–50.

[8] Behrens, Kristian, Gilles Duranton, and Frédéric L. Robert-Nicoud (2014). "Productive Cities: Sorting, Selection and Agglomeration." *Journal of Political Economy* 122 **(3)**, 507–53.

[9] Behrens, Kristian, and Frédéric L. Robert-Nicoud (2014). "Agglomeration Theory with Heterogeneous Agents." In: Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics*, vol. 5, Elsevier: North-Holland, forthcoming.

[10] Bleakly, Hoyt and Jeffrey Lin (2012). "Portage and Path Dependence." *Quarterly Journal of Economics* 127 **(2)**, 587–644.

[11] Cheshire, Paul and Stephen Sheppard (2002). "The welfare economics of land use planning." *Journal of Urban Economics* 52, 242–269.

[12] Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2014). "The Costs of Agglomeration: Land Prices in French Cities." *In progress*, Wharton School, University of Pennsylvania.

[13] Davis, Donald R., and David E. Weinstein (2002). "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review* 92**(5)**, 1269–1289.

[14] Desmet, Klaus, and Esteban Rossi-Hansberg (2013). "Urban Accounting and Welfare." *American Economic Review* 103**(6)**, 2296–2327.

[15] Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson (1977). "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods." *American Economic Review* 67**(5)**, 823–839.

[16] Duranton, Gilles and Diego Puga (2014). "Urban Land Use." In: Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics*, vol. 5, Elsevier: North-Holland, forthcoming.

[17] Duranton, Gilles and Diego Puga (2004). "Micro-foundations of Urban Agglomeration Economies." In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, Elsevier: North-Holland, pp. 2063–2117.

[18] Eaton, Jonathan, and Samuel Kortum (2002). "Technology, Geography, and Trade." *Econometrica* 70**(5)**, 1741–1779.

[19] Flatters, Frank, J. Vernon Henderson, and Peter Mieszkowski (1974). "Public Goods, Efficiency, and Regional Fiscal Equalization." *Journal of Public Economics* 3**(2)**, 99–112.

[20] Helsley, Robert W. and William C. Strange (2014). "Coagglomeration and the Scale and Composition of Clusters." *Journal of Political Economy*, forthcoming.

[21] Helsley, Robert W. and William C. Strange (1994). "City Formation with Commitment." *Regional Science and Urban Economics* 24**(3)**, 373–390.

[22] Henderson, J. Vernon, and Anthony J. Venables (2009). "The Dynamics of City Formation". *Review of Economic Dynamics* 12**(2)**, 233–254.

[23] Henderson, J. Vernon (1988). *Urban Development: Theory, Fact, and Illusion.* New York and Oxford: Oxford University Press.

[24] Henderson, J. Vernon (1986). "The Timing of Regional Development." *Journal of Development Economics* 23**(2)**, 275–292.

[25] Henderson, J. Vernon (1974). "The Sizes and Types of Cities." *American Economic Review* 64**(4)**, 640–656.

[26] Hilber, Christian A.L. and Frédéric Robert-Nicoud (2013). "On the Origins of Land Use Regulations: Theory and Evidence from US Metro Areas." *Journal of Urban Economics* 75, 29–43.

[27] Kanemoto, Yoshitsugu (1990). "Optimal Cities with Indivisibility in Production and Interactions Between Firms." *Journal of Urban Economics* 27**(1)**, 46–59.

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

[29] Papageorgiou, Yorgos Y., and David Pines (2000). "Externalities, Indivisibility, Nonreplicability, and Agglomeration." *Journal of Urban Economics* 48**(3)**, 509–535.

[30] Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90**(6)**, 1257–1278.

[31] Saiz, Albert (2010). "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics* 125**(3)**, 1253–1296.

[32] Vermeulen, Wouter (2011). "Agglomeration Externalities and Urban Growth Controls." SERC Discussion Paper #0093, Spatial Economics Research Centre, London School of Economics.

