Agglomeration Externalities and Urban Growth Controls

Wouter Vermeulen

Agglomeration Externalities
and Urban Growth Controls

by

Wouter Vermeulen

CPB Netherlands Bureau for Economic Policy Analysis

VU University

Spatial Economics Research Centre (SERC)

Abstract

Should constraints on urban expansion be relaxed because of external agglomeration economies? In a system of heterogeneous cities, we demonstrate that second-best land use policy consists of a tax on city creation and a subsidy (tax) on urban development in cities in which the marginal-average productivity gap is above (below) average. However, the implementation of this policy requires coordination at the system level. A tax on city creation does not raise welfare if development taxes are set decentrally by competitive urban developers, nor does correction of these taxes raise welfare if a tax on city creation is unavailable. In the resulting constrained optimal allocation, urban development is subsidized in all cities. The quantitative significance of these findings is explored in an application of our model.

Abstract in Dutch

Wat betekent het bestaan van externe agglomeratievoordelen voor ruimtelijke-ordenings-beleid? We verkennen deze vraag in een gestileerd model voor een systeem van heterogene steden. Als instrumenten die direct op de externaliteit aangrijpen ontbreken, dan kan ruimtelijke ordening op twee manieren behulpzaam zijn. Op de eerste plaats is beleid dat het aantal steden beperkt welvaartsverhogend, omdat de steden die overblijven dan groter worden en de agglomeratievoordelen beter benutten. Omgekeerd leidt beleid dat de opzet van nieuwe steden faciliteert, zoals het (voormalige) groeikernenbeleid, tot een verlies aan agglomeratievoordelen elders in het systeem. Op de tweede plaats kan een combinatie van restricties en subsidies op nieuwbouw in bestaande steden ervoor zorgen dat de steden waar het externe agglomeratie-effect het grootst is, ook de meeste mensen binnen halen.

Vervolgens onderzoeken we de vraag in hoeverre competitie tussen steden leidt tot een maatschappelijk wenselijke uitkomst. Implementatie van het second-best beleid veronderstelt de mogelijkheid om het aantal steden te beperken en het vereist coördinatie van het ruimtelijke beleid van steden op nationaal niveau. Afzonderlijke steden houden immers geen rekening met het feit dat een versoepeling van hun beleid elders een verlies aan agglomeratievoordelen tot gevolg heeft. Zo leidt competitie tot een evenwicht waarin elke stad nieuwbouw subsidieert. Als het beperken van het aantal steden of de nationale coördinatie van ruimtelijk beleid onmogelijk blijkt, dan is deze uitkomst ook vanuit maatschappelijk oogpunt het beste alternatief.

Om een gevoel te krijgen voor het kwantitatieve belang van de mechanismen in het model, eindigt het paper met een numerieke toepassing.

1   Introduction

Economies of agglomeration are key to the success of cities. Their external nature leaves scope for policy intervention: markets would not by themselves take sufficient advantage of sharing, matching and learning in dense areas. In the absence of instruments that directly address the relevant failures, land use policy is an obvious second-best candidate, being an important determinant of the size and shape of cities. How exactly this policy should be geared to the presence of agglomeration externalities is the central topic of this paper.

In the setting of a system of heterogeneous cities, we consider two channels through which land use policy could further the exploitation a Marshallian externality in production. The first is a constraint on city creation, which raises the size of remaining cities in the system. The developer of a new city does not internalize the loss of external agglomeration economies in existing places from which its inhabitants are drawn, so that there are too many cities in a market equilibrium. This result runs counter to the conventional wisdom that cities are too large and that their creation should be supported by national governments, such as in the New Town programs that have been adopted in various European countries.

As a second channel, we analyse the adjustment of constraints and subsidies on residential development in existing cities. Provided that the number of cities is either fixed or optimized, the gain of relaxing land use policy in one place should be offset against foregone agglomeration benefits elsewhere. Hence, land use policy should be adjusted so as to direct households to places in which the external agglomeration benefit is comparably high. The potential gains depend on the degree of heterogeneity across cities and on the strength of nonlinearities in the relationship between productivity and city size.

Is it possible to decentralize the second-best land use policy to competitive urban developers who maximize the total differential rent in their own city? Such developers do not have the incentive to consider the costs imposed elsewhere when relaxing land use policy. Uncoordinated competition thus leads to an outcome in which development is subsidized in each city, until the opportunity costs of developed land offset the external productivity gains at the margin. We demonstrate that this outcome can only be improved upon if it is possible to constrain the creation of new cities and if land use policy in each city can be coordinated through planning at the national level. Limits on the number of cities do not enhance welfare if coordination of urban land use policies fails, nor can welfare be raised by correcting the

outcome of uncoordinated competition amongst urban developers if there is no constraint on city creation.

These results are of interest in view of the considerable heterogeneity in institutions across countries. For instance, land use policy is coordinated at the national level in countries like the UK and the Netherlands, but planning is mainly a local affair in the US. In particular, the process of city creation is typically coordinated by national governments in Europe, yet incorporation of municipalities is largely the product of bottom-up government in the US (Fischel, 2001). Our analysis thus suggests an explanation for why many US cities fail to charge the full cost of infrastructure provision, giving rise to what is commonly perceived as excessive urban sprawl.[1] In the absence of a coordinated national land use policy, these subsidies on urban development may be a rational way to enhance the exploitation of external agglomeration economies.

The theoretical part of this paper is augmented with a numerical application that illustrates the main mechanisms in our model and provides a crude sense of their quantitative significance. We contrast equilibrium outcomes under the second-best land use policy and under uncoordinated competition amongst urban developers, as well as under the outcome that obtains when land use policy ignores the existence of agglomeration externalities altogether. By evaluating welfare in these alternative outcomes, we are able to shed some light on the potential gains of tailoring land use policy to the presence of external agglomeration economies.

2   Background

Optimal land use policy in the presence of agglomeration externalities is rarely addressed in the urban economics literature. One notable exception is Rossi-Hansberg (2004), who considers zoning of land to business and residential use within an open city, when agglomeration externalities are localized as in the Lucas and Rossi-Hansberg (2002) model. He finds that zoning raises productivity by enhancing the spatial concentration of firms, which is insufficient in the equilibrium allocation. Our paper adds to these results by

---

[1] Brueckner (2000) argues that under common financing arrangements in the US, the infrastructure-related tax burden on new homeowners is typically less than the actual infrastructure costs they generate, because the cost of new sewers and schools is shared among all of the city's residents rather than charged directly to those who require the new infrastructure.

investigating how land use policy may be used to enhance the concentration of employment in a system of cities.

In the context of a system of two cities that differ in an exogenous amenity, Anas and Pines (2008a) address second-best urban growth boundaries in the presence of unpriced traffic congestion. They find that the size of the high-amenity city should be restricted and it should be enhanced it in the low-amenity city. Treating agglomeration as the mirror image of congestion and assuming that the externality increases with city size, this is consistent with our result that development should be subsidized (taxed) in cities with a high (low) marginal average productivity gap in a system in which the number of cities is fixed. Considering a system of an endogenous number of replicable cities, Anas and Pines (2008b) show that imposing an urban growth boundary in all cities is a second-best policy in the presence of unpriced traffic congestion. This mirrors our finding that urban development should be subsidized in all cities when a tax on city creation is unavailable.

Kanemoto (2011) considers the benefits from investment in urban transportation in a system of identical cities that is augmented with a rural sector, where external agglomeration economies derive from monopolistic competition with differentiated products. Keeping the total urban population and the number of cities fixed, he shows that external productivity gains from infrastructure investment in one city are fully offset by losses elsewhere in the system. This result closely resembles our finding that welfare cannot be raised by subsidizing development relative to its first-best level in a system of a fixed number of identical cities. Migration from the rural area mitigates the offsetting effect in his model, while the impact of city creation depends on whether cities are too large or too small. As in our analysis, results for an optimal number of cities are identical to the case of a fixed number of cities.

Glaeser and Gottlieb (2008) explore the role for national governments in strengthening the economies of particular localities or regions. They show in a spatial general equilibrium framework that it may be welfare enhancing to relocate people to areas where agglomeration economies are comparably high, which is consistent with our finding that development should be subsidized (taxed) when the gap between the marginal and the average product of labour is above (below) average in a system with a fixed number of cities. Welfare may be similarly raised by relocating people to areas where the elasticity of congestion with respect to population size is comparably low. The analytical framework is applied to both infrastructure investment and land use policy. Empirically, however, the authors find that nonlinearities in agglomeration and congestion externalities are difficult to identify, so that 'place-making policy' is deemed a risky affair.

None of the aforementioned papers considers constraints on city creation as an explicit policy instrument to enhance exploitation of external agglomeration economies. In contrast, Abdel-Rahman and Anas (2004) survey a literature contending that city formation is inefficiently slow. This literature assumes that there is no coordination of intercity migration, while scale economies make moving to a new place attractive only if a number of households migrate at the same time. Existing cities may thus become grossly overpopulated before new cities emerge, calling for government involvement in the setup of new cities. The New Towns that have been founded in several European countries may be regarded as a way to address this problem (Anas *et al*., 1998). However, Henderson and Venables (2009) show that the coordination problem may also be overcome by competitive forward-looking investors who supply the sunk capital, like public infrastructure, housing and business capital, that is indispensable for urban growth.

The few existing papers that perform an applied welfare analysis of urban land use use constraints ignore the existence of agglomeration externalities (see e.g. Cheshire and Sheppard, 2002, or Walsh, 2007). Whether this leads to an underestimation of costs depends, as our analysis shows, on how the cities under consideration compare to others in terms of the marginal-average productivity gap, as well as on the institutions that govern the setup of new cities. The application of our model sheds some light on the quantitative significance of the external productivity effects relative to the direct costs of land use policy.

3 Theory

We conceptualize land use policy in a system of cities as a set of possibly negative city-specific taxes on conversion of agricultural land to urban use and a tax on creating new cities.[2] This means that we abstract from direct regulations of urban density, for instance through minimum lot size zoning, and only consider their impeding effect on urban expansion. Conditional on this policy, competitive markets allocate land and all other goods and services. Land use policy is set by a benevolent planner at the system level or, equivalently, by an absentee owner of all land in the system. First-order conditions for optimal policy are derived and we consider the decentralization to planners or absentee landowners (urban developers) at

---

[2] Taxes may be interpreted as shadow prices of direct land use regulation, which is more often observed in practice. In the case of negative taxes, direct land use constraints would not be binding, but as discussed in the introduction, subsidies on urban development are a common phenomenon at least in the US.

the city level. Throughout this section, direct benefits of constraints on urban expansion are ignored, so in a first-best world the optimal development tax would be zero everywhere. This simplifying assumption allows us to focus on the implications of agglomeration externalities for land use policy in a second-best world.

