Urban Sustainability, Agglomeration Forces and the Technological Deus ex Machina

Erik Verhoef¹
Peter Nijkamp¹
Joe Daniel²

¹ Department of Spatial Economics, Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam, and Tinbergen Institute,
² Department of Economics, University of Delaware, USA
Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl
Abstract:
This paper aims to analyze and depict urban equilibrium from the perspective of a complex force field between (positive) agglomeration economies and (negative) environmental externalities. Based on a simplified representation of a linear urban economy, an archetypical model based on general equilibrium principles is designed and its properties are investigated by using numerical simulations. The model includes a spacious industrial centre in which agglomeration externalities are differentiated over space, and a residential area that suffers from pollution which too is differentiated over space. Environmental technology choice by firms is endogenized. This model is able to generate interesting and sometimes counter-intuitive results for the city under consideration.

1Peter Nijkamp and Erik Verhoef are affiliated to the Tinbergen Institute, Amsterdam-Rotterdam.
4The research of Erik Verhoef has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.
1. Towards an urban world

The historical evolution of our world is marked by a continual shift from a rural society to urban modes of living. According to UN estimates the urban population world-wide has grown from approximately 15 percent in 1900 to 55 percent in 2000 (with large differences between the developing and the developed world).

And even though in a recent book O’Sullivan (2000) points at uneasy, ambiguous and sometimes controversial views on the benefits of urbanization, the world-wide trend towards further urbanization is undeniable. A growth in urbanization may manifest itself in two ways, viz. a growth in existing cities (up to the level of mega-cities exceeding 10 million inhabitants, such as Mexico City and Sao Paolo) and the emergence of new cities (with a rapid population growth, either planned in a new town context or unplanned or spontaneous). Both types of urbanization run nowadays in parallel, with the result that in many countries the urbanization rate exceeds 70 percent.

Is there any reason why urbanization should come to a standstill? In the past, urban economists have tried to derive the optimal size of cities by minimizing the overall urban costs per inhabitant, but reality has been rather harsh in that modern cities never stopped growing. Of course, there have been temporary slowdowns and declines, but from a structural perspective it is noteworthy that urban areas (including suburban areas) have shown a remarkably stable growth pattern to the detriment of rural areas. Even the de-urbanization hypothesis has never exhibited a strong trend toward a global decline of cities (be it with a few exceptions).

The city as a nucleus of human activity has generated many economies of density, most of them positive, but also some very negative (e.g., organized crime, environmental decay, traffic congestion etc.). And as a consequence we witness nowadays an increasing concern about the urban environment as part of the global sustainability debate (see e.g. Stanners and Bourdeau, 1995). In the literature we find many debates on counter-urbanization versus urban sprawl, and this reflects the fact that positive and negative externalities are preponderant forces with varying degrees of dominance. This ambiguity in the urban economics literature calls for a solid analytical economic framework which should comprise the essential characteristics of the urban economy, with particular emphasis on the forces that shape (dis)economies in urban activity patterns. A prominent place should here be given to environmental externalities which may erode or even overrule agglomeration economies in the city. Against this background the present paper aims to present the fundamentals of an
urban economic model encapsulating both scale economies and environmental externalities. The conditions for a sustainable equilibrium can then be derived. A new element in the model is the inclusion of an energy saving and pollution reducing (i.e. environmental) technologies, which might act as a remedy against urban environmental decline. The functional mechanism of the latter recovery factor for the urban economy necessitates a resort to the endogenous growth literature. A further extension of this model implies that also scale economies in the abatement sector may emerge, suggesting that city size offers an appropriate platform for increasing returns to scale in abatement activities. This phenomenon may affect the conditions for sustainable urban development towards a larger city size. The model will be illustrated by some simple numerical simulations. Before outlining the contours and the specifications of our model, we will offer in the next sector a concise overview of recent contributions to the analysis of the urban environment.

2. In search of sustainable cities

Urban environmental problems are not an exclusive policy concern in modern times, but have played a role ever since the emergence of human settlements (see Banister et al. 1999). The scale and intensity of environmental decay is however, increasingly recognized as a major threat for a healthy urban future. In particular after the publication of the Brundtland Report (see WCED 1987) we have witnessed an increasing interest in environmental (and climatological) policy issues, not only world-wide but also locally. The avalanche of sustainability studies (see e.g. Van den Bergh 1996 for an overview) has also called for a renewed interest in urban environmental quality (see e.g. Banister et al. 1997, Breheny 1992, Capello et al. 1999, Finco and Nijkamp 2001, Haughton and Hunter 1994, Nijkamp and Perrels 1994, Selman 1996, Pearce 1999 and Satterthwaite 1997, 2000).

At present, two strands of literature may distinguished on urban sustainability. The first class of contributions stem from urban ecology and looks at sustainable cities from the perspective of a multidimensional set of environmental, social and cultural quality indicators of cities. In this category one finds also such concepts as the green city, the garden city or the eco-city. Often reference is made to the compliance with a priori specified threshold values (such as carrying capacity, noise levels, critical emission levels etc.). These ideas have gained much interest in policy circles, such as the Local Agenda 21 and OECD.

Another class of interest in urban sustainability originates from the urban economics literature. Most emphasis is laid here on efficiency principles in a rather abstract way. There are however some notable exceptions. For example, almost two decades ago, already
Orishimo (1982) demonstrated that city size is related to urban environmental qualities, while the actual externalities are determined by urban land use, the urban transport system and the urban way of life. It should be noted that cities are not the sole sources of environmental decay. In fact, cities provide a wide range of promising possibilities for energy-efficient activity patterns (e.g., district heating, public transport) and environmentally-benign modes of production (e.g., combined heat and power, waste water treatment) and consumption (e.g., solar energy, insulation of apartment buildings). From this perspective, a city may be seen as an efficient way of spatially organizing human activity. Any other spatial organization of our world is likely less efficient from an economic, environmental and energy angle.

