

TI 2012-137/VIII  
Tinbergen Institute Discussion Paper



# Dynamic Congestion and Urban Equilibrium

*Sergejs Gubins*  
*Erik T. Verhoef*

*Faculty of Economics and Business Administration, VU University Amsterdam, and  
Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Dynamic Congestion and Urban Equilibrium

**Sergejs Gubins<sup>\*</sup>, Erik T. Verhoef**

Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam,  
The Netherlands

Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands

## Abstract

We consider a monocentric city where a traffic bottleneck is located at the entrance of the central business district. The commuters' choices of the departure times from home, residential location, and lot size, are all endogenous. We show that elimination of queuing time under optimal road pricing induces individuals to spend more time at home and to have larger houses, causing urban sprawl. This is opposite to the typical results of urban models with static congestion, which predict cities to become denser with road pricing.

*JEL classification:* D62, R21, R41, R48

*Keywords:* dynamic traffic congestion, urban equilibrium, road pricing, bottleneck model, monocentric model

---

<sup>\*</sup> Corresponding author. Tel.: +31 205988106.

E-mail addresses: s.gubins@vu.nl (S. Gubins), e.t.verhoef@vu.nl (E.T. Verhoef).

Financial support from The Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged. This paper is part of TRISTAM project (Traveler Response and Information Service Technology – Analysis and Modeling). Verhoef furthermore gratefully acknowledges support from ERC Advanced Grant #246969. We thank for useful discussions and comments Thomas de Graaff, seminar participants at Tinbergen Institute, and conference participants at Kuhmo-Nectar (Berlin, July 2012), ERSA (Bratislava, August 2012) and NARSC (Ottawa, November 2012).

## 1. Introduction

Peak-hour traffic congestion is a notorious phenomenon in modern cities. The majority of drivers during the morning peak period are commuters who are bound for a particular destination often within a business district, and who have a certain preferred time of arrival at work, usually 9am when business hours start. Congestion, which is the result of peoples' common desires to be at the same place and at the same time, is one of the most noticeable negative side-effects of colocation in space. Alleviation of congestion remains a pressing issue.

Congestion in an urban setting, where both spatial and travel behavior are endogenous, has proven to be a challenging topic for economic analysis (e.g., Ross and Yinger 2000). By far the most common way of modeling congestion in an urban setting is in terms of static flow congestion, where the timing of travel is not the result of optimizing behavior, and where traffic flows and speeds are constant over time. Essentially, that approach renders travel as an atemporal activity. For example, Solow and Vickrey (1971) relate the cost of travelling at a certain location to the total number of drivers passing it. That specification of travel cost does not allow the consideration of potential benefits of avoiding the peak by travelling earlier or later. But congestion in reality is a dynamic phenomenon, with time-varying speed and traffic flows. This has inspired transportation economists to develop dynamic models of traffic congestion, in which choice of departure time is endogenous, and where dynamic patterns of travel are key features. As Vickrey showed in 1969, such models may produce insights that diverge substantially from those traditional static models.

To develop a model that accounts for the dynamic nature of travel in a city, we integrate two strands of models in the urban economics and transportation economics literature. A classic Alonso-Muth-Mills monocentric city model, in some important respects probably "the most successful model in urban economics" (Glaeser 2008, p. 18), defines the spatial equilibrium, in which no individual can unilaterally reallocate within a city to gain a higher utility. Vickrey's dynamic bottleneck model of peak period congestion, probably the most widely used dynamic model of traffic congestion in transportation economics, allows for the analysis of a dynamic equilibrium, in which no driver can gain a higher utility by unilaterally changing the departure time (see, for example, the exposition in Small and Verhoef 2007). We aim to unite those two models in order to study the relationship between spatial allocation and travel time decisions.

In our model, homogeneous inhabitants<sup>1</sup> with a common preferred time of arrival at work, commute from their homes to a workplace located in the central business district (CBD) of a city. The city is congestion-free, apart from a road bottleneck right at the entrance to the CBD. One might think of a bridge or a junction that a driver has to pass to get to the workplace. When the inflow of cars exceeds bottleneck's capacity, a queue grows in front of the bottleneck. By adjusting the timing of departure from home, a driver can incur different queuing times, and arrive at work at the time moments with different levels of inconvenience. For instance, a driver can depart very early in the morning and avoid the queue altogether, or arrive at the most desirable moment but not after having spent quite some

---

<sup>1</sup> All city inhabitants are drivers and we use those two words interchangeably.

time in the queue. Following the framework of Vickrey (1973), and later Tseng and Verhoef (2008) and Fosgerau and Engelson (2011), we explicitly define the utility that an individual derives from spending time at home, at work, and in a car. That will later allow us to incorporate the spatial aspect of the model in a structured manner.