3.1   Equilibrium outcome conditional on land use policy

In order to characterize the outcome of a general equilibrium at the system level, we first establish the outcome on land and labour markets in an open city conditional on land use policy. The system of cities consists of a set of open cities on which a population constraint is imposed. The number of cities in the system will initially be treated as given, but city creation will be endogenized in section 3.3.

Urban land market equilibrium in an open city

Households are assumed to be homogeneous and they derive utility $U(s, z, A_i)$ from the consumption of land s, a composite commodity z, and a city-specific exogenous amenity A_i, where i indexes the set of cities in the system. Households must reach a system-wide utility level u, which will be treated as exogenous. This condition implicitly defines $Z(u, s, A_i)$ as the consumption of z required for a household in city i with land consumption of s. Households receive a lump-sum transfer T and in the central business district (CBD) of the city, they exchange one unit of labour for a wage w_i. Respecting its budget constraint, the maximum amount a household can afford to pay for a unit land at distance r from the CBD equals:

$$\Psi_i(w_i, r, T) = \max_s \frac{w_i + T - tr - Z(u, s, A_i)}{s}, \tag{1}$$

where t denotes transport costs per unit of distance and the price of z is normalized to one. Expression (1) defines the *bid rent function*, which determines land rents in a spatial equilibrium. At each distance r, the first order condition associated with this maximization problem is:

$$-\frac{\partial Z(u, s, A_i)}{\partial s} = \frac{w_i + T - tr - Z(u, s, A_i)}{s}, \tag{2}$$

and the argument that solves it is denoted by $s_i(w_i,r,T)$. Expression (2) states the usual condition that the marginal rate of substituting the composite commodity for land should equal their rate of exchange at market prices.

Land is converted from agricultural to urban use if the bid rent exceeds annualized opportunity costs C_i (conversion costs and value of land in alternative use) and a development tax τ_i. Hence, at the urban fringe it must hold that:

$$\Psi_i(w_i,r,T) = C_i + \tau_i. \tag{3}$$

This condition implicitly defines the function $r_i^F(w_i,\tau_i,T)$, which characterizes the distance from the CBD to the fringe of the city.

Finally, for future reference, we derive the surface of the residential area in this city as:

$$S_i(w_i,\tau_i,T) = \int_0^{r_i^F(w_i,\tau_i,T)} L_i(r)\,dr, \tag{4}$$

where $L_i(r)$ denotes the length of the arc at distance r from the CBD that is available for residential use.

Urban labour market equilibrium in an open city

Since each household supplies one unit of labour, the urban labour supply equals the household density integrated over the residential sector of the city:

$$N_i(w_i,\tau_i,T) = \int_0^{r_i^F(w_i,\tau_i,T)} \frac{L_i(r)}{s_i(w_i,r,T)}\,dr. \tag{5}$$

Equation (5) implicitly defines the *inverse labour supply function* $w_i(N,\tau_i,T)$, which gives the wage that employers in the CBD should offer in order to attract N units of labour supply.

Labour is the single input in the production of a good that is traded on international markets for a price normalized to unity, employing a production technology of the following shape:

$$F(N) = g(N)N, \tag{6}$$

where $g(N)$ may be thought of as an increasing concave function of urban employment. The marginal product of labour is $g(N) + g'(N)N$. Individual firms are so small that they ignore the impact of wage setting on N and pay labour its average product $g(N)$. Hence, the inverse labour demand function is given by:

$$w(N) = g(N),\qquad\qquad(7)$$

which results in the conventional Marshallian externality that is compatible with competitive equilibrium.[3]

A stable urban labour market equilibrium is characterised by the conditions that N_i and w_i are chosen such that labour supply equals demand and that at this point, the cost of attracting labour rises more steeply than its average product. We denote the equilibrium labour supply by $N_i(\tau_i, T)$ and the equilibrium wage by $w_i(\tau_i, T)$.[4]

General equilibrium in a system of cities

A general equilibrium in open city i occurs when the urban land and labour markets clear simultaneously. Bid rents $\Psi_i(r, \tau_i, T)$, lot sizes $s_i(r, \tau_i, T)$, the urban fringe $r_i^F(\tau_i, T)$, and residential surface $S_i(\tau_i, T)$, as functions of r, τ_i and T only, are obtained by substituting $w_i(\tau_i, T)$ into expressions (1), (3) and (4) respectively.

The system of cities consists of a set of open cities indexed by i, $i = 1, ..., n$, which jointly accommodate an exogenous number of households $\bar{N}$. Note that we require all households to live in cities, there is no rural sector in our model. The implied *population constraint* reads:

$$\sum_{i=1}^{n} N_i(\tau_i, T) = \bar{N}.\qquad\qquad(8)$$

This condition implicitly determines the transfer level $T(\vec{\tau})$, where we denote $\vec{\tau} \equiv (\tau_1, ..., \tau_n)$. The *system general equilibrium* is defined as the set of open city general equilibrium allocations evaluated at $T = T(\vec{\tau})$.

It should be noted here that we consider *compensated equilibria* in which the population constraint is met through adjustment of T, while keeping u at a constant level.[5] The reason is that keeping households equally well off in each scenario allows us to evaluate different policies in welfare economic terms by employing a money metric that is readily

---

[3] This approach may be interpreted as a reduced form of more complex and possibly more realistic models of agglomeration externalities, see for instance discussions in Duranton and Puga (2004) or Kanemoto (2011).
[4] Existence and uniqueness properties depend on functional forms, so we will only verify them in the calibrated version of our model.
[5] Each compensated equilibrium corresponds to a *competitive equilibrium* in which T is held fixed and in which market interaction leads to an equilibrium utility level such that condition (8) is fulfilled.

available and easy to interpret: the *social surplus*, defined below as the difference between the value of the produce and all costs needed to attain it.

Social surplus

The costs of ensuring attainment of the target utility level for all households consist of transport costs, the purchase of composite commodities, the purchase of agricultural land and the cost of converting it. For city i, they may be written as:

$$C_i(\tau_i) = \int_0^{r_i^F(\tau_i, T(\vec{\tau}))} \left(\frac{tr + Z\left(u, s_r\left(r, \tau_i, T(\vec{\tau})\right), A_i\right)}{s_r\left(r, \tau_i, T(\vec{\tau})\right)} + C_i \right) L_i(r) \, dr. \tag{9}$$

Social surplus is obtained by subtracting these costs from $F\left(N_i(\vec{\tau})\right)$, the value of the produce in this city. Summing over all cities in the system, we rewrite it as:

$$SS(\vec{\tau}) = \sum_{i=1}^{n} \int_0^{r_i^F(\tau_i, T(\vec{\tau}))} \left(\Psi_i\left(r, \tau_i, T(\vec{\tau})\right) - C_i \right) L_i(r) \, dr - \bar{N} T(\vec{\tau}). \tag{10}$$

This may be seen by writing the total produce as the average product of labour times the household density, integrated over the entire residential area.

The exogenous utility level may be chosen in such a way that no transfers are required for attaining it if $\vec{\tau}$ maximizes social surplus. In that case, social surplus equals the *total differential land rent*, defined as the difference between the total land revenue and its opportunity costs. The problem of a planner who sets $\vec{\tau}$ so as to maximize social surplus is thus seen to coincide with the problem of a profit-maximizing absentee landlord who owns all the land in the system.

3.2 Optimal land use policy for a fixed number of cities

The first-best allocation in this model is well-known: wages should be subsidized with a city-specific amount $N_i(\vec{\tau}) g'\left(N_i(\vec{\tau})\right)$, so as to bridge the gap between marginal and average product. Since this subsidy internalizes all external effects, urban development and city creation should not be taxed or subsidized. The first-best allocation could be decentralized to urban developers who finance the wage subsidy exactly from the total differential land rent, following the Henry George theorem (see e.g., Abdel-Rahman and Anas, 2004). In practice, however, the nature of agglomeration externalities is highly complex and difficult to observe, which may explain why urban wage subsidies are rarely used as an instrument. Hence, we

8

assume that appropriate first-best instruments are unavailable in this paper, and we investigate the implications for optimal land use policy.[6]

The following lemma characterizes the effect on surplus of a marginal rise in the development tax.

Lemma 1: In a system general equilibrium for a fixed number of cities, the total cost of a marginal change in development tax in city i is given by:

$$\frac{\partial SS(\vec{\tau})}{\partial \tau_i} = \tau_i \frac{\partial S(\tau_i, T(\vec{\tau}))}{\partial \tau_i} + \Delta_i \frac{\partial N(\tau_i, T(\vec{\tau}))}{\partial \tau_i} - \Lambda \frac{\partial N(\tau_i, T(\vec{\tau}))}{\partial \tau_i}, \tag{11}$$

where $\Delta_i \equiv N_i(\vec{\tau}) g'(N_i(\vec{\tau}))$ is the gap between the marginal and the average product of labour and Λ is the shadow price of the population constraint (8). This shadow price may be decomposed as $\Lambda = \Lambda_1 + \Lambda_2$, where

$$\Lambda_1 \equiv \sum_{j=1}^{n} \tau_j \frac{\partial S_j(\tau_j, T(\vec{\tau}))}{\partial T} \bigg/ \sum_{j=1}^{n} \frac{\partial N_j(\tau_j, T(\vec{\tau}))}{\partial T} \tag{12}$$

is the gain in land surplus at the fringe of constrained cities from adding one additional household to the system and

$$\Lambda_2 \equiv \sum_{j=1}^{n} \Delta_j \frac{\partial N_j(\tau_j, T(\vec{\tau}))}{\partial T} \bigg/ \sum_{j=1}^{n} \frac{\partial N_j(\tau_j, T(\vec{\tau}))}{\partial T} \tag{13}$$

is the part of the gain in productivity that is not internalized in wages.

A proof of this lemma is deferred to the Appendix.