Clearly, in the eyes of the public the negative externalities are most visible. And it is therefore no surprise that a ‘flight to the suburbs’ (urban sprawl or a movement to green areas) has taken place. But, nevertheless, cities have tried to turn the tide by various rehabilitation and recovery programmes, and in various case surprising successes have been achieved. In a recent article on ‘Are Cities Dying?’, Glaeser (1998) raises the question whether cities will survive, not only from an economic but also from a social and environmental perspective. The author rightly emphasizes the importance of agglomeration economies (e.g., in the urban labour market and the urban ICT sector). The city is the place ‘par excellence’ for spillovers of communication and information and hence for learning economies. Similar ideas have also been advocated by Fujita et al. (1999). Clearly, there are also diseconomies (such as congestion, pollution concentration, diseconomies of density, crime, urban anonymity etc.). It is a challenge of urban policy to strike a balance between such conflicting forces. For the time being there is no reason to believe that cities would vanish from earth, but such a belief ought to be substantiated by a more firm economic and less anecdotical foundation. Especially the role of modern technology - and perhaps of network organization – ought to be given more prominent attention (see e.g. Mokyr 1990, Wigand et al. 1997, and Evans and Wurster 2000). An important concern is however, that not all cities will survive to the same extent and will flourish in terms of equal economic and environmental opportunities. Due attention for and due public investment in furtherance of promising agglomeration economies is a sine qua non for urban sustainability. But to map out in a consistent economic way all the forces at work is a major challenge, particularly as each model of the urban economy tends to become immediately extremely complex, so that its analytical properties can hardly be traced anymore. Therefore, there is a need for a simple but rich model that is able to offer a rather representative mapping of a complex sustainable city. Such a model will be presented in the next section.
3. A spatial general equilibrium modelling framework for studying agglomeration and environmental externalities in an urban context

3.1. Introduction

As will be clear from the above discussion, for the study of the sustainability of cities, and of – environmental – policies for cities, a rather complex conceptual framework is required. Not only are different, often counteracting, forces at work (e.g. positive external effects such as agglomeration externalities and technological spill-overs versus negative ones such as environmental externalities), but, in addition, these typically vary over space in intensity. Indeed, if this were not the case, an important raison d’être for cities would be eliminated. For instance, if agglomeration advantages enjoyed would not somehow fall with the distance to other producers, there would be less or even no reason for firms to cluster together and pay high land rents in a dense city centre. The conceptual framework should thus at least be capable of dealing with the spatial dimension. Furthermore, as sustainability (typically) refers to a long-run goal, a partial equilibrium analysis may be problematic as it would ignore long run indirect effects of environmental policies on, for instance, land rents and on urban labour market conditions, and the resulting repercussions on input choice in the sector considered. Especially if the environmental externalities caused by firms are directly related to one or more of the inputs used – as will be assumed in our model – a general equilibrium approach seems preferable.

The analytic framework for formalizing the issues raised in the previous sections that will be presented in this section therefore concerns a continuous space urban general equilibrium model. Given the inherent complexity of this type of models, we will have to make many simplifying assumptions, just to keep the exercise manageable and the results interpretable. We therefore emphasize here that the key purpose of the model is to describe a continuous-space system that captures what we consider to be essential aspects of the problem at hand – environmental externalities, agglomeration externalities, input substitution, general equilibrium, all from a spatial perspective – in the most basic form, and to map out the possible consequences of environmental policies in such a system. The model is not intended to describe a realistic city – it is only intended to describe and analyse economic principles that would be relevant in a realistic city, in a coherent framework.

The model builds upon earlier work in Verhoef and Nijkamp (2001). In that paper, the trade-off between agglomeration economies and environmental externalities was extensively investigated. In the present model we have a considerably richer representation of the
production sector, especially by dropping the assumption of a spaceless industrial district, by considering agglomeration externalities as a spatially differentiated phenomenon, and by introducing endogenous technology choice of firms with a view to energy saving. At the same time, the model has a much simpler representation of the residential sector and households’ decision making: we will assume here fixed labour supply, and also assume that residential lot sizes are given.

Some introductory remarks are in order. First, $z$ will be used to denote a one-dimensional continuous urban space. Our model will produce a symmetrical city. Unless explicitly stated otherwise, we will be considering only one half of this city, knowing that the other half will be identical. The (endogenously determined) centre of the central industrial district (CID) defines $z=0$ (without loss of generality), and the CID stretches to the endogenously determined boundary with the suburban residential district (SRD), denoted as $z^\#$. (the spatial demarcation of industrial production and the residential area is not assumed beforehand, but will follow endogenously from the model specification). At the separation point $z^\#$, the equilibrium industrial and residential ‘bid-rents’ $r_I$ and $r_R$ must be equalized and cross, following the standard rule that in a competitive equilibrium situation, land should go to the highest bidder. The SRD stretches from $z^\#$ to $z^*$, with $z^*$ being the a priori unknown endogenous city boundary. At $z^*$, the equilibrium residential bid-rent $r_R$ must be equalized to the exogenous and constant agricultural bid-rent $r_A$, for the same reason as above. In equilibrium, no household would have an incentive to move beyond $z^*$, as it would increase commuting costs without reducing land-rent. Neither would a firm or household have an incentive to move to the SRD or CID, respectively, as the prevailing rent would be higher than their bid-rent. Moreover, within the CID, profits must be constant over space in equilibrium, and equal to zero by our assumption of perfect competition. Within the SRD, utility must be constant over space. Both imply that the actual equilibrium rent $r$ should be equal to the bid-rents $r_I$ in the CID and $r_R$ in the SRD. These conditions, too, reflect the idea that in a spatial equilibrium with endogenous land prices, no economic actor would have an incentive to relocate and could benefit from outbidding another actor occupying a certain lot. The equilibrium land-rent is therefore given by $r(z) = \text{MAX}\{r_I(z), r_R(z), r_A\}$.

---

1 The bid-rent is defined as the maximum rent a firm or household can pay per unit space at different locations in order to obtain the maximum profit level (for firms; typically zero) or utility level (for households) that is achievable in the city.
It is assumed that all excess urban land rents above $r_A$ are redistributed among the city’s population. Alternatively, we could have used the ‘absentee land-lord assumption’, which would seem less realistic as it assumes that none of the land rents generated in the city would be used for consumption in the city. It would also be implausible to assume that all land rents generated in the city are redistributed among the population, as this would imply that the endogenous city size can be expanded costlessly. The present representation would correspond to the situation where the public authority of the city buys the urban land against the relatively low rural land price, implying an equivalent per-unit-of time price of $r_A$, and redistributes all excess rents generated in the city among its population. It is a convenient assumption in the sense that it easily allows us to consider households with identical initial endowments. If, for instance, residential land were privately owned, someone paying a high rent would at the same time receive that high rent, leaving him no worse off in budgetary terms than someone paying a low rent. Similarly, it is assumed that all (environmental) tax revenues generated in the city, $TAX$, are redistributed in a uniform, lump-sum way among the population.