Appendix A: Proof of Proposition 2

The Lagrangian of the planner's optimisation problem is given by

$$\mathcal{L} = \int_0^1 \left[a\ell(a)^{\epsilon+1} - \ell(a)^{\gamma+1}\right] \mathrm{d}G(a) + \mu \left[L - \int_0^1 \ell(a)\mathrm{d}G(a)\right] + \int_0^1 \lambda(a)\ell(a)\mathrm{d}G(a),$$

which yields the following first-order conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mu} &= L - \int_0^1 \ell(a)\mathrm{d}G(a) = 0 \\
\frac{\partial \mathcal{L}}{\partial \ell(a)} &= a(\epsilon+1)\ell(a)^{\epsilon} - (\gamma+1)\ell(a)^{\gamma} - \mu + \lambda(a) = 0 \\
& \quad \lambda(a) \geq 0, \quad \ell(a) \geq 0, \quad \lambda(a)\ell(a) = 0
\end{aligned}$$

The second-order conditions associated with the choice of $\ell_0(a)$ are given by

$$\frac{1}{\ell_0(a)} \left[a(1+\epsilon)\epsilon\ell_0(a)^{\epsilon} - (1+\gamma)\gamma\ell_0(a)^{\gamma}\right] < 0. \tag{16}$$

(i) Bounds for $\ell_0(a)$. From the FOC for sites with positive population and $\mu_0 > 0$ we have

$$(1+\epsilon)a\ell_0(a)^{\epsilon} - (1+\gamma)\ell_0(a)^{\gamma} > 0$$

and, from the SOC,

$$\epsilon(1+\epsilon)a\ell_0(a)^{\epsilon} - \gamma(1+\gamma)\ell_0(a)^{\gamma} < 0$$

so that we obtain the following bounds for $\ell_0(a)$:

$$\frac{\epsilon}{\gamma}\frac{1+\epsilon}{1+\gamma}a < \ell_0(a)^{\gamma-\epsilon} < \frac{1+\epsilon}{1+\gamma}a$$

Note that the foregoing implies that $\ell_0(a) < \ell_m(a)$ because $\frac{1+\epsilon}{1+\gamma}a < a$.[14] We can derive sharper bounds for $\ell_0(a)$ as follows. Differentiating (9) with respect to $\ell(a)$ yields the following differential equation in $\ell(a)$:

$$0 = (1+\epsilon)\ell_0(a)^{\epsilon} + \frac{\ell_0'(a)}{\ell_0(a)}\left[a(1+\epsilon)\epsilon\ell_0(a)^{\epsilon} - (1+\gamma)\gamma\ell_0(a)^{\gamma}\right]. \tag{17}$$

The term in square brackets is negative by (16) and $\ell_0(a)$ is positive by definition. Together, these and (17) yield $\ell_0'(a) > 0$, i.e., optimal city sizes increase naturally with a.[15] It then follows by continuity and from $\mu_0 > 0$ and the first-order condition that there exists an $a_0 \in [0,1)$ such that $\ell_0(a) > 0$ if and only if $a \geq a_0$, whereas $l_0(a) = 0$ for all $a < a_0$.

(ii) Proof that the smallest city is of individually optimal size and additional bound. Totally differentiate the Lagrangian to get

$$\mathrm{d}\mathcal{L}_0 = \mu_0 \mathrm{d}L - [a_0 l_0(a_0)^{\epsilon+1} - l_0(a_0)^{\gamma+1}]g(a_0)\mathrm{d}a_0 + \mu_0 l_0(a_0) g(a_0)\mathrm{d}a_0$$

Since $\mathrm{d}\mathcal{L}_0/\mathrm{d}L = \mu_0$ from the envelope theorem, this is equivalent to

$$0 = l_0(a_0)\left[a_0 l_0(a_0)^{\epsilon} - l_0(a_0)^{\gamma} - \mu_0\right] g(a_0)\frac{\mathrm{d}a_0}{\mathrm{d}L} \tag{17}$$

which implies that $\mu_0 = a_0 l_0(a_0)^{\epsilon} - l_0(a_0)^{\gamma} = u_0(a_0)$. Plugging this into the first-order condition for $\ell(a_0)$, it follows that

$$a_0(\epsilon+1)\ell_0(a_0)^{\epsilon} - (\gamma+1)\ell_0(a_0)^{\gamma} - \mu_0 = 0$$