Expression (11) decomposes the total cost of a rise in τ_i into two city-specific or local effects and an effect at the system level. The first term reflects the loss of surplus at the fringe of city i. It equals the reduction in urban land multiplied by the gap between its value and the opportunity cost, which by condition (3) is identical to τ_i in equilibrium. The second term, the fall in population multiplied by the marginal-average productivity gap, reflects the loss in urban productivity insofar as it is not internalized in wages. However, the households that leave city i will locate in another city in the system and induce an expansion of urban land use and the workforce. Surplus will rise to the extent that the value of land in urban use exceeds opportunity costs and the marginal product of labour exceeds the average product in destination cities. The parameter Λ may thus be interpreted as a shadow price of constraint

---

[6] Of course, if the precise nature of agglomeration externalities is known and instruments that address them in a more direct fashion are available, these should be preferred over adjustment of land use policy.

(8), since it is equal to the gain in surplus from one additional household in the system. Note that in the absence of external agglomeration economies, this shadow price vanishes if development taxes are set at their first-best level (i.e., zero).

By evaluating the total derivative in (11) at $\vec{\tau} = \vec{0}$, we obtain the following proposition, which sheds light on how the presence of agglomeration externalities changes the desirable level of growth controls relative their first-best levels.

Proposition 1: In a system general equilibrium for a fixed number of cities, we have:

$$\left.\frac{\partial SS(\vec{\tau})}{\partial \tau_i}\right|_{\vec{\tau}=\vec{0}} > 0 \Leftrightarrow \Delta_i < \overline{\Delta}, \tag{14}$$

where

$$\overline{\Delta} \equiv \sum_{j=1}^{n} \Delta_j \frac{\partial N_j\left(\tau_j, T(\vec{\tau})\right)}{\partial T} \bigg/ \sum_{j=1}^{n} \frac{\partial N_j\left(\tau_j, T(\vec{\tau})\right)}{\partial T} \tag{15}$$

is a weighted average of Δ_j over all cities in the system.

This proposition states that social surplus rises by decreasing (increasing) development taxes relative to their first-best levels if the marginal-average productivity gap is larger (smaller) than a weighted average. The weights relate to the cost of drawing more households to a city. Intuitively, in the absence of first-best instruments to internalize agglomeration economies, land use constraints may thus be used to direct households from cities where the marginal-average productivity gap is relatively low towards cities where it is relatively high. By implication, agglomeration externalities do not affect the optimal development tax in a city in which this gap equals the weighted system average. Furthermore, the gain in surplus that may be brought about by tailoring land use policy to these externalities depends on variation in the marginal-average productivity gap across cities. In a system of identical cities, this instrument would not be helpful, since the gain from relaxing the development tax in one place would be exactly offset by the loss in productivity elsewhere.

Lemma 1 may also be used to characterize the first-order conditions for optimal land use policy, as stated in the following proposition.

Proposition 2: In a system general equilibrium for a fixed number of cities, first-order conditions for optimal land use policy are given by:

$$\tau_i \frac{\partial S(\tau_i, T(\vec{\tau}))}{\partial \tau_i} + \Delta_i \frac{\partial N(\tau_i, T(\vec{\tau}))}{\partial \tau_i} - \Lambda \frac{\partial N(\tau_i, T(\vec{\tau}))}{\partial \tau_i} = 0 \qquad (16)$$

for each city *i*.

It should be noted that, depending on the functional form of $g(N)$, the set of conditions (16) may not suffice to identify a unique global optimum.

3.3   City creation

Suppose that the set of *n* cities consists of *k* heterogeneous and *m* identical or *replicable* cities, for which all urban parameters are the same as for the *k*th city, with $n = k + m$. As long as we keep the number of cities fixed, this is a special case of the model that was previously discussed and all results we obtained continue to apply. Now we endogenize the number of replicable cities *m*. There is an annualized fixed cost *K* to creating a new city, which augments the conversion costs per unit of land. Replicates are set up by competitive urban developers, but the system planner determines development taxes and a tax μ on city creation.[7] The variable *m* will be treated as continuous and strictly positive.

Free entry of urban developers ensures the following zero-profit condition:

$$\int_0^{r_k^F(\tau_k, T)} \left(\Psi_k(r, \tau_k, T) - C_k \right) L_k(r) \, dr = K + \mu. \qquad (17)$$

For each τ_k and μ, this condition fixes a unique transfer level $T(\tau_k, \mu)$.[8] In this setting, the population constraint may be written as:

$$\sum_{i=1}^{k} N_i \left(\tau_i, T(\tau_k, \mu) \right) + m N_k \left(\tau_k, T(\tau_k, \mu) \right) = \bar{N}, \qquad (18)$$

which implicitly determines the number of replicates $m(\vec{\tau}, \mu)$. Intuitively, as long as $T > T(\tau_k, \mu)$, developers will find it profitable to build a new replicate city, since the total differential land rent it generates exceeds the fixed cost and tax. However, as the number of replicates increases, the transfer must decline in order for the population constraint to be satisfied. A similar reasoning applies for the case $T < T(\tau_k, \mu)$. Hence, we may define the

---

[7] In this context, competitive means that developers treat the transfer (or utility level) as given. This means that we rule out strategic growth controls, such as studied in Brueckner (1995) or Helsley and Strange (1995).
[8] This can be seen by noting that the total differential land rent is monotonously increasing in the transfer and that it is infinitely small (large) for a an infinitely large negative (positive) transfer.

system general equilibrium for an endogenous number of cities as a set of op city general equilibrium allocations for k heterogeneous cities and $m(\vec{\tau}, \mu)$ replicates, evaluated at $T(\tau_k, \mu)$. Social surplus equals:

$$SS(\vec{\tau}, \mu) = \sum_{i=1}^{k} \int_{0}^{r_i^F(\tau_i, T(\tau_k, \mu))} \left(\Psi_i(r, \tau_i, T(\tau_k, \mu)) - C_i\right) L_i(r) dr$$

$$+ m(\vec{\tau}, \mu) \int_{0}^{r_k^F(\tau_k, T(\tau_k, \mu))} \left(\Psi_k(r, \tau_k, T(\tau_k, \mu)) - C_k\right) L_k(r) dr \qquad (19)$$

$$- \left(k + m(\vec{\tau}, \mu)\right) K - \bar{N} T(\tau_k, \mu).$$

First-order conditions for optimal land use policy are characterised in the following proposition.

Proposition 3: In a system general equilibrium for an endogenous number of cities, first-order conditions for optimal land use policy are given by:

$$\mu = N_k(\tau_k, T(\tau_k, \mu)) \widetilde{\Lambda}, \qquad (20)$$

where $\widetilde{\Lambda}$ is the shadow price associated with population constraint (18), and:

$$\tau_i \frac{\partial S(\tau_i, T(\tau_k, \mu))}{\partial \tau_i} + \Delta_i \frac{\partial N(\tau_i, T(\tau_k, \mu))}{\partial \tau_i} - \widetilde{\Lambda} \frac{\partial N(\tau_i, T(\tau_k, \mu))}{\partial \tau_i} = 0 \qquad (21)$$

for each city i.

While a formal proof of these results is deferred to the Appendix, they may be well understood by revisiting the case of a fixed number of cities. First, observe that expression (21) is virtually identical to expression (16), the only difference being that the shadow price refers to the slightly reformulated population constraint. The conditions that characterize optimal development taxes are the same for a system with a fixed number of cities as for a system in which this number is chosen optimally. This is an immediate application of the Envelope Theorem. If the number of cities is chosen optimally, then entry or exit of cities should not contribute to social surplus at the margin. Hence, the marginal effect of a rise in τ_i on the number of cities in the system should leave surplus unaffected. This number may thus as well be treated as constant. By implication, Proposition 1 holds as well: development taxes should be adjusted relative to their first-best levels in order to direct households towards cities with a high marginal-average productivity gap.

Now consider the margin of city creation. Each new city reduces the population in the existing system with $N_k(\tau_k, T(\tau_k, \mu))$ households, thus reducing productivity and surplus at the fringe of existing cities. The reduction in surplus equals the shadow price of the population constraint times this number of households. It is these costs that the tax in expression (20) internalizes. Intuitively, by reducing the number of cities in the system, this tax improves exploitation of agglomeration externalities in existing cities. It should be noted that if the variation in the marginal-average productivity gap across cities is sufficiently small, the tax on city creation dominates development taxes as an instrument to internalize agglomeration externalities. In the limiting case of a system of identical cities, this tax is even a perfect substitute for the wage subsidy that would be implemented in a first-best world.

3.4 Decentralization to urban developers

What land use policy will result if development taxes are set by competitive urban developers? A competitive urban developer would maximize the total differential land rent in his city, while treating it as open by taking T (or utility) as given. The next proposition characterizes the first-order condition associated with this problem.

Proposition 4: In an open city general equilibrium, the first-order condition for the development tax that optimizes the total differential land rent in city i is given by:

$$\tau_i \frac{\partial S(\tau_i, T)}{\partial \tau_i} + \Delta_i \frac{\partial N(\tau_i, T)}{\partial \tau_i} = 0. \tag{22}$$

A proof is deferred to the Appendix. Since $\partial S(\tau_i, T)/\partial \tau_i$ is negative and the second term of (22) is negative as well, it can be seen that τ_i should be adjusted downwardly relative to its first-best level. This contrasts the adjustment as described in Proposition 1, where development taxes should increase in some cities and decrease in others, according to their marginal-average productivity gap. Intuitively, competitive urban developers would subsidize land until at the margin, the gap between opportunity costs and urban land value, multiplied by the change in land use induced by subsidizing it even further, offsets the part of productivity gains that is not internalized in wages.

The implication of Proposition 4 is that implementation of optimal land use policy in a system with an endogenous number of cities, as characterized by Proposition 3, requires

coordination at the system level. It cannot be decentralized to urban developers who maximize the total differential land rent in their specific city. In particular, by ignoring the term $-\widetilde{\Lambda}\,\partial N\left(\tau_i,T\left(\tau_k,\mu\right)\right)\big/\partial\tau_i$, which makes the difference between conditions (21) and (22), urban developers fail to take into account the cost that a development subsidy imposes on other cities in the system. Households that are attracted by the subsidy reduce productivity and surplus at the fringe of other cities. In an attempt to reap the benefits of agglomeration, developers engage in suboptimal 'land tax competition'.

We now turn to the case in which coordination of development taxes at the system level is impossible. Planning at the system level cannot correct or overrule the development taxes that are set by competitive urban developers according to condition (22). Is it still beneficial in this situation to restrict city creation?