Next, we turn to the resource and environmental sector in the city. Pollution in the CID is assumed to result proportionally from the use of one of the inputs, energy. Like land, also energy is bought against a given price on an open ‘world market’. The same holds for environmental technologies, which are assumed to be of a pure energy-saving nature, and thus to simultaneously affect both the internal (energy) costs of a firm, and – via the proportional relation – the external costs resulting from the pollution emitted. These technologies are assumed not to be produced within the city itself, but to be offered against an exogenous price in the ‘rest of the world’. Some share of the urban production will therefore not be consumed in the urban area, but will be exported in exchange for the energy input, environmental technologies, and for the purchase of urban land – both industrial and residential – against the agricultural rent.

Some final assumptions and remarks are to be made. All consumers and producers are assumed to be price-takers. Households are identical, and so are firms. The industrial product can be transported costlessly. The price of the industrial good can be used as the numéraire. However, since also the ‘terms of trade’ for agricultural land, the energy input and environmental technologies are assumed to be exogenous, four prices can be set beforehand. We will now turn to the various actors in the city and the resulting equilibrium issues.
3.2. Consumers

Our closed city has $N$ households, which we will treat as a continuum of single economic entities. A household’s utility depends on the consumption of the industrial good, $y$, on the consumption of space or the size of the residence, $s$, and on the environmental quality, $Eq$. A household’s financial budget consists of the endogenous wage rate times the fixed amount of labour supplied per household (which we set at unity), plus the redistributed excess urban land rents. In equilibrium, this budget is fully spent on the consumption of $y$ and $s$. All prices are treated parametrically by the (price-taking) households; $w$ denotes the wage rate, $p$ the price of the industrial good, and $r(z)$ the rent. Commuting does not require financial outlays, but costs time at a given rate $t$ (there is no congestion). We make this simplifying assumption to avoid having to include a market for commuters’ transport in our general equilibrium framework. Also strictly for convenience, we will ignore differences in commuting distance to firms whose locations within the CID actually differ (also wages will be assumed to be equalized within the CID). We simply assume that all commuters travel to the centre of the CID, so that the total commuting time $T_c(z)$ is equal to $t\cdot z$. Finally, we assume that residential lot sizes $s_R$ are given, so that there is no substitution between the consumption of space and other consumption goods.

The fact that labour supply and space consumption are exogenous leads to a trivial household’s optimization problem, which we present here only because we need it for the determination of equilibrium rents in the SRD. The household’s maximization problem is dependent on the residential location $z$, and can be written as follows:

\[
\begin{align*}
\text{Max } U(y(z), s, T_c(z)) \\
\text{s.t. } \frac{R + \text{TAX}}{N} + w - p \cdot y(z) - r(z) \cdot s_R = 0
\end{align*}
\]

with:

\[
R = \int_0^z (r(z) - r_A)dz
\]

A spatial equilibrium requires that $U(z)$ be constant over $z$ for all $z^* \leq z \leq z^*$ (and exceeds $U(z)$ for all other $z$). Given the structure of the utility function and the spatial pattern of environmental quality $Eq(z)$, this implies a particular equilibrium pattern of land-rents. We can be more explicit about this when postulating a specific form for the utility function. In the numerical model used below, we will be using a very convenient linearly additive structure (which can be used because all consumed quantities except $y$ are determined exogenously):
\[ U(z) = \alpha_y \cdot y(z) + \alpha_s \cdot s_R - \alpha_r \cdot r(z) \cdot z + \alpha_e \cdot Eq(z) \]  \tag{3} 

where each \( \alpha \) is a parameter. The conditional demand for \( y \) follows directly from the constraint in (1): 

\[ y(z) = \frac{R + TAX - w \cdot r(z) \cdot s_R}{N \cdot p} \]  \tag{4} 

and the indirect utility – the maximum utility achievable under given prices and wage – can be written as:

\[ V(z) = \frac{R + TAX - w \cdot r(z) \cdot s}{N \cdot p} + \alpha_y \cdot s_R - \alpha_r \cdot r(z) \cdot z + \alpha_e \cdot Eq(z) \]  \tag{5} 

The condition that \( V(z) \) be constant over place implies:

\[ V'(z) = \alpha_y \cdot \frac{-r'(z) \cdot s_R}{p} - \alpha_r \cdot r(z) + \alpha_e \cdot Eq'(z) = 0 \]  \tag{6a} 

which implies the following slope for the residential bid-rent gradient:

\[ r'_R(z) = \frac{p \cdot \left( -\alpha_r \cdot r(z) + \alpha_e \cdot Eq'(z) \right)}{\alpha_y \cdot s_R} \]  \tag{6b} 

With a linear decay function for the effect of CID pollution on environmental quality \( Eq(z) \), as we are assuming for the sake of convenience, (6b) will be constant over \( z \), implying that the equilibrium rent gradient must be linear in the SRD. A ‘compact’ city requires (6b) and hence its numerator to be smaller than zero: the per unit of distance change in commuting disutility should outweigh the change in environmental disutility. The location and size of the SRD, given \( r_I(z) \) and \( r_A \), is then defined by the solution of the following two equations in \( z^# \) and \( z^* \):

\[ \frac{z^* - z^#}{s_R} = N \]  \tag{7a} 

and:

\[ r_A + r'_R \cdot (z^# - z^*) = r_I(z^#) \]  \tag{7b} 

Given this solution, the equilibrium rent in the SRD is most easily written as:

\[ r(z) = r_R(z) = r_A + r'_R \cdot (z - z^*) \quad \forall \ z^# \leq z \leq z^* \]  \tag{8} 