is equivalent to $a_0\epsilon\ell_0(a_0)^{\epsilon} - \gamma\ell_0(a_0)^{\gamma} = 0$. This establishes that the smallest city is at its individually optimal size, i.e., $\ell_0(a_0) = \ell_o(a_0)$. It then follows from the first-order condition for city a_0 that the Lagrange multiplier is

$$\mu_0 = \frac{1}{\theta\epsilon}\left(\frac{\epsilon}{\gamma}a_0\right)^{\theta\gamma}, \quad \text{with} \quad \frac{\mathrm{d}\mu_0}{\mathrm{d}a_0} > 0.$$

To see that $l_0(a) > l_o(a)$ for all $a > a_0$, take the first order conditions and rewrite them as follows:

$$\ell(a)[a\epsilon\ell(a)^{\epsilon-1} - \gamma\ell(a)^{\gamma-1}] + a\ell(a)^{\epsilon} - \ell(a)^{\gamma} = \mu_0 = u_0(a_0).$$

[14] One can also check that $\lambda < 0$ when evaluated at the maximum city size, which is incompatible with optimality since the planner can always reduce existing city sizes and create new cities to increase utility in that case.

[15] The transcendental differential equation in (16) does not admit a closed form solution.

If $\ell(a) = \ell_o(a)$, this boils down to

$$a\ell_o(a)^\epsilon - \ell_o(a)^\gamma = u_o(a) > \mu_0 = u_0(a_0)$$

since utility is increasing in a and by definition of individual optimal size. Since $\gamma > \epsilon$, it then follows that $l_0(a) > l_o(a)$ for all $a > a_0$.

We thus have shown that $l_o(a) \leq l_0(a) < l_m(a)$, with $l_o(a_0) = l_0(a_0)$.

(iii) Increasing net utility. In order to obtain the result that utility is increasing in a, use the first-order condition to write $(1+\epsilon)u(\ell_0(a), a) = \mu_0 + (\gamma - \epsilon)\ell_0(a)^\gamma$. Totally differentiating with respect to a, and recalling that μ_0 is independent of a, yields

$$(1+\epsilon)\left[\frac{\mathrm{d}u_0}{\mathrm{d}l_0}\frac{\mathrm{d}l_0}{\mathrm{d}a} + \frac{\mathrm{d}u_0}{\mathrm{d}a}\right] = (\gamma - \epsilon)\gamma l_0^{\gamma-1}\frac{\mathrm{d}l_0}{\mathrm{d}a}.$$

We know that $\frac{\mathrm{d}u_0}{\mathrm{d}l_0} < 0$ since $l_0(a) > l_o(a)$ and $\frac{\mathrm{d}l_0}{\mathrm{d}a} > 0$. Hence, $\frac{\mathrm{d}u_0}{\mathrm{d}a} > 0$ and $\frac{\mathrm{d}u_0}{\mathrm{d}l_0}\frac{\mathrm{d}l_0}{\mathrm{d}a} + \frac{\mathrm{d}u_0}{\mathrm{d}a} > 0$.

(iv) Decreasing returns in the urban system. Rewrite equation (16) as $\mathcal{L}_0 = \mu_0 L + \Delta_0$, where the remainder

$$\Delta_0 \equiv \int_{a_0}^{1} \left[a\ell_0(a)^\epsilon - \ell(a)^\gamma - \mu_0\right]\ell_0(a)\mathrm{d}G(a)$$

is strictly positive. To prove the latter, note that from the first-order conditions (9) we have

$$a\ell_0(a)^\epsilon - \ell(a)^\gamma - \mu_0 = -\left[a\epsilon\ell_0(a)^\epsilon - \gamma\ell(a)^\gamma\right] > 0,$$

where the last inequality comes from the fact that the term is zero at $\ell_0(a) = \ell_o(a)$, but we know that $\ell_0(a) > \ell_o(a)$ and that we are on the decreasing part of the utility curve. Since $\mathrm{d}\mathcal{L}/\mathrm{d}L = \mu_0$ from the envelope theorem, it immediately follows that

$$\frac{\mathcal{L}_0}{L} > \frac{\mathrm{d}\mathcal{L}_0}{\mathrm{d}L},$$

which establishes that there are decreasing returns to scale in the urban system.