Proposition 5: In a system general equilibrium for an endogenous number of cities, if development taxes are set according to condition (22) in all cities, then the shadow price of the population constraint vanishes and social surplus cannot be raised by imposing a tax on city creation.

A proof of this proposition is provided in the Appendix. The intuitive explanation is that by setting land use policy according to condition (22), urban developers second-best optimize the number of households in their city, so that there is no gain (loss) of a marginal rise (fall) in this number. Hence, in this equilibrium, the external effect of creating new cities vanishes as well and taxing it does not raise social surplus.

It turns out that the reverse holds as well. Welfare cannot be raised by correcting development taxes that urban developers would choose in existing cities, i.e. cities for which the fixed cost has already been incurred, if taxing city creation is either legally or practically unfeasible.

Proposition 6: In a system general equilibrium for an endogenous number of cities, if there is no tax on city creation, then for existing cities, the first-order condition for the development tax that optimizes social surplus (21) is identical to the first-order condition for the development tax that optimizes its total differential land rent (22).

We consider the development tax in city i unequal to k, since existing replicate cities can be included in the class of heterogeneous cities without loss of generality. Intuitively, a rise in τ_i does not affect other existing cities, since T only depends on τ_k and μ trough the zero-profit condition (17). The only impact beyond city i is an increase in the number of replicates. However, a marginal change in m does not affect social surplus if μ is set equal to zero: the rise in total differential land rent is exactly offset by fixed setup costs. Hence, in this equilibrium, a change in τ_i only affects surplus through a change in the total differential land rent in city i, so a competitive urban developer would internalize all its costs and benefits. This statement cannot be extended to the marginal city, since the developer could effectively regulate entry by intervening with its development tax. A more formal proof is again provided in the Appendix.

The overall implication of Propositions 4, 5 and 6 is that for the implementation of an optimal planning policy that satisfies the conditions specified in Proposition 3, a strong degree of coordination or planning at the system level is required. The creation of new cities has to be taxed or constrained and the owners or governments of specific cities have to be refrained from engaging in harmful land tax competition. If it is impossible to meet either of these conditions, a constrained optimum may be attained by allowing cities to impose development taxes that serve their own interests and by allowing for free entry of new urban developers.

Finally, while so far we have maintained the assumption that the number of households in the system is fixed, it is useful to briefly consider the implications of relaxing it. The opposite extreme would be the presence of a rural sector in which returns to scale are constant, so that there is a perfectly elastic supply of households to the system of cities. In this case, constraints on city creation raise the number of households in the rural sector, yet they leave the size of remaining cities unaffected. Hence, even if taxing city creation were legally and practically feasible, it would be ineffective in furthering the exploitation of external agglomeration economies. Since there would still be a cost in terms of foregone surplus in new cities, it is optimal not to impose a tax on city creation at all. However, in the absence of such a tax, it follows from Proposition 6 that allowing for decentralized setting of development taxes yields an optimal outcome. In other words, if there is a perfectly elastic supply of households from a rural sector, then the decentralized outcome is second-best.

4   Application to the Dutch Randstad Area

Figure 1 shows a map of the Dutch Randstad Area that highlights the seven cities to which we will apply our analysis.[9] The number of households and the surface of the residential area in each city are indicated in Table 1. While the distance between these cities is relatively small, commuting patterns suggest that they should not be regarded as one integrated spatial labour market (see e.g., RPB, 2006). The two largest cities of Amsterdam and Rotterdam, only 75 kilometres apart, are comparable in terms of population size and in terms of total land cover, while the third largest city of The Hague is more than half as large as Amsterdam. Hence, an urban primate hardly exists and the even size distribution of these cities defies Zipf's law.

If the origins of this remarkable urban constellation are historical, land use regulation has contributed significantly to maintaining or even reinforcing its character (see also Vermeulen and Van Ommeren, 2009). Figure 1 indicates boundaries of the so-called Green Heart, an area consisting of predominantly open space that Dutch planning policy has striven to preserve since the 1950s, and of Buffer zones that are especially designated to prevent the conglomeration of cities. The gap between the value of residential and agricultural land at these boundaries indicates that they do impose binding constraints on land use. The fourth column of Table 1 shows tentative estimates of this gap for each of the seven cities.[10] At the fringe of Amsterdam, the annual rent of a square meter of land in residential use appears to be almost 13€ higher than the sum of the agricultural land rent and annualized conversion costs, which corresponds to a regulatory tax rate on housing of almost 30%. However, the implied regulatory tax rate in Rotterdam is only about 7%, pointing to a substantial degree of heterogeneity across cities in the extent to which land use constraints are binding.[11]

---

[9] The delineation of these cities, as indicated in Figure 1, is based on the concept of *urban agglomerations* (in Dutch: 'grootstedelijke agglomeraties'), which has been proposed by Statistics Netherlands (CBS, 2005). These urban agglomerations consist of several municipalities and the (morphological) criterion for inclusion of a municipality is contiguity of build-up land.

[10] RIGO (2003) estimates the costs and benefits of residential development projects at various sites in the Randstad area. We have selected sites near the fringe of each city. For instance, for the city of Amsterdam we have used the site 'Bovenkerkerpolder' at its southern boundary. The report provides estimates of location-specific costs of acquiring and converting the land, as well as its value after development, so an indication of the shadow price may be obtained by subtracting the former from the latter. While the standard project used by RIGO contains a social housing share of 30%, we assume that land is only developed in its most profitable use. The regulatory tax rate is obtained by dividing the shadow price by the value of land inclusive of residential structures. Of course, these estimates are contingent on the specific choice of development sites, as well as a range of additional assumptions, so they should be interpreted with considerable caution.

[11] A comparison with regulatory tax rates reported for selected US metropolitan areas in Glaeser *et al.* (2005) indicates that land use controls in Amsterdam, The Hague and Leiden are at least as restrictive as in Boston, New York and Washington D.C., whereas only San Francisco and San Jose appear to conduct policies that are significantly more restrictive than the city of Amsterdam.

Preservation of open space is an important motivation for the land use constraints that Figure 1 illustrates. Focussing on land use policy as a second-best instrument in the presence of external agglomeration economies, our theoretical analysis has ignored any external value of open space. The purpose of this application is to illustrate the theoretical findings and provide a quantitative intuition. Hence, we simply assume that the use-constrained land has an existence value to society as open space that is exactly equal to the observed regulatory tax, i.e. development taxes are set at their first-best levels by assumption. In the baseline scenario, we similarly assume that there is no tax on city creation. We then use the calibrated model to show how land use policy should be adjusted relative to this first-best policy as a consequence of agglomeration externalities and we also compute the cost of failing to do so.

4.1 Choice of functional forms and parameters

Household utility is the product of a CES component in land and the composite commodity and the amenity level in the city of residence:

$$u(s, z, A) = A\left(\alpha z^\rho + \beta s^\rho\right)^{1/\rho},$$ (23)

where $\alpha + \beta = 1$. We assume an elasticity of substitution $\sigma = 1/(1-\rho)$ of 0.5, so households are less willing to substitute away from land than in the Cobb-Douglass case and constraints on the availability of land have a stronger impact on wellbeing. The parameters in this function are calibrated on data for the city of Amsterdam: A is normalized to unity for this city, u is chosen such that the number of households N_i predicted by the model equals the observation in Table 1 at $T = 0$ and the choice of α ensures that the surface of the residential area S_i is consistent with this table as well. Table 2 reports these and other system-wide parameters.

The urban production function is given as:

$$F(N) = CN^\kappa N,$$ (24)

where κ is the elasticity of average labour productivity with respect to urban scale – the number of households or jobs in the city. Rosenthal and Strange (2004) survey the early literature on this elasticity as indicating that doubling city size raises productivity by an amount that ranges from roughly 3 to 8%. However, these studies did not control for unobserved factors, such as the composition of the local workforce, that recent work has shown to result in downward bias (see in particular Combes *et al*., 2008). Therefore, we

somewhat conservatively choose $\kappa = 0.02$.[12] The constant C is chosen such that in the baseline equilibrium, the predicted wage in Amsterdam equals the average disposable household income in this city in 2002, as estimated from a household survey.

In each city i, the amount of developable land at distance r from the CBD is given by:

$$L_i(r) = 2\pi\omega_i r, \tag{25}$$

where ω_i is taken from Table 1 – it turns out that this parameter is quite homogeneous over cities. The annualized shadow price τ_i and the sum of the agricultural land rent and the annualized conversion costs C_i, given in the final two columns of Table 1, are also observed. The transport cost parameter t is calibrated such that the average percentage decline in land rents per unit of distance from the CBD in the model coincides with an estimated semi-elasticity of land rents with respect to distance in Amsterdam.[13] This land rent gradient may not only reflect transport costs to the centre, but also the distribution of jobs and amenities within the city.

The amenity level in cities other than Amsterdam is calibrated so that the number of households in the model matches the data in Table 1. Given our choice of production function, this assumption immediately yields wages in the other cities and given the choice of other parameters, we also obtain the surface of the residential area. The outcome is shown in Table 3. It turns out that there are substantial differences in the attractiveness of cities in other dimensions than the wage level. For instance, Amsterdam and Rotterdam are almost equally large and hence, in our model, wages are almost equal in these two cities. However, given the much lower shadow price in Rotterdam, land rents are calibrated to be substantially lower here. The lower willingness to pay for land must be driven by a consumption disadvantage, as captured by the amenity level in our model. The difference reported in Table 3 is equivalent to a 8.7% reduction in income.

The almost perfect correlation of actual with predicted surface of the residential area in Tables 1 and 3 respectively provides a crude validation of our calibration. If we divide by the number of households to obtain actual and predicted average lot size, the correlation is 0.86.

We assume that the smallest city in the system, Haarlem, is replicable. In the baseline scenario on which we calibrate our model, there is no tax on city creation. This means that the

---

[12] In an applied general equilibrium analysis of US county-level employment, Chatterjee (2006) also chooses a scale-elasticity of 0.02, following essentially the same line of reasoning. This study illustrates that such a seemingly small elasticity can still have a substantial impact on the spatial distribution of jobs.

[13] Land rents were obtained from housing transactions by estimating the local willingness to pay for land. The 'Dam square' was chosen as centre of Amsterdam. We found that land rents declined with 10% per kilometre of distance from this square.

fixed cost K of creating a new city must equal the surplus in Haarlem and that the number of replicates m is large enough to accommodate the total number of households in the Ranstad area.