Labour supply is fixed at the population size:

\[ L^s = N \]  \tag{9} 

The total consumption of the industrial product \( Y \) should satisfy:
Finally, we define environmental quality such that a virgin state corresponds to $Eq = Eq^V$. Denoting the environmental quality at the edge of the CID as $Eq(z^*)$, to be defined as a function of emissions in equation (20) below, a spatially differentiated externality can be represented by letting $Eq(z)$ be a decreasing function of $Eq(z^*)$ and an increasing function of $z$. We will be using the following piecewise linear relation:

$$Eq(z) = \min \{Eq^V, Eq(z^*) + Eq^0 \cdot (z - z^*) \cdot (Eq^V - Eq(z^*))\} \quad \forall z > z^*$$ (11)

where $Eq^0$ is a parameter. We thus assume a linear distance-decay relation for the impact of every unit of pollution generated in the CID. Throughout the paper – including in fact the derivation of (7a)-(8) above – we will assume that $Eq(z^*) < Eq^V$, so that a strictly linear function $Eq(z)$ can be used.

1.3. Firms

There is a continuum of firms, each of which infinitesimally small relative to the market and taking all prices as given. The industrial output is homogeneous, and agglomeration externalities in our model thus arise from a more efficient production when the scale of aggregate production increases. The agglomeration benefits enjoyed are assumed to be dependent on the firm’s location relative to that of others, and are summarized in an efficiency measure $a(z)$. An individual firm takes the equilibrium pattern of $a(z)$ as given, but $a(z)$ is endogenous on the city level, which represents the existence of external agglomeration economies.

Firms have a CES production technology with three inputs (labour, energy and space), which allows us to consider input substitution. A firm’s production function is assumed to exhibit constant returns to scale, and therefore qualifies for application of Euler’s theorem. Therefore, also when the urban aggregate production function exhibits increasing returns to scale due to agglomeration externalities, we can model the firms’ behaviour using a ‘derived production function’ with constant returns to scale, in which the efficiency measure $a(z)$ is treated parametrically. The following derived aggregate production function applies:

$$q(z) = a(z) \cdot \left( \left( \delta_L \cdot l(z) \right)^{\rho} + \left( \delta_E \cdot En(z) \right)^{\rho} + \left( \delta_S \cdot s(z) \right)^{\rho} \right)^{\frac{1}{\rho}}$$ (12)

where $l(z)$ is the labour input, $En(z)$ the generalized energy input (a bundle of pure energy and energy technology; see below), and $s(z)$ the land input; and $\delta_L$, $\delta_E$, $\delta_S$ and $\rho$ are parameters (at least at the firm level), where $\rho$ ($\rho \neq 0, \leq 1$) defines the elasticity of substitution $\sigma$ according to
\(\sigma = 1/(1-\rho)\). In particular, \(\rho = -\infty\) corresponds to a Leontief production function, \(\rho = 1\) to a linear production function, and the limit of \(\rho \to 0\) would reproduce a Cobb-Douglas production function. A convenient parameter often used when working with CES production functions is \(\pi = \rho/(\rho - 1)\), which we too will be using below.

Verhoef and Nijkamp (2000) consider two different types of agglomeration economies for the non-spatial CID they considered, namely one in which the non-localized efficiency measure \(A\) increases in aggregate production \(Q\) (‘Type Q agglomeration economies’) and one in which it increases in aggregate labour supply \(L\) (‘Type L agglomeration economies’). Both formulations have been used in the literature. Sullivan (1986), for instance, uses Type Q agglomeration effects, whereas Arnott (1979) and Fujita (1989, Section 8.2) use Type L.

In our present model, we will consider ‘Type Q agglomeration economies’ only. As stated, we will assume that these depend on a firm’s location relative to that of others. Specifically, the agglomeration economies enjoyed are assumed to increase in proximity-weighted total production, which we represent by using the following specification:

\[
a(z) = a^0 + a^1 \left( \int_0^z \frac{q(x)}{a^2 + [z - x]^2} dx + \int_z^\infty \frac{q(x)}{a^2 + z + x} dx \right) \tag{13}
\]

with \(a^0 = a^0_0 + a^0_{CID}\)

where \(a^0 - a^3\) are parameters. The first of these, \(a^0\), represents the efficiency that would be obtained by a firm producing in the CID when agglomeration externalities were absent, and thus reflects the efficiency that would be obtained by a firm that would produce in complete isolation (\(a^0_0\)), in addition to factors such as the quality of the CID’s transport and communication infrastructures, (not formally modelled) service sectors, etc. (\(a^0_{CID}\)). The parameter \(a^1\) gives the relative importance of proximity weighted production, given by the terms between the large parentheses. These two terms distinguish between firms on one’s own side of the CID (the first term) and those on the other side (the second term). Note that this is an instance where we have to consider the second half of our city, too; otherwise we would identify the wrong location as the one with the highest \(a(z)\). Otherwise, these two terms reveal that production elsewhere in the CID is weighted by the inverse of distance raised to some power \(a^3\) to determine the impact on \(a(z)\).

The function \(a(z)\) thus in a simple but conceptually appealing way reflects spatially differentiated production-dependent agglomeration externalities as occurring in the CID, which may be the result of all sorts of technology and knowledge spill-overs as they may occur between firms that cluster in space. For reasons of analytical and numerical simplicity,
these spill-overs are lumped together in a single, one-dimensional space-varying efficiency measure. Despite the resulting artificial and conceptual nature of the resulting measure \( a(z) \), we believe this is a meaningful way of endogenizing agglomeration externalities in a spatial general equilibrium model of perfect competition.