(v) Margins of decreasing returns. Totally differentiating the full-employment condition yields

$$\mathrm{d}L = -\ell_0(a_0)g(a_0)\mathrm{d}a_0 + \int_{a_0}^{1} \mathrm{d}\ell_0(a)\mathrm{d}G(a),$$

where the first term in the RHS is the extensive margin and the second term is the intensive margin. Totally differentiating the expression for μ_0 yields

$$\mathrm{d}\mu_0 = \left[-1 + \frac{\epsilon}{\gamma}\frac{1+\epsilon}{1+\gamma}a\ell_0(a)^{-\gamma+\epsilon}\right](1+\gamma)\gamma\ell_0(a)^{-1+\gamma}\mathrm{d}\ell_0(a),$$

which implies that the sign of $\mathrm{d}\mu_0$ is the opposite of the sign of $\mathrm{d}\ell_0(a)$ by virtue of the second-order condition (the term in the square parenthesis is negative). Thus,

$$\frac{\mathrm{d}a_0}{\mathrm{d}L} < 0 \quad \text{and} \quad \frac{\mathrm{d}\ell_0(a)}{\mathrm{d}L} > 0$$

for all $a \in (a_0, 1)$.

Appendix B: Proof of Proposition 3

(i) Too few cities and increasing oversize. We prove this point by contradiction. Assume that $a_E \leq a_0$, i.e., more sites are developed in equilibrium than at optimum. Then $\ell_E(a_E) = \ell_o(a_E) = (a_E\epsilon/\gamma)^{1/\theta}$ from the individually optimal smallest size implies $\ell_E(a_0) > \ell_0(a_0) = \ell_o(a_0)$ by the stability condition for ℓ_E. In turn, $\ell_E(a) > \ell_0(a)$ for all $a \geq a_0$ by $u_0'(a) > 0$ for all $a > a_0$, whereas u_E is constant. Thus, equilibrium city sizes must increase more quickly than optimal city sizes, which also implies that the oversize $\ell_E(a)/\ell_0(a)$ is an increasing function of a. Given the foregoing, we have

$$L = \int_{a_0}^{1} \ell_0(a)\mathrm{d}G(a) < \int_{a_0}^{1} \ell_E(a)\mathrm{d}G(a) < \int_{a_E}^{1} \ell_E(a)\mathrm{d}G(a),$$

which violates the full-employment condition. This establishes a contradiction and $a_E > a_0$ must therefore hold. It then follows, by continuity and the population constraint and from $\ell_E(a_E) < \ell_0(a_E)$, that there must be $\widehat{a} > a_E$ such that $\ell_E(a) > \ell_0(a)$ for all $a > \widehat{a}$.

(ii) Uniqueness. Turning to uniqueness, we totally differentiate the full employment condition and get

$$\mathrm{d}L = -\ell_E(a_E)g(a_E)\mathrm{d}a_E + \int_{a_E}^{1} d\ell_E(a)\mathrm{d}G(a).$$

Differentiating the spatial equilibrium condition yields

$$\mathrm{d}u_E = \left[-\gamma\ell_E(a)^{\gamma} + \epsilon a\ell_E(a)^{\epsilon}\right] \mathrm{d}\ln \ell_E(a),$$

for all $a \geq a_E$. The term in the square parenthesis above is negative by the local stability condition. These together imply that a_E is unique and

$$\frac{\mathrm{d}a_E}{\mathrm{d}L} < 0 \quad \text{and} \quad \frac{\mathrm{d}\ell_E(a)}{\mathrm{d}L} > 0.$$

(iii) Pareto optimality. Assume that the equilibrium is not Pareto optimal, i.e. assume that there exists a $a_C \neq a_E$ such that $u(\ell_C(a), a) = u_C$, all $a \geq a_C$, $\ell_C(a) = 0$ for all $a < a_C$, and $u_C \geq u_E$. Then $\ell_C(a) < \ell_E(a)$ for all $a > a_C, a_E$ by the local stability condition. In turn, the full-employment condition requires $a_C < a_E$, but this implies $u_C(\ell_C(a_E), a_E) < u_E$, which contradicts $u_C \geq u_E$. Thus $u_C < u_E$.