An important assumption that we introduce is that the marginal utility of spending time at home depends on the size of the house, which is endogenously defined in the model. *Ceteris paribus*, the larger the house one lives in, the more utility one derives from spending additional time in it. This assumption comes naturally from the observed preferences for larger housing, as shown by hedonic price studies (see, e.g., Bajari and Kahn 2008). We impose a time aspect by stating that an individual derives utility from a house by spending time in it. While a house might yield utility via various channels, e.g., by being a storage place, a status symbol, or an investment tool, we believe it is safe to assume that at least some part of the utility from having a house depends on the amount of time an individual spends in it. The usual pattern in the monocentric city model, with land consumption increasing the further away one lives from the CBD, thus produces variation in scheduling preferences across drivers, as a driver who lives closer to the CBD values time spent at home at a lower rate than one who lives at the city fringe. As we show later, depending on the size of the house, a driver chooses a certain time of departure from home, and vice versa. In this manner, we connect the transport model of bottleneck congestion, that applies a dynamic equilibrium condition, with the urban model of a monocentric city, that applies a spatial equilibrium condition.

Arguably, the most striking result of this paper is that first-best time-dependent road pricing in this context leads to a lower density, and hence increased urban sprawl, even without redistributing the collected road toll revenues back to the city inhabitants. The intuition is that first-best time-dependent road pricing induces drivers to spend more time at home, as road tolling eliminates queuing time. More time spent at home provides stronger incentives for having a larger house, and thus the city expands. In a similar fashion, an expansion of bottleneck capacity leads to the same effect. Our outcome is the opposite of the typical result of spatial urban models with static flow congestion. There, the Pigouvian toll, the typical first-best remedy for the negative congestion externality, increases the generalized transportation costs, and therefore reduces the city's geographical extent (Anas, Arnott, and Small 1998). Nevertheless, our result is consistent with other more familiar results in the sense that improved transportation causes urban sprawl.

There are few economics papers that model dynamic congestion in space. Arnott and DePalma (2011) report on their progress to solve the so-called "corridor problem", in which monocentric city inhabitants might experience congestion at each point on the road, as opposed to the single point congestion considered in our paper. Finding a complete solution to the dynamic equilibrium of flow congestion without a toll appears prohibitively difficult, even in a setting with an exogenously distributed population. A paper by Arnott (1998) is the only one that considers both dynamic congestion in the form of Vickrey's bottleneck and a (discrete) locational choice endogenously. The specified preference for the lot size though does not relate to the scheduling behavior and, as noted by Fosgerau and de Palma (2012), space in the model by Arnott (1998) has been "essentially [...] assumed away". Fosgerau and de Palma (2012) consider a continuous space city with a central Vickrey bottleneck, with time-varying

marginal utilities of spending time at home and at work and without considering spatial equilibrium. Contrary to what we will find, they show that inhabitants located near the bottleneck tend to lose from optimal pricing. To the best of our knowledge, our study is the first to consider both locational choice (that leads to commuting) and scheduling choice (that affects residential location) of city inhabitants endogenously.

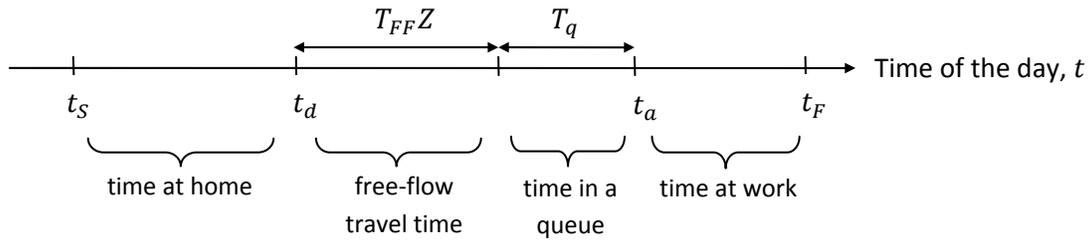
The paper is organized as follows. Section 2 presents the model setup, and Section 3 discusses how to find the market equilibrium. Section 4 then shows the resulting equilibrium patterns of land consumption, population density and rents over space; as well as the dynamic travel patterns by location. Section 5 considers first-best time-varying road pricing. Section 6 checks the robustness of the numerical outcomes by performing sensitivity analyses. Section 7 concludes.

## 2. Model setup

Consider a closed linear city of  $n$  homogeneous, atomistic, car owning, utility-maximizing inhabitants. All inhabitants earn an identical wage  $m$  in a spaceless central business district, hereafter the CBD, which is located in the city centre. We conveniently employ a “half-a-linear-city” spatial setting, with the CBD being at the spatial edge of our single dimension, due to the linear city’s symmetry. The inhabitants, with a common preferred time of arrival at work  $t^*$ , commute in the morning by car from their homes to the CBD, on a single road at a constant free-flow speed. At the entrance of the CBD, the road has a traffic bottleneck with a fixed capacity  $s$ . When the inflow of cars at any moment exceeds  $s$ , a “first-in first-out” traffic jam builds up in front of the CBD. Without queuing, the total commuting (free-flow) travel time of a driver is proportional to the distance from her home to the CBD, as in a conventional monocentric model. With queuing, the total travel time of a driver is the sum of her free-flow travel time to the CBD, plus the waiting time that she incurs in the queue.