The following conditional demand functions can be derived when solving the firms’ cost minimizing problem under the constraint implied by the production function in (12):

\[
I(z) = w^{\pi-1} \cdot \delta_L^{-\pi} \cdot \left( \frac{w}{\delta_L} \right)^{\pi} + \left( \frac{p_{En}}{\delta_E} \right)^{\pi} + \left( \frac{p(z)}{\delta_S} \right)^{\pi} \cdot \frac{q(z)}{a(z)} \tag{14a}
\]

\[
E_n(z) = p\frac{En}{En} \cdot \delta_E^{-\pi} \cdot \left( \frac{w}{\delta_L} \right)^{\pi} + \left( \frac{p_{En}}{\delta_E} \right)^{\pi} + \left( \frac{p(z)}{\delta_S} \right)^{\pi} \cdot \frac{q(z)}{a(z)} \tag{14b}
\]

\[
s(z) = r(z)^{\pi-1} \cdot \delta_S^{-\pi} \cdot \left( \frac{w}{\delta_L} \right)^{\pi} + \left( \frac{p_{En}}{\delta_E} \right)^{\pi} + \left( \frac{p(z)}{\delta_S} \right)^{\pi} \cdot \frac{q(z)}{a(z)} \tag{14c}
\]

where \( p_{En} \) gives the price of the generalized energy input. As stated, the generalized energy input is composed of a ‘pure energy’ part, denoted \( E \), and a technology part, denoted \( T \). We assume that these two inputs are purchased against given prices \( p_E \) and \( p_T \) on the world market, and that they can be combined in a constant-returns-to-scale Cobb-Douglas production function to produce the generalized energy input \( En \). This gives us an intuitive formulation that enables us to study substitution between pure energy and energy saving technologies that may occur in response to regulatory policies. Note that this formulation makes sure that the overall bi-layered production function still exhibits constant returns to scale, as required for our assumptions of perfect competition and zero profits.

The generalized energy input is thus produced according to:

\[
En(z) = E(z)^{\frac{\epsilon}{\epsilon}} \cdot T(z)^{1-\epsilon} \tag{15}
\]

The conditional demand functions for \( E \) and \( T \) can again be derived from the cost minimization problem, and read:

\[
E(z) = \frac{\epsilon}{p_E + \tau_E} \cdot \left( \frac{p_E + \tau_E}{\epsilon} \right)^{\frac{\epsilon}{\epsilon}} \cdot \left( \frac{p_T}{1-\epsilon} \right)^{1-\epsilon} \cdot En(z) \tag{16a}
\]

\[
T(z) = \frac{1-\epsilon}{p_T} \cdot \left( \frac{p_E + \tau_E}{\epsilon} \right)^{\frac{\epsilon}{\epsilon}} \cdot \left( \frac{p_T}{1-\epsilon} \right)^{1-\epsilon} \cdot En(z) \tag{16b}
\]
where \( \tau_E \) denotes the (uniform) energy tax that the urban regulator will be assumed to use to affect energy consumption and hence pollution in the CID. The total tax revenues generated with this tax \( \tau_E \) are given by:

\[
TAX = \int_0^z \tau_E \cdot \frac{E(z)}{s(z)} \, dz
\]  

(17)

From (16a) and (16b), it can be inferred that the generalized energy price \( p_{En} \) perceived by a firm – given that it selects the cost-minimizing input combination \( E(z) \) and \( T(z) \) – is indeed independent of its location \( z \) (as used in (14a-c) above), and can be written as:

\[
p_{En} = \left( \frac{p_E + \tau_E}{\epsilon} \right)^{\epsilon} \cdot \left( \frac{p_T}{1-\epsilon} \right)^{1-\epsilon}
\]  

(18)

In a similar way, the average costs that can be derived form (14a-c) in combination with the zero-profit condition and the constancy of prices over the CID imply that the following condition should hold:

\[
p = \frac{1}{a(z)} \cdot \left( \frac{w}{\delta_L} \right)^{\epsilon} + \left( \frac{p_{En}}{\delta_E} \right)^{\epsilon} + \left( \frac{r(z)}{\delta_S} \right)^{\epsilon} \quad \forall \ 0 \leq z \leq z^a
\]  

(19)

Equation (19) shows that \( r(z) \) cannot be constant over space in the CID in equilibrium when \( a(z) \) is not constant – which it generally will not be by (13).

From equation (16a), and assuming that pollution is proportional to the use of the pure energy input with a factor \( e \), and that from the point of emission, the decay of the environmental externality immediately sets in, \( Eq(z^a) \) can be written as:

\[
Eq(z^a) = Eq^v - \int_0^z \epsilon \cdot \frac{E(z)}{s(z)} \cdot \left( 1 - Eq^v \cdot (z^a - z) \right) \, dz
\]  

(20)

Note that we assume that pollution moves only one way in our one-dimensional model, namely from the point of emission to the SRD. This is not an essential assumption, but it does imply that we ignore the impact of emissions from the implicit ‘other half of the CID’ for the SRD we have modelled. For a fully symmetric ‘complete’ city, this assumption would correspond with the situation where the emitted pollutant would disperse exclusively by flowing from a point with a high concentration \( (z=0) \) to points with lower concentrations, so that emissions from the one side of the CID would indeed never reach the SRD on the other side of the city. Clearly, other assumptions on the nature of dispersion and decay of pollutants could be made just as well in our model.
Total production $Q$, total labour demand $L^D$, total energy use $EN$ and its components $E$ and $T$, and total land use $S$ by the industrial sector can be written as:

\[ Q = \int_{z^*}^{z^*} \frac{q(z)}{s(z)} \, dz \]  

(21a)

\[ L^D = \int_{z^*}^{z^*} \frac{l(z)}{s(z)} \, dz \]  

(21b)

\[ EN = \int_{z^*}^{z^*} \frac{EN(z)}{s(z)} \, dz \]  

(21c)

\[ E = \int_{z^*}^{z^*} \frac{E(z)}{s(z)} \, dz \]  

(21d)

\[ T = \int_{z^*}^{z^*} \frac{T(z)}{s(z)} \, dz \]  

(21e)

\[ S = \int_{z^*}^{z^*} \frac{s(z)}{s(z)} \, dz \equiv \ z^a \]  

(21f)

Equilibrium on the labour market finally requires:

\[ L^D = L^S = N \]  

(22)