Appendix C: Proof of Proposition 4

At a spatial equilibrium without mobility frictions, people live in city a as long as

$$u_E = u(\ell_E(a), a) \equiv a\ell_E(a)^\epsilon - \ell_E(a)^\gamma,$$

for some $u_E > 0$. At the socially optimal allocation, the following must hold:

$$\mu_0 = (1+\epsilon)a\ell_E(a)^\epsilon - (1+\gamma)\ell_E(a)^\gamma, \tag{1}$$

for some $\mu_0 > 0$. Implementing the socially optimal allocation with an income tax for mobile individuals is to impose a tax on nominal income with a common and constant marginal rate

$$u_t = u(\ell_E(a), a, t) \equiv (1-t)a\ell_E(a)^\epsilon - \ell_E(a)^\gamma, \tag{1}$$

for some $t \in (0, 1)$ and some $u_t > 0$. Choosing

$$t = \frac{\gamma - \epsilon}{1+\gamma} \tag{1}$$

achieves precisely the optimal allocation $\ell_0(a)$. To see this, insert (4) into (4) and observe that the outcome is equivalent to (4) for $u_t(1+\gamma) = \mu_0$.

We can then compare the equilibrium allocation with income tax with the socially optimal allocation. Using the first-order condition for $\ell_0(a)$, (4), and the definition of individual utility yields $u_0(a) = [\mu_0 + (\gamma - \epsilon)\ell_0(a)^\epsilon]$. Summing across the whole population yields the following aggregate welfare evaluated at the socially optimal allocation:

$$U_0 = \frac{\mu_0}{1+\gamma}L + \frac{\gamma-\epsilon}{1+\gamma}\int_{a_0}^1 a\ell_0(a)^{1+\epsilon}\mathrm{d}G(a).$$

By the same token, summing u_t in (4) and redistributing the income tax evenly across the whole population yields

$$U_t = \frac{\mu_0}{1+\gamma}L + \left(1 - t - \frac{1+\epsilon}{1+\gamma}\right)\int_{a_t}^1 a\ell_t(a)^{1+\epsilon}\mathrm{d}G(a),$$

where we have made use of (4).

The 'only if' part of Proposition 4 follows by inspection of (4) and (4). To see the sufficiency part, assume that $t = (\gamma - \epsilon)/(1+\gamma)$ but $a_t \neq a_0$. Then, this equilibrium is not Pareto optimal by an argument identical to that in the proof of Proposition 3, establishing a contradiction. Then $a_t = a_0$ and $\ell_t(a) = \ell_0(a)$ by (4) and (4).

Appendix D: Proof of Proposition 5

When developers are atomistic, condition (15) simplifies to

$$a(1+\epsilon)\ell_D(a)^\epsilon - (1+\gamma)\ell_D(a)^\gamma - u_D = 0, \tag{-1}$$

which is isomorphic to (9), with u_D instead of μ_0. Then $\ell_D(a) = \ell_0(a)$ if $u_D = \mu_0$. Note that $u_D > \mu_0$ is not possible (the decentralised allocation cannot Pareto dominate the optimal allocation by definition). Assume $u_D < \mu_0$. Then cities are more populated and fewer than at the optimal allocation. This implies that land developers located on empty land with $a \approx a_0$ can attract workers by offering them a utility higher than u_D and yet make profits. To see this, use (4) to solve for u_D and substitute into the definition for land developer profit to get $\pi(\ell_D(a), a) = \ell_D(a)[\gamma\ell_D(a)^\gamma - a\epsilon\ell_D(a)^\epsilon]$, which is positive by $\ell_D(a) > \ell_o(a)$. Thus $u_D = \mu_0$ and the equilibrium and optimal solutions coincide. To establish the 'only if' part of (i) and part (ii), assume that $\ell_D(a) = \ell_0(a)$. Then it must be that $u_D = \mu_0$. By (9) and (4), this implies $\partial u_D/\partial\ell(a) = 0$, which contradicts the assumption in (ii) that land developers behave strategically.