The inverse urban labour demand and supply functions that were defined in equations (5) and (7) are shown in Figure 2 for the calibrated model. For each city, the inverse labour supply function cuts the inverse labour demand function twice. At its first intersection, the average product of labour rises faster than the cost of ensuring that the marginal household attains utility u, so this equilibrium cannot be stable. However, at the second intersection, the marginal cost curve rises faster than the average product curve, so that this is the only stable equilibrium.

Figure 2 also shows the marginal product of labour. The difference between this curve and the average product, defining Δ_i, is seen to rise with city size. Hence, Proposition 1 implies that the planner should relax land use constraints in the largest cities and make them more stringent in the smallest ones, inducing a more skewed size distribution. However, the variation in Δ_i across cities appears to be moderate, so that it seems unlikely that much can be gained from these policy adjustments in terms of surplus.

4.2   Comparative statics at the baseline equilibrium

For each city, Table 4 decomposes the total impact on surplus from a marginal rise in the development tax, as derived in Lemma 1, relative to the baseline equilibrium. The first column reports the reduction in surplus at the urban fringe, which by assumption would be exactly offset by the existence value of the land that is preserved from development. By implication it must hold that $\Lambda_1 = 0$, since the same offsetting effect is present at the fringe of the cities to which the households from city i are diverted. Hence, the total impact on social surplus shown in the fourth column results from $\Delta_i \, \partial N_i / \partial \tau_i$ and $-\Lambda_2 \, \partial N_i / \partial \tau_i$, the loss and gains in uninternalized productivity in city i and elsewhere in the system that are reported in columns two and three.

Bearing in mind the positive relationship between city size and the marginal-average productivity gap in Figure 2, the signs of $\partial SS / \partial \tau_i$ for the different cities confirm Proposition 1: social surplus rises from a decline in the development tax in relatively large cities and a rise in the development tax in relatively small cities. However, in all cities, the magnitude of the net effect is small relative to its constituent terms. Gains and losses in productivity appear to

cancel out by and large. This is an implication of the limited variation in the marginal-average productivity gap, resulting from the constant and relatively low scale elasticity.

The comparative static results in Table 4 also shed light on the relative importance of foregone agglomeration benefits, compared to the magnitude of $\tau_i \, \partial S_i / \partial \tau_i$. The foregone surplus on constrained land exceeds the external productivity loss in all cities and in most of them, the difference is substantial. Hence, a misperception of the external value of open space, arbitrarely assumed to offset these costs in our analysis, could easily dominate the social cost of land use constraints.

At the baseline equilibrium, Λ_2 equals 467 euro, or 1.9% of the average disposable household income in Amsterdam. This external effect from reducing the number of households in the system may be regarded as substantial. Although Table 4 indicates that development taxes are of little help in addressing it, a significant scope for policies that constrain city creation is implied.

4.3   Counterfactual analysis

Table 5 shows how land use policy should be adjusted relative to the first-best as a consequence of external agglomeration benefits. The third column reports the second-best policy that is characterised in Proposition 3. Columns one and four contrast this policy to the baseline and the first-best policy, in which labour is paid its marginal product. The constrained optimal policy, in which development taxes are set by competitive developers according to Proposition 4 and in which there is no tax on city creation, is characterised in the second column. The table also documents the equilibrium allocations of households over cities, for which Figure 3 provides a graphical illustration, the total usage of urban land and social surplus.

As expected, the second-best land use policy consists of a substantial tax on city creation, being almost as large as the fixed setup cost, and a minor adjustment of development taxes in line with Proposition 1. This policy leads to a major reduction in the number of replicates, thus boosting the number of households in the remaining cities: the population size in these cities rises with 46% on average. Thus, although there is hardly any need to adjust development taxes, direct land use constraints should be substantially adjusted relative to the baseline scenario, in order to allow for these additional households. The resulting allocation is nearly identical to the distribution of households in the first-best equilibrium. In the

constrained optimum, development taxes are substantially lower than their first-best levels in all cities, leading to an average population increase of 35%. Since growth of existing cities reduces the profitability of city creation, the number of replicates is significantly reduced relative to the baseline scenario, though not to the same extent as in the second-best outcome.

The total amount of land in urban use reduces substantially by moving from the baseline to the first or second-best equilibrium. The reason is that by concentrating the population into a smaller number of larger cities, these policies will push up land prices and hence induce a substitution towards the consumption of the composite commodity. This effect is also present in the constrained optimal allocation, but it is more than offset by the distorting impact of land subsidies on city size. Hence, in this equilibrium, average density goes down rather than up. These results are also noteworthy in the context of perceived excessive sprawl in US cities, as discussed in the introduction.

Table 5 finally reports the social surplus generated in each equilibrium. The gain that can be realized by moving from the baseline scenario to the first-best allocation amounts to 233 million Euro annually, or 0.31% of the value of the total produce. Unsurprisingly, the second-best land use policy that leads to an almost identical allocation, also yields virtually the same surplus gain. However, 80% of this surplus gain can still be realised in the constrained optimum, in which either the tax on city creation is unavailable or coordination of development taxes at the system level is impossible.

5   Conclusions and discussion

We have explored the properties of land use policy as a second-best instrument in the presence of urban agglomeration externalities. It has been shown that a tax on city creation enhances exploitation of agglomeration benefits by raising the average size of remaining cities, while city-specific taxes or subsidies on development further improve the allocation by directing people towards those cities where the externality is largest. Land tax competition by urban developers may undermine the second-best allocation. Taxing city creation is undesirable if this kind of competition cannot be curbed through coordination at the system level, whereas decentralized setting of development taxes yields a constrained optimal outcome if the tax on city creation is unavailable.

Compared to the first-best wage subsidy, which would address the marginal-average productivity gap in each city specifically, the tax on city creation is crude in that it ignores the variation in this gap across cities, yet a development tax or subsidy is crude in that it distorts the size of cities in terms of total land use. Hence, constraining city creation turns out to be the dominant second-best instrument if variation in the marginal-average productivity gap is limited. This insight appears to be new in the urban economics literature, yet three qualifying remarks are in order. In the first place, variation in the scale elasticity across cities may be larger in reality than assumed in our calibration, which would shift the balance between the instruments. In the second place, our model ignores the rural sector. Constraints on city creation become ineffective in furthering the exploitation of external agglomeration economies if they merely raise the number of households in the countryside, so that the constrained optimal policy becomes the second-best. Finally, a corrective policy in opposite direction may be called for if coordination problems hamper the process of city formation. Put differently, the need for governments to support the process of city creation is mitigated or may even be reversed in the presence of external agglomeration economies.

While the focus of this paper has been on agglomeration externalities in production, the same analysis could equally be applied to agglomeration externalities in consumption, which are believed to be of increasing importance for urban success (Glaeser *et al*., 2001). Local goods and services that are produced with increasing returns to scale will be underprovided, if they are priced at average rather than marginal costs. For instance, the provision of public transportation may involve substantial fixed investments that are partly recouped from user fees. Any fixed costs faced by theatres or restaurants will also likely contribute to the price of tickets and menus. In bigger cities such fixed costs are spread over a larger number of users, yet the marginal resident does not internalize the benefit this entails for other households in the city. Land use policies should be adjusted to such external agglomeration economies in consumption in the same way as in the case of an agglomeration externality in production, thus enhancing the quality and variety of local consumer goods and services.[14]

Urban growth may also induce negative externalities like traffic congestion, pollution and crime. The urban production function in our theoretical analysis could be altered to reflect these externalities, without fundamentally changing the analysis. What would change is the

---

[14] See for instance Abdel-Rahman and Anas (2004) for a discussion of the isomorphisms between urban models with a Marsshallian externality in production, models with a local public good and models with a demand for product variety.

comparative statics: city creation should be subsidized and if such were impossible, urban development should be taxed if negative externalities were predominant. Whether the positive or negative externalities of cities dominate is in the end an empirical matter.

Appendix: Proofs of propositions

Proof of Lemma 1: In order to compute the total derivative of $SS(\vec{\tau})$ with regard to τ_i, it is convenient to consider:

$$SS(\vec{\tau},T) = \sum_{i=1}^{n} \int_{0}^{r_i^{F}(\tau_i,T)} \left(\Psi_i(r,\tau_i,T) - C_i \right) L_i(r) dr - \bar{N}T, \tag{A1}$$

where $SS(\vec{\tau}) = SS(\vec{\tau}, T(\vec{\tau}))$. Differentiation yields:

$$\frac{\partial SS(\vec{\tau})}{\partial \tau_i} = \frac{\partial SS(\vec{\tau},T)}{\partial \tau_i} + \frac{\partial SS(\vec{\tau},T)}{\partial T} \frac{\partial T(\vec{\tau})}{\partial \tau_i}. \tag{A2}$$

This expression decomposes the impact of a change in τ_i into a city-specific and a system-level effect. We proceed by determining all of its components. First, it follows from Leibniz's rule that:

$$\frac{\partial SS(\vec{\tau},T)}{\partial \tau_i} = \tau_i \frac{\partial S_i(\tau_i,T)}{\partial \tau_i} + \int_{0}^{r_i^{F}(\tau_i,T)} \frac{\partial \Psi_i(r,\tau_i,T)}{\partial \tau_i} L_i(r) dr, \tag{A3}$$

where the first term has been simplified using the boundary condition (3) and the definition for surface of the residential area (4). The second term may be simplified in a number of steps:

$$\int_{r=0}^{r_i^{F}(\tau_i,T)} \frac{\partial \Psi_i(r,\tau_i,T)}{\partial \tau_i} L_i(r) dr = \frac{\partial w_i(\tau_i,T)}{\partial \tau_i} \int_{r=0}^{r_i^{F}(\tau_i,T)} \frac{\partial \Psi_i(w_i(\tau_i,T),r,T)}{\partial w} L_i(r) dr$$

$$= \frac{\partial g(N_i(\tau_i,T))}{\partial N} \frac{\partial N_i(\tau_i,T)}{\partial \tau_i} N_i(\tau_i,T) \tag{A4}$$

$$= \Delta_i \frac{\partial N_i(\tau_i,T)}{\partial \tau_i}.$$

The first step uses the fact that a change in τ_i affects the bid rent only through its impact on wages. The second step uses the derivative of the bid rent to the wage from (1) and the condition that labour gets paid its average product. The third step uses the equality $F'(N) = g(N) + g'(N)N$ and the definition of Δ_i.