1.4. General spatial equilibrium

It will not come as a surprise that the system described above has no easily manageable closed-form analytical solutions. A formal proof for existence and uniqueness of a spatial equilibrium would be beyond the scope of this paper. A very loose way for making existence plausible would be to count equations and unknowns. There are 14 unknown scalar variables ($z^#, z^*, Y, Eq(z^*), TAX, p_{En}, w, Q, L^S, L^D, En, E, T, S$) for which an equal number of (linearly independent) equations is available (7a-b, 10, 20, 17, 18, 22, 9, 21a-f); and there are 12 unknown functions of $z$ ($y(z), V(z), r_H(z), Eq(z), q(z), a(z), l(z), En(z), s(z), E(z), T(z), r_I(z)$) in the same number of equations (4, 5, 8 (after substitution of 6b), 11, 12, 13, 14a-c, 16a-b, 19). The other scalars ($N, p, p_E, p_T, r_A, e, \rho, \alpha's, \delta s, a's, e, Eq^V, Eq^0, t$ and $s_R$) are all exogenously given parameters, and $\tau_E$ can be treated as exogenous for the determination of equilibria. As stated, this counting of equations and unknowns is by no means conclusive, but at least suggests that the model could have a unique equilibrium. This, of course, requires a solution for the above 26 unknowns, satisfying the set of 26 equations mentioned.
The constant-returns-to-scale assumption in the bi-layered production function furthermore guarantees that Euler’s theorem applies, which means that all inputs can indeed be paid their marginal value productivity without running into economic losses or profits for the firms. As the tax revenues and excess rents (above \( r_A \)) are redistributed among the population, the trade balance condition and exhaustion of income for consumptive purposes will also be satisfied. To see why, first observe that a balance of trade would require:

\[
p \cdot (Q - Y) = r_A \cdot z^* + p_E \cdot E + p_T \cdot T
\]  

(23a)

and that the aggregate zero profit condition implies:

\[
p \cdot Q = \int r(z) - r_A \, dz + r_A \cdot z^* + w \cdot N + p_E \cdot E + p_T \cdot T + \text{TAX}
\]  

(23b)

Households spend the total wage sum \( w \cdot N \) plus the sum of tax revenues and excess rents on the consumption of the industrial product and on the consumption of residential land:

\[
w \cdot N + \int r(z) - r_A \, dz + \text{TAX} = p \cdot Y + \int r(z) - r_A \, dz + r_A \cdot (z^* - z^*)
\]  

(23c)

Substitution of (23c) into (23b) immediately yields (23a).

One condition for a spatial equilibrium with a compact city was already discussed in Section 3.1: the residential bid-rent should be negatively sloped. For the assumed monocentric city to be stable, the profits attainable for production outside the CID must be negative. Because we want to maintain focus on a monocentric city, we will simply assume that this is the case; i.e., that the equivalent of \( a^0 \) applying outside the CID (for instance ‘just’ \( a^0 \)) is sufficiently low to prevent profits from being positive in absence of agglomeration externalities at equilibrium land rents. It is clear that endogenization of this condition in the current model would allow us to study the formation of sub-centres in the same modelling framework – an issue that we do not want to include in the present exposition but postpone to future work.

We will not engage in a further inquiry into the existence, uniqueness and properties of equilibria of the formal model, but instead turn to a discussion of the results of a simulation model that was built fully consistent with the above model, and that allows a more insightful exposition of the properties of the model.
4. A numerical simulation model

4.1. Parametrization and base-case equilibrium

The numerical simulation model represents a fully imaginary city that operates exactly according to the model developed in Section 3. The model solves the set of equations defining equilibrium using the rather intuitive economic logic of starting with an exogenously defined disequilibrium, and then equilibrating markets one by one (sometimes in loops) until the convergence criteria are met. Table 1 shows the parameter values chosen. These were set such that the interpretation of results is made as easy as possible, by creating as much ‘balance’ as possible in the parametrization of the utility and production functions.

<table>
<thead>
<tr>
<th>Prices</th>
<th>Utility function</th>
<th>Upper layer production function</th>
<th>Lower layer production function</th>
<th>Environmental externalities function</th>
<th>Agglomeration externalities function</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>$\alpha_s = 1$</td>
<td>$\rho = -0.35$</td>
<td>$\epsilon = 0.5$</td>
<td>$Eq^V = 1$</td>
<td>$a_{NO}^V = a_{NO}^D = 0.5$</td>
<td>$N = 250$</td>
</tr>
<tr>
<td>$p_E = 0.1$</td>
<td>$\alpha_s = 1$</td>
<td>$\sigma = 1/(1-\rho) = 0.74$</td>
<td>$Eq^A = 0.005$</td>
<td>$a^E = 1$</td>
<td>$a^E' = 0.15$</td>
<td>$t = 0.001$</td>
</tr>
<tr>
<td>$p_T = 0.1$</td>
<td>$\alpha_t = 1$</td>
<td>$\pi = \rho/(\rho-1) = 0.26$</td>
<td>$e = 0.001$</td>
<td>$a^T = 1$</td>
<td>$a^T' = 1.5$</td>
<td></td>
</tr>
<tr>
<td>$r = 0.1$</td>
<td>$\alpha_s = 1$</td>
<td>$\delta_s = 10$</td>
<td></td>
<td>$a^s = 1$</td>
<td>$a^s' = 1.5$</td>
<td></td>
</tr>
<tr>
<td>$\tau_E = 0$</td>
<td>$\alpha_s = 1$</td>
<td>$\delta_s = 10$</td>
<td></td>
<td>$a^s = 1$</td>
<td>$a^s' = 1.5$</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Parametrization of the numerical model

The base-case of our model concerns the market equilibrium in absence of environmental taxation (i.e., $\tau_E = 0$). Table 2 shows the equilibrium values of the main endogenous (scalar) variables for the parametrization in Table 1. Given the conceptual character of the model, most of these equilibrium values have no particular meaning for a single equilibrium considered in isolation, but will become relevant only when performing comparative static analyses, for instance, when comparing the base-case equilibrium to one that results with environmental taxation. The only variables that do have some meaningful interpretation as characteristics of the base-case are the income shares, showing that 74% of the available money income is spent on the industrial good and the rest on land rents; factor shares, showing that 30% of the total production costs concerns labour, 33% the polluting input (energy), and 37% land rents; the factor shares within the generalized energy input, showing an equal distribution between pure energy and technology; the fact that the average efficiency level $a^{av}$ (averaged over production, not space) is 1.5, which is one-and-a-half times as much as the level that would be obtained without agglomeration externalities ($a^d = 1$); and the fact that around 50% of the land is used for production and 50% for residential purposes.
Table 2. Some key variables in the base-case equilibrium