In order to determine the second term of expression (A2), we note that implicit differentiation of the population constraint (8) yields:

$$\frac{\partial T(\vec{\tau})}{\partial \tau_i} = -\frac{\partial N_i(\tau_i,T)}{\partial \tau_i} \bigg/ \sum_{j=1}^{n} \frac{\partial N_j(\tau_j,T)}{\partial T} . \tag{A5}$$

Furthermore, we again apply Leibniz's rule in order to derive:

$$\frac{\partial SS(\vec{\tau},T)}{\partial T} = \sum_{j=1}^{n} \left(\tau_j \frac{\partial S_j(\tau_j,T)}{\partial T} + \int_{0}^{r_j^F(\tau_j,T)} \frac{\partial \Psi_j(r,\tau_j,T)}{\partial T} L_i(r) dr \right) - \bar{N} , \tag{A6}$$

where the first term has been simplified in the same way as in (A3). Noting that:

$$\frac{\partial \Psi_j(r,\tau_j,T)}{\partial T} = \frac{\partial \Psi_j(w_j(\tau_j,T),r,T)}{\partial T} + \frac{\partial \Psi_j(w_j(\tau_j,T),r,T)}{\partial w_j} \frac{\partial w_j(\tau_j,T)}{\partial T} , \tag{A7}$$

we may simplify the second term of expression (A6) as:

$$\int_{r=0}^{r_j^F(\tau_j,T)} \frac{\partial \Psi_j(r,\tau_j,T)}{\partial T} L_j(r) dr = N_j(\tau_j,T) + \Delta_j \frac{\partial N_j(\tau_j,T)}{\partial T} , \tag{A8}$$

using essentially the same steps as in (A4). Substitution yields:

$$\frac{\partial SS(\vec{\tau},T)}{\partial T} = \sum_{j=1}^{n} \left(\tau_j \frac{\partial S_j(\tau_j,T)}{\partial T} + \Delta_j \frac{\partial N_j(\tau_j,T)}{\partial T} \right) . \tag{A9}$$

Collecting all components and substituting them back into (A2), while using the definition of Λ, we obtain expression (11), which was to be demonstrated.

Proof of Proposition 3: We establish this proposition by computing derivatives of social surplus with regard to μ and τ_i and equating them to zero. First, social surplus is redefined as:

$$SS(\vec{\tau},\mu,T) = \sum_{i=1}^{k} \int_{0}^{r_i^F(\tau_i,T)} \left(\Psi_i(r,\tau_i,T) - C_i \right) L_i(r) dr$$

$$+ m(\vec{\tau},\mu) \int_{0}^{r_k^F(\tau_k,T)} \left(\Psi_k(r,\tau_k,T) - C_k \right) L_k(r) dr \tag{A10}$$

$$- \left(k + m(\vec{\tau},\mu) \right) K - \bar{N}T ,$$

where $SS(\vec{\tau},\mu) = SS(\vec{\tau},\mu,T(\tau_k,\mu))$. This allows us to write:

$$\frac{\partial SS(\vec{\tau},\mu)}{\partial \mu} = \frac{\partial SS(\vec{\tau},\mu,T)}{\partial \mu} + \frac{\partial SS(\vec{\tau},\mu,T)}{\partial T} \frac{\partial T(\tau_k,\mu)}{\partial \mu} . \tag{A11}$$

The proof will proceed by deriving all components of this expression. First, by making use of the zero profit condition (17), we note that:

$$\frac{\partial SS(\vec{\tau},\mu,T)}{\partial \mu} = \mu \frac{\partial m(\vec{\tau},\mu)}{\partial \mu}. \tag{A12}$$

Implicit differentiation of the population constraint (18) yields:

$$\frac{\partial T(\tau_k,\mu)}{\partial \mu} = -\frac{\partial m(\vec{\tau},\mu)}{\partial \mu} \frac{N_k(\tau_k,T(\tau_k,\mu))}{\sum_{j=1}^{k} \partial N_j(\tau_j,T(\tau_k,\mu))/\partial T + m(\vec{\tau},\mu)\partial N_k(\tau_k,T(\tau_k,\mu))/\partial T}. \tag{A13}$$

Following essentially the same derivation as in Lemma 1, we obtain:

$$\frac{\partial SS(\vec{\tau},\mu,T)}{\partial T} = \sum_{j=1}^{k}\left(\tau_j \frac{\partial S_j(\tau_j,T)}{\partial T} + \Delta_j \frac{\partial N_j(\tau_j,T)}{\partial T}\right) + \\ + m(\vec{\tau},\mu)\left(\tau_k \frac{\partial S_k(\tau_k,T)}{\partial T} + \Delta_k \frac{\partial N_k(\tau_k,T)}{\partial T}\right). \tag{A14}$$

Substitution into (A11) yields:

$$\frac{\partial SS(\vec{\tau},\mu)}{\partial \mu} = \left(\mu - N_k(\tau_k,T(\tau_k,\mu))\widetilde{\Lambda}\right)\frac{\partial m(\vec{\tau},\mu)}{\partial \mu}, \tag{A15}$$

where $\widetilde{\Lambda}$ is the shadow price of the population constraint (18), given by:

$$\widetilde{\Lambda} \equiv \frac{\sum_{j=1}^{k}\left(\tau_j \dfrac{\partial S_j(\tau_j,T)}{\partial T} + \Delta_j \dfrac{\partial N_j(\tau_j,T)}{\partial T}\right) + m(\vec{\tau},\mu)\left(\tau_k \dfrac{\partial S_k(\tau_k,T)}{\partial T} + \Delta_k \dfrac{\partial N_k(\tau_k,T)}{\partial T}\right)}{\sum_{j=1}^{k} \partial N_j(\tau_j,T(\tau_k,\mu))/\partial T + m(\vec{\tau},\mu)\partial N_k(\tau_k,T(\tau_k,\mu))/\partial T}. \tag{A16}$$

Hence, the first-order condition for optimal taxation of city creation reads:

$$\mu = N_k(\tau_k,T(\tau_k,\mu))\widetilde{\Lambda}, \tag{A17}$$

or city creation should be taxed with the (shadow) price of taking away one household from the existing system of cities, multiplied by the size of the marginal city.

Now consider the derivative of social surplus with regard to τ_i, for i unequal to k. Since $T(\tau_k,\mu)$ is independent of τ_i, we may write:

$$\frac{\partial SS(\vec{\tau},\mu)}{\partial \tau_i} = \frac{\partial SS(\vec{\tau},\mu,T)}{\partial \tau_i}, \tag{A18}$$

the right-hand side of which equals:

$$\frac{\partial SS(\vec{\tau},\mu,T)}{\partial \tau_i} = \frac{\partial}{\partial \tau_i} \int_0^{r_i^F(\tau_i,T)} \left(\Psi_i(r,\tau_i,T) - C_i\right)L_i(r)\,dr + \mu \frac{\partial m(\vec{\tau},\mu)}{\partial \tau_i}. \tag{A19}$$

The first term of this expression is already familiar from the proof of Lemma 1. The second term is obtained by implicit differentiation of (18), yielding:

$$\frac{\partial m(\vec{\tau}, \mu)}{\partial \tau_i} = -\frac{\partial N_i\left(\tau_i, T(\tau_k, \mu)\right)/\partial \tau_i}{N_k\left(\tau_k, T(\tau_k, \mu)\right)} \,. \tag{A20}$$

Substituting into (A18), while making use of (A17), we obtain:

$$\frac{\partial SS(\vec{\tau}, \mu)}{\partial \tau_i} = \tau_i \frac{\partial S\left(\tau_i, T(\tau_k, \mu)\right)}{\partial \tau_i} + \Delta_i \frac{\partial N\left(\tau_i, T(\tau_k, \mu)\right)}{\partial \tau_i} - \widetilde{\Lambda} \frac{\partial N\left(\tau_i, T(\tau_k, \mu)\right)}{\partial \tau_i} \,, \tag{A21}$$

yielding the familiar first-order condition:

$$\tau_i \frac{\partial S\left(\tau_i, T(\tau_k, \mu)\right)}{\partial \tau_i} + \Delta_i \frac{\partial N\left(\tau_i, T(\tau_k, \mu)\right)}{\partial \tau_i} - \widetilde{\Lambda} \frac{\partial N\left(\tau_i, T(\tau_k, \mu)\right)}{\partial \tau_i} = 0 \,. \tag{A22}$$

Differentiating social surplus with regard to τ_k yields the impact on surplus of a simultaneous marginal increase of the development tax in all replicate cities. We have:

$$\frac{\partial SS(\vec{\tau}, \mu)}{\partial \tau_k} = \frac{\partial SS(\vec{\tau}, \mu, T)}{\partial \tau_k} + \frac{\partial SS(\vec{\tau}, \mu, T)}{\partial T} \frac{\partial T(\tau_k, \mu)}{\partial \tau_k} \,. \tag{A23}$$

The first term of this expression is given by:

$$\frac{\partial SS(\vec{\tau}, \mu, T)}{\partial \tau_k} = \left(1 + m(\vec{\tau}, \mu)\right) \frac{\partial}{\partial \tau_k} \int_0^{r_k^F(\tau_k, T)} \left(\Psi_k\left(r, \tau_k, T\right) - C_k\right) L_k(r) \, dr + \mu \frac{\partial m(\vec{\tau}, \mu)}{\partial \tau_k} \,. \tag{A24}$$

Again, by implicit differentiation of (18) we obtain:

$$\frac{\partial m(\vec{\tau}, \mu)}{\partial \tau_k} = -\left(\sum_{i=1}^{k} \frac{\partial N_i\left(\tau_i, T(\tau_k, \mu)\right)}{\partial T} + m(\vec{\tau}, \mu) \frac{\partial N_k\left(\tau_k, T(\tau_k, \mu)\right)}{\partial T}\right) \frac{\partial T(\tau_k, \mu)/\partial \tau_k}{N_k\left(\tau_k, T(\tau_k, \mu)\right)} \tag{A25}$$
$$- \left(1 + m(\vec{\tau}, \mu)\right) \frac{\partial N_k\left(\tau_k, T(\tau_k, \mu)\right)/\partial \tau_k}{N_k\left(\tau_k, T(\tau_k, \mu)\right)} \,.$$

Hence, using (A14) and (A17), the second term of expression (A24) equals:

$$\mu \frac{\partial m(\vec{\tau}, \mu)}{\partial \tau_k} = -\frac{\partial SS(\vec{\tau}, \mu, T)}{\partial T} \frac{\partial T(\tau_k, \mu)}{\partial \tau_k} - \left(1 + m(\vec{\tau}, \mu)\right) \widetilde{\Lambda} \frac{\partial N\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} \,. \tag{A26}$$

Since the first term of this expression cancels out against the third term of (A23), substitution yields:

$$\frac{\partial SS(\vec{\tau}, \mu)}{\partial \tau_k} = \left(1 + m(\vec{\tau}, \mu)\right) \left(\tau_k \frac{\partial S\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} + \Delta_k \frac{\partial N\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} - \widetilde{\Lambda} \frac{\partial N\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} \right) \,.$$

$$\tag{A27}$$

Hence, we obtain the first order condition:

$$\tau_k \frac{\partial S\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} + \Delta_k \frac{\partial N\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} - \widetilde{\Lambda} \frac{\partial N\left(\tau_k, T(\tau_k, \mu)\right)}{\partial \tau_k} = 0 \,, \tag{A28}$$

which completes the proof.