<table>
<thead>
<tr>
<th>Prices</th>
<th>Utility</th>
<th>Upper layer production function</th>
<th>Lower layer production function</th>
<th>Agglomeration externalities</th>
<th>Size of the city</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w = 0.134)</td>
<td>(Y = 51.2)</td>
<td>(Q = 114)</td>
<td>(E = 186)</td>
<td>(a^\tau = 1.50)</td>
<td>(z^\tau = 123)</td>
</tr>
<tr>
<td>(p_{Ea} = 0.200)</td>
<td>(Eq(z^\tau) = 0.873)</td>
<td>(S = 123)</td>
<td>(T = 186)</td>
<td>(z^* = 248)</td>
<td></td>
</tr>
<tr>
<td>(\tau_E = 0)</td>
<td>(U^\nu = 1.43)</td>
<td>(En = 186)</td>
<td>(En = 186)</td>
<td>(En = 186)</td>
<td></td>
</tr>
</tbody>
</table>

**Income shares:**
- Ind. good: 74%
- Labour: 30%
- Housing: 26%
- Space: 37%

**Factor shares (in Energy):**
- Pure energy: 50%
- Tax: 0%
- Technology: 50%

**Factor shares:**
- Labour: 30%
- Energy: 33%
- Space: 37%

**Table 2. Some key variables in the base-case equilibrium**

![Figure 1](image-url). Land-rents (upper-left panel), utility and profits (upper-right panel), environmental quality (lower-left panel), and the efficiency indicator (lower-right panel) as a function of space in the base-case equilibrium (solid lines) and the second-best optimum (dotted lines).
The solid lines in Figure 1 depict the space patterns of some of the model’s key variables. The lower-right panel, for instance, shows the equilibrium pattern of \( a(z) \), which reflects the efficiency-enhancing impact of proximity-weighted production. The intuitive feature that \( a(z) \) obtains the highest value in the centre of the CID at \( z=0 \), is reflected in the spatial pattern of land-rents, which too reach a maximum at \( z=0 \) (upper-left panel). The CID extends to the intersection with the residential bid-rent function. As required for equilibrium, the industrial bid-rent intersects the residential one from above at \( z^# \) (not shown in the figure, but easily identifiable), and the same again holds for the residential and agricultural bid-rents at \( z^* \) (again not shown in the figure).

Two other important spatial equilibrium conditions are depicted in the upper-right panel: as required, utility \( U \) is constant over space (at the level \( U^{av}=1.43 \) shown in Table 2) within the SRD, and profits \( \Pi \) are zero and (hence) constant over space within the CID. Finally, the lower left panel shows that, as could be expected, environmental quality indeed increases over space, but nowhere in the city reaches the ‘virgin’ quality \( Eq^V = 1 \).

### 1.2. A second-best optimum: energy taxation

Evidently, our conceptual model is particularly useful for comparative static analyses, comparing the base-case equilibrium to equilibria as they would arise under some form of policy intervention, taking spatial general equilibrium interactions fully into account. In this paper, we consider one such policy, namely one that figures predominantly in the environmental economics literature: energy taxes. From the outset, we emphasize that a tax on energy use is a second-best instrument for two reasons in the current setting. First, the model has two important externalities: apart from the environmental externality, there is the agglomeration externality which means that the free market would most likely fail to achieve a Pareto efficient (spatial) equilibrium. Verhoef and Nijkamp (2001) studied the simultaneous regulation of environmental and agglomeration externalities. Secondly, due to the decay of pollution, a unit of energy used in the centre of the CID would lead to lower external costs in the SRD than a unit used near \( z^# \). Optimal pollution taxes would thus vary over space. Because we assume that the polluting input is taxed, which can freely be traded within the CID, such spatial tax differentiation is however impossible, and hence a second type of second-best distortion enters the picture.

The value added of using a spatial general equilibrium model for the analysis of the effects of such a policy instrument in an urban setting, is justified particularly convincingly if even a simple conceptual model would lead to qualitatively different results than would be
anticipated on the basis of ‘logical reasoning’, or simple partial equilibrium models, be they spatial or not. We therefore first give an intuitive reasoning of the qualitative effects of an energy tax in the present setting, which will next be proven wrong – or at least not necessarily correct – using the simulation model.

One would expect that a tax on the pure energy input would lead to a substitution away from pure energy to technology in the lower level production function (which indeed will be the case), and in the upper level production function away from the generalized energy input (which has become more expensive) to the other inputs, land and labour. The use of more land for production would imply that the density in the CID decreases, while its size increases, leading to a reduction in beneficial agglomeration externalities. As argued in Verhoef and Nijkamp (2001), there thus would be a conflict between optimizing environmental externalities and agglomeration externalities, where the pursuit of the former goal would go at the expense of the latter.

![Figure 2. Utility as a function of $\tau_E$ (normalized in base-case equilibrium)](image)

Now, which are the ‘true’ effects of an energy tax in the present model? To that end, we first find the optimal level of the second-best energy tax. As there is no closed-form solution for the second-best energy tax, we found this by numerical search, namely by considering the equilibrium utility level as a function of the energy tax. Figure 2 shows that the optimal value is near $\tau_E=0.09$ (visible irregularities in this and subsequent figures are due to numerical imprecision). Table 3 shows the same equilibrium levels of endogenous variables for the second-best optimum as Table 2 did for the base-case, while the dotted curves in Figure 1
depict the pattern of some key spatially differentiated variables in the second-best optimum. We re-emphasize that parameters were not chosen to represent any realistic situation, but much more to create sufficient differentiation between equilibria so that, for instance, the curves in Figure 1 lie sufficiently far apart. Note that this implies differences between the free-market and second-best optimal equilibria that would be considered unrealistically large by most readers (including ourselves).