Proof of Proposition 4: We write the total differential land rent in city i as:

$$TDR_i\left(\tau_i,T\right)=\int_0^{r_i^F\left(\tau_i,T\right)}\left(\Psi_i\left(r,\tau_i,T\right)-C_i\right)L_i\left(r\right)dr. \tag{A29}$$

By applying Leibniz's rule and the same steps as in (A3) and (A4), we obtain:

$$\frac{\partial TDR\left(\tau_i,T\right)}{\partial\tau_i}=\tau_i\frac{\partial S\left(\tau_i,T\right)}{\partial\tau_i}+\Delta_i\frac{\partial N\left(\tau_i,T\right)}{\partial\tau_i}. \tag{A30}$$

The proof is competed by setting this derivative equal to zero.

Proof of Proposition 5: We start by observing that in the derivation of the optimal tax on city creation in the proof of Proposition 3, no specific assumptions about $\bar{\tau}$ have been made, so development taxes may be chosen such that they satisfy (22). Substitution of this condition into expression (A16) yields $\widetilde{\Lambda}=0$. Substitution into (A17) yields $\mu=0$.

Proof of Proposition 6: For city $i\neq k$, condition (A19) reduces to condition (A30) if we substitute $\mu=0$. Hence, the first-order condition that obtains by setting (A18) to zero must be the same as the first-order condition that obtains by setting (A30) to zero. From inspection of (A23), it can be seen that the same argument does not work for city k, since a marginal change in τ_k also affects $T\left(\tau_k,\mu\right)$.

References

Abdel-Rahman H. M. and Anas, A. (2004). Theories of systems of cities. In Henderson V. J. and Thisse, J. F. (eds.), *Handbook of Regional and Urban Economics*, *Volume 4*. Amsterdam: North-Holland/Elsevier.

Anas, A., Arnott, R. and Small, K. A. (1998). Urban spatial structure. *Journal of Economic Literature*, **36**, 1426–64.

Anas, A. and Pines, D. (2008a). Anti-sprawl policies in a system of congested cities. *Regional Science and Urban Economics*, **38**, 408–23.

Anas, A. and Pines, D. (2008b). Public goods and congestion in a system of cities: how do fiscal and zoning policies improve efficiency? Paper presented at the 3rd Kuhmo-Nectar Conference and Summer School, VU University, Amsterdam, July 2008.

Brueckner, J. K. (1995). Strategic control of growth in a system of cities. *Journal of Public Economics*, **57**, 393–416.

Brueckner, J. K. (2000). Urban Sprawl: Diagnosis and Remedies. *International Regional Science Review*, **23**, 160–71.

CBS (2005). *Grootstedelijke agglomeraties en stadsgewesten afgebakend.* Voorburg / Heerlen: Centraal Bureau voor de Statistiek.

Chatterjee, S. (2006). A Quantitative Assessment of the Role of Agglomeration Economies in the Spatial Concentration of U.S. Employment. Federal Reserve Bank of Philadelphia Working Paper no. 06 – 20.

Cheshire, P. C. and Sheppard, S. (2002). Welfare Economics of Land Use Regulation. *Journal of Urban Economics*, **52**, 242–69.

Combes P.-P., Duranton G. and Gobillon L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, **63**, 723–42.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson V. J. and Thisse, J. F. (eds.), *Handbook of Regional and Urban Economics, Volume 4*. Amsterdam: North-Holland/Elsevier.

Fischel, W. A. (2001). *The homevoter hypothesis*. Cambridge, MA: Harvard University Press.

Glaeser, E. L., Kolko, J. and Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, **1**, 27–50.

Glaeser, E. L., Gyourko J. and Saks, R. E. (2005). Why is Manhattan so Expensive? Regulation and the Rise in Housing Prices. *Journal of Law and Economics*, **48**, 331–69.

Helsley, R. W. and Strange W. C. (1995). Strategic growth controls. *Regional Science and Urban Economics*, **25**, 435–60.

Henderson, J. V. and Venables, A. J. (2009). The dynamics of city formation. *Review of Economic Dynamics*, **12**, 233–54.

Kanemoto, Y. (2011). Second-Best Cost–Benefit Analysis with a Microfoundation of Urban Agglomeration. Tokyo, Japan: GRIPS Discussion Paper 11-03.

Lucas, R. E. Jr. and Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, **70**, 1445–76.

RIGO (2003). *Investeringskosten Woningbouwlocaties Deltametropool; Deelstudie voor kosten en baten verstedelijking Deltametropool.* Amsterdam: RIGO Research en Advies BV.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In Henderson V. J. and Thisse, J. F. (eds.), *Handbook of Regional and Urban Economics*, *Volume 4*. Amsterdam: North-Holland/Elsevier.

Rossi–Hansberg, E. (2004). Optimal Urban Land Use and Zoning. *Review of Economic Dynamics*, **7**, 69–106.

RPB (2006). *Vele steden maken nog geen Randstad.* Rotterdam: NAi uitgevers.

Vermeulen, W. and Van Ommeren, J. N. (2009). Does land use planning shape regional economies? A simultaneous analysis of housing supply, internal migration and local employment growth in the Netherlands. *Journal of Housing Economics*, **18**, 294–310.

VROM, LNV, V&W and EZ (2004). *Nota ruimte. Ruimte voor ontwikkeling*. Ministerie van Ministerie van Volkshuisvesting, Ruimtelijke ordening en Milieu, Ministerie van Landbouw, Natuurbeheer en Visserij, Ministerie van Verkeer en Waterstaat and Ministerie van Economische Zaken, SdU, Den Haag.

Walsh, R. (2007). Endogenous open space amenities in a locational equilibrium. *Journal of Urban Economics*, **61**, 319–44.

Table 1: City-specific data

City	N_i (households)	S_i (hectares)	ω_i (%)	τ_i (€ / m^2)	C_i (€ / m^2)
Amsterdam	527086	7964	18	12.71	4.06
Rotterdam	476514	8553	18	2.36	3.76
The Hague	305393	5355	24	8.20	3.96
Utrecht	191066	3837	22	5.41	4.12
Leiden	114024	2467	25	9.11	4.19
Dordrecht	102280	2751	15	4.20	4.23
Haarlem	89029	2205	26	5.43	3.89

Note: The number of households, the surface of the residential area and the share of the total surface that is in residential use are obtained from Statistics Netherlands. The population data refer to 2002 and the land use data refer to 2003. The annualized shadow price of land use constraints and sum of opportunity and conversion costs were derived by the author based on data in RIGO (2003), see the main text for details.

Table 2: System-wide parameters

	Description of parameter	Value
α	preference parameter composite commodity	0.9977
β	preference parameter land	0.0023
σ	elasticity of substitution	0.5
u	utility level	13034
T	transfer	0
C	constant in production function	18441
κ	scale elasticity	0.02
t	transport cost (€ / m)	0.450
K	annualized fixed cost (million €)	68.50
N	total number of households in the system	3197340
m	number of replicates	15.63
	discount rate	0.05

Note: Transport costs and the number of households in the Randstad area are based on actual data, the other parameters have been calibrated as explained in the main text. Land rents on which t is estimated were derived from housing transactions data of the Dutch Realtors Association (NVM) and household data are from Statistics Netherlands. We use the same delineation of the Randstad area as in Vermeulen and Van Ommeren (2009).

Table 3: The baseline equilibrium

City	w_i (€)	A_i	S_i (hectare)
Amsterdam	24000	1	7964
Rotterdam	23952	0.913	10429
The Hague	23739	0.901	6021
Utrecht	23518	0.861	4373
Leiden	23276	0.866	2411
Dordrecht	23226	0.848	2516
Haarlem	23161	0.824	2274

Note: The wage in Amsterdam equals the average disposable household income in this city as observed from the 'Woning Behoefte Onderzoek' survey in 2002. The Amsterdam amenity level is normalized to one. The surface of the residential area in this city corresponds to its observed value, as reported in Table 1. See the main tax for details on how outcomes for the other cities were obtained.

Table 4: Comparative statics at baseline equilibrium (in M€)

| City	$\tau_i \dfrac{\partial S_i}{\partial \tau_i}$	$\Delta_i \dfrac{\partial N_i}{\partial \tau_i}$	$-\Lambda_2 \dfrac{\partial N_i}{\partial \tau_i}$	$\dfrac{\partial SS}{\partial \tau_i}$
Amsterdam	-82.65	-17.79	17.31	-0.49
Rotterdam	-28.50	-19.91	19.40	-0.51
The Hague	-72.90	-19.07	18.75	-0.32
Utrecht	-47.09	-16.04	15.92	-0.12
Leiden	-64.24	-14.81	14.85	0.04
Dordrecht	-25.17	-10.22	10.27	0.05
Haarlem	-47.53	-15.09	15.20	0.12

Note: Comparative statics at the baseline equilibrium are obtained through simulation with the calibrated model. See the main text for details.

Table 5: Counterfactual analysis

City	Baseline	Constrained	Second-best	First-best
Development tax (€)				
Amsterdam	12.71	10.26	12.67	12.71
Rotterdam	2.36	0.99	2.34	2.36
The Hague	8.20	6.31	8.19	8.20
Utrecht	5.41	3.81	5.41	5.41
Leiden	9.11	7.24	9.13	9.11
Dordrecht	4.20	2.73	4.22	4.20
Haarlem	5.43	3.92	5.45	5.43
Tax on city creation (M€)	0	0	67.07	0
Number of households				
Amsterdam	527086	597130	631293	631867
Rotterdam	476514	516159	560539	561155
The Hague	305393	364872	391518	391700
Utrecht	191066	233685	253879	253840
Leiden	114024	162304	175751	175622
Dordrecht	102280	127465	139652	139483
Haarlem	89029	128231	141524	141342
Number of replicates	15.63	8.32	6.38	6.38
Total urban land use (ha)	71543	72802	67725	67724
Surplus (G€)	2.20	2.39	2.43	2.43

Note: Characteristics of counterfactual scenarios, simulated with the calibrated model. See the main text for a description of the alternative scenarios.

Figure 1: Randstad Holland
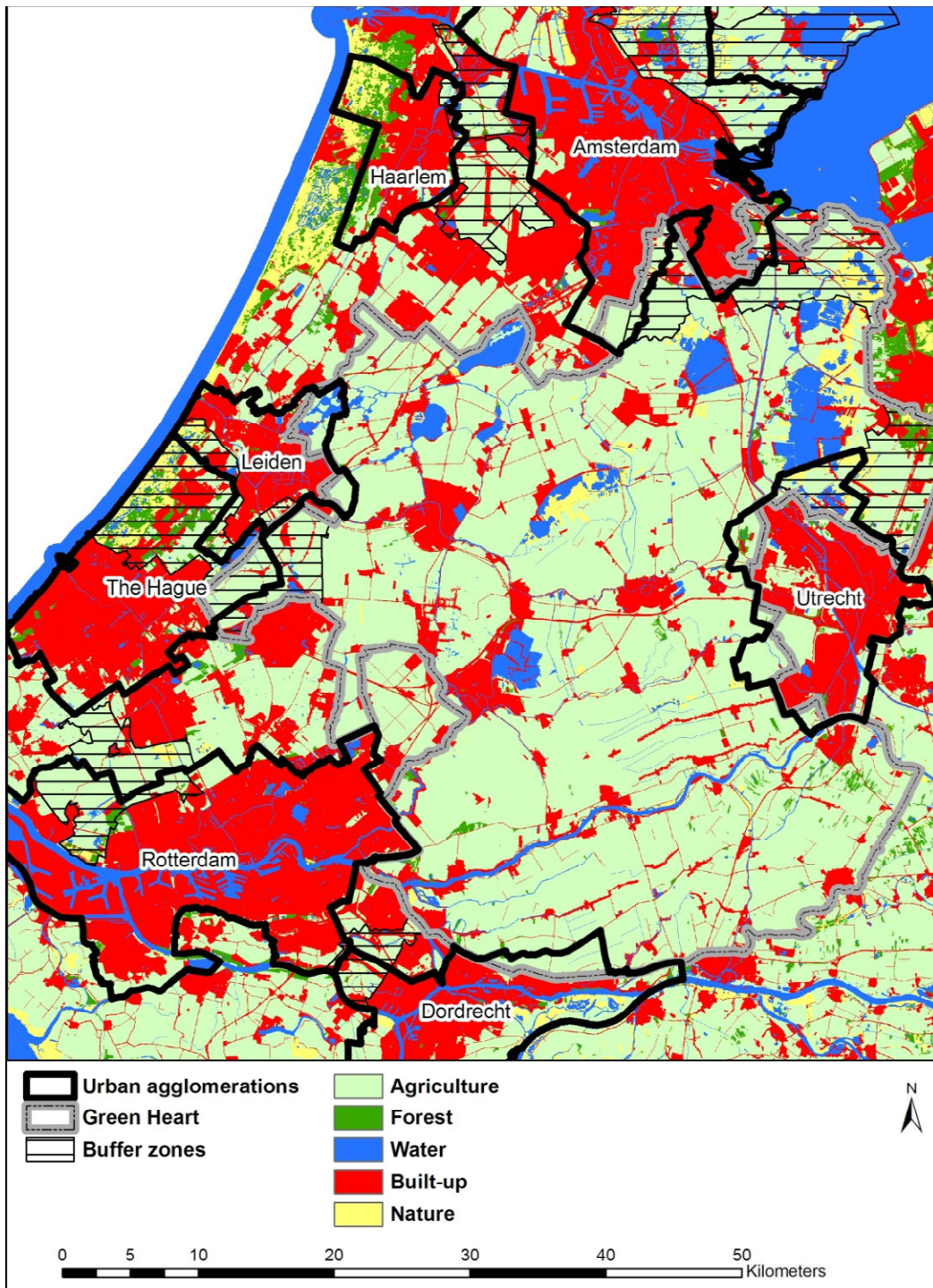
Note: Land use in 2003 and boundaries of the urban agglomerations are obtained from Statistics Netherlands. Boundaries of the Green Heart area and Buffer zones are based on VROM *et al.* (2004). This map was produced by Spinlab, VU University.

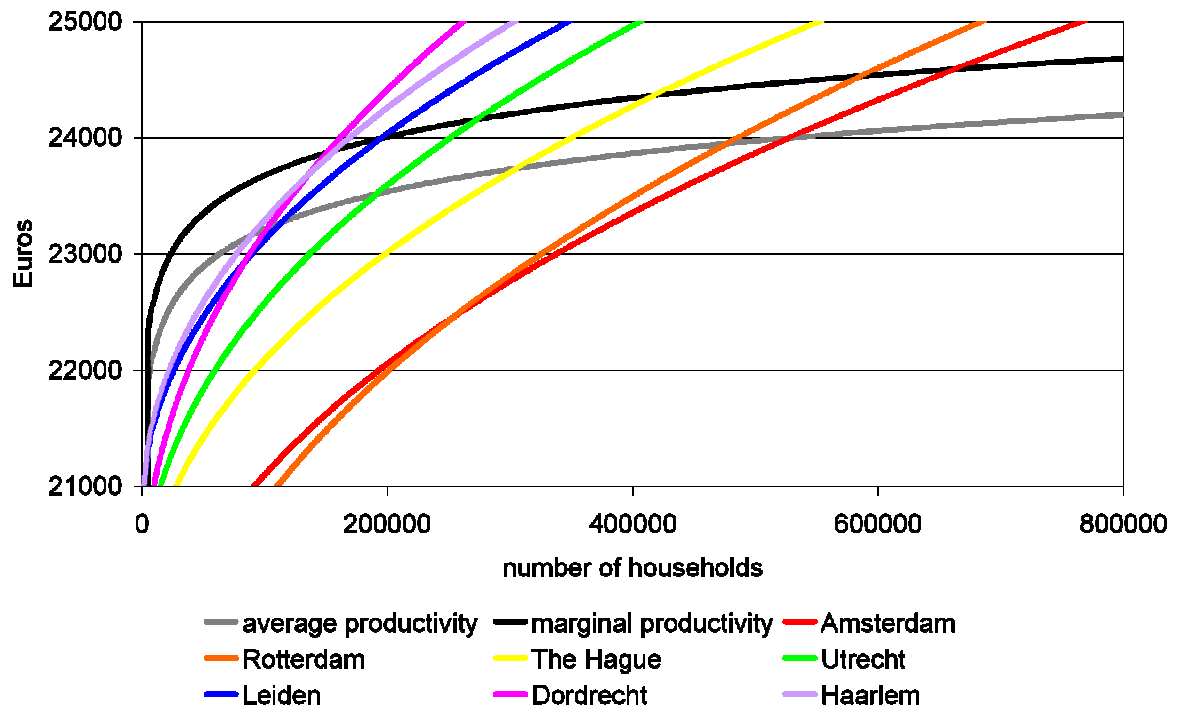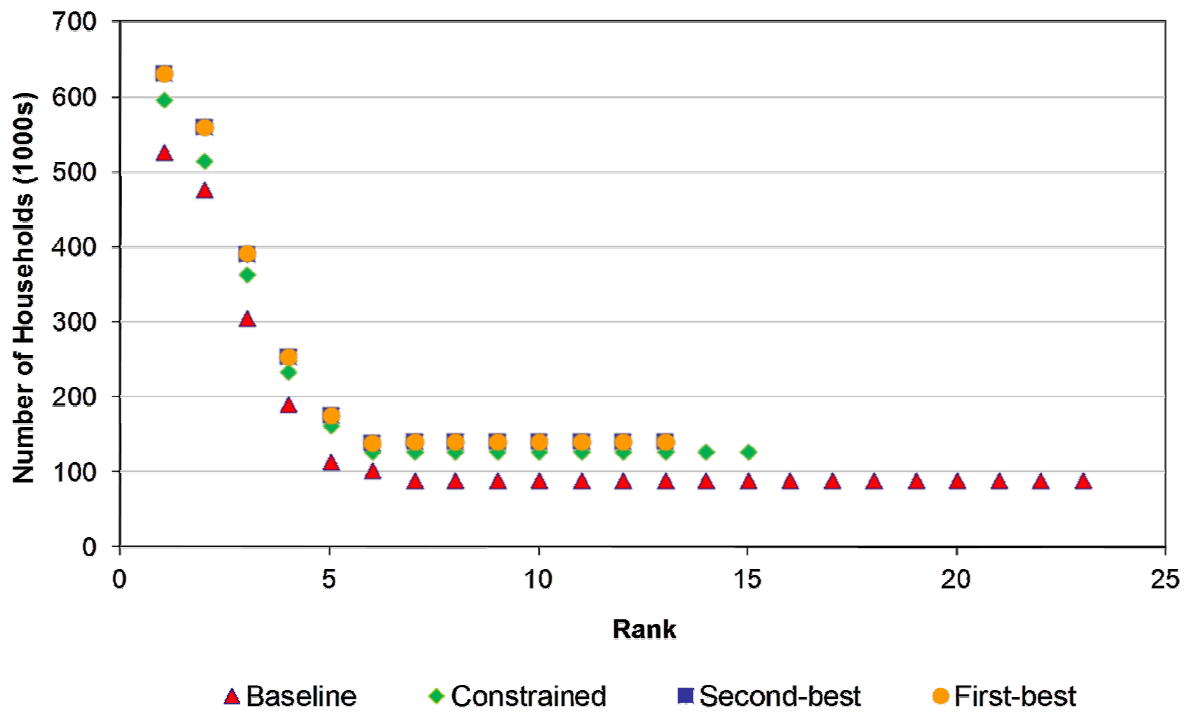*Figure 2: Urban labour demand and supply*

Figure 3: Counterfactual analysis