<table>
<thead>
<tr>
<th>Prices</th>
<th>Utility</th>
<th>Upper layer production function</th>
<th>Lower layer production function</th>
<th>Agglomeration externalities</th>
<th>Size of the city</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w = 0.069$</td>
<td>$Y = 35.3$</td>
<td>$Q = 72.2$</td>
<td>$E = 65.4$</td>
<td>$a^o = 1.69$</td>
<td>$z^o = 53.3$</td>
</tr>
<tr>
<td>$p_{En} = 0.276$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r(z^o) = 0.280$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_E = 0.09$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Income shares:
- Ind. good: 60%
- Housing: 40%

Factor shares:
- Labour: 24%
- Energy: 35%
- Tax: 24%
- Technology: 50%

Table 3. Some key variables in the second-best optimum

The results show that in the second-best optimum, the use of pure energy has indeed been reduced, which is partly the result of a shift from pure energy towards energy saving technologies, partly the result of using less of the generalized energy input per unit of

Figure 3. Inputs and outputs as a function of $\tau_E$ (normalized in base-case equilibrium)
production, and partly the result of a lower equilibrium output level. We emphasize that this is not as obvious as might be expected, since the second-best tax could have been negative if the indirect effect on agglomeration externalities were negative and would have outweighed the beneficial direct impact on the environmental externality.

Figure 3 shows that as $\tau_E$ is raised, the proportional decline in the pure energy input is the strongest among all inputs and outputs. The decline in the use of the generalized energy input is significantly smaller, which is consistent with the fact that the use of energy saving technologies has a smaller decrease than that of overall output. The absolute use of the labour input has remained constant, which is the result of the general equilibrium nature of the model in combination with the assumption of fixed labour supply. Consistent with relative increase of labour per unit of output, the endogenous wage rate has decreased; see also Figure 4 (recall that $p$, $p_E$ and $p_T$ are exogenous).

![Figure 3](image)

**Figure 4. Input prices as a function of $\tau_E$ (normalized in base-case equilibrium)**

Probably the greatest surprise in the results is that, contrary to expectation, the absolute use of land for production ($z^\#$) decreases due to the energy tax, and that – consistent with the decrease in average land input per unit of product implied in Figure 3, and the resulting increase in density in the CID – the average efficiency parameter $a_{av}$ as well as the levels of $a(z)$ shown in Figure 1 increase. Intuitively, one would expect the opposite, namely a substitution of energy towards land, implying more, and less dense, land use in the CID instead.
The explanation for the paradox is not so difficult, but illustrates in a nice way the importance of using a spatial general equilibrium framework for studying the type of questions under consideration here. What happens is the following. The improved environmental quality implies that the residential bid-rent should become steeper for household equilibrium to hold. This reflects that more central housing locations have become relatively more attractive, as the benefits of reduced emissions increase when approaching the CID due to the assumed decay function. With a given agricultural land rent and a relatively inelastic demand for housing per household (perfectly inelastic in our model), this implies that \( r(z^g) \) must increase compared to the initial equilibrium – even before knowing the new equilibrium value of \( z^g \). As the general shape of \( r(z) \) will not change, this in turn implies an upward shift of the \( r(z) \) function. This will subsequently lead to a lower conditional demand for land for production, which increases density in the CID, which leads to higher agglomeration benefits (due to the assumed distance decay), which in turn will drive up CID land rents even further, until a new equilibrium is reached.

Of course, because we have a general equilibrium model in which literally everything affects everything directly or indirectly, a full explanation of why and how equilibrium values and patterns of endogenous variables change between equilibria is practically impossible to do. But the above explanation captures the key mechanisms. It can at the same time be noted that in the new equilibrium, the wage rate has decreased by so much that the anticipated initial substitution away from energy to land is more than compensated for by a substitution away from land towards labour.

5. Retrospect

In the above, we have presented a conceptual spatial general equilibrium model, and a numerical simulation model based on this, for the purpose of demonstrating a number of points that we consider important for the study of the sustainability of cities. The conceptual nature of the model is an important aspect to be emphasized: the model developed is not intended to describe a realistic city – it is only intended to describe and analyse in a coherent framework the economic principles that would be relevant in a realistic city. The exposition above served to offer an example of how such a model could be built, and which would be the type of analysis one could carry out with it.

We argued that – and why – a rather complex conceptual framework is required. Not only are different, often counteracting, forces at work (e.g. positive external effects such as agglomeration externalities and technological spill-overs versus negative ones such as...
environmental externalities), but these in addition typically vary over space in intensity. The conceptual framework should thus at least be capable of dealing with the spatial dimension. Furthermore, as sustainability (typically) refers to a long-run goal, a partial equilibrium analysis may be problematic, as it would ignore long run indirect effects of environmental policies on, for instance, land rents and on urban labour market conditions, and the resulting repercussions on input choice in the sector considered. Especially if the environmental externalities caused by firms are directly related to one or more of the inputs used – as will be assumed in our model – a general equilibrium approach seems preferable.

The numerical example demonstrated the importance of using such a modelling framework for the analysis of environmental pollution in cities and associated policies. Simple as that, the more complex the real system one is dealing with, the less predictable its behaviour becomes, implying that ‘loose reasoning’ on the basis of intuitive arguments only may cause one to make inferences that are opposite to what may happen in reality when all mutual interactions between the system’s elements are taken into account. In our example, it turned out that environmental policies would not necessarily lead to a reduction in (average) agglomeration benefits, but that the opposite may in fact occur.

We have also tried to show that one can actually get quite far in modelling spatial systems according to general equilibrium principles. Despite the model’s simplicity, it is in fact notable that apparently, it is not too difficult to construct a general equilibrium framework for an urban economy in which both agglomeration and environmental externalities exist and vary over continuous space; three inputs are used, among which labour which is supplied by households competing for the same land as producers; and (energy) technology choice is endogenous.

At the same time, we are well aware of the limitations of the conceptual model presented here. Among the long list of possible extensions, we would for instance have dynamics and heterogeneity (of firms and household, but also multi-sectoral urban economies, possibly with an R&D sector) high on our research agenda. Another extremely interesting topic would be the spatial lay-out of ‘free-market’ versus ‘optimal’ cities, especially if multiple production (sub-)centres are allowed for. An entirely different strand of research would look into empirical evidence and try to identify which are the key economic forces that explain the existence of cities, to determine to which extent these forces are externalities, and if so, to see what the existence of such externalities would imply for the expected efficiency and desirability of environmental policies in cities.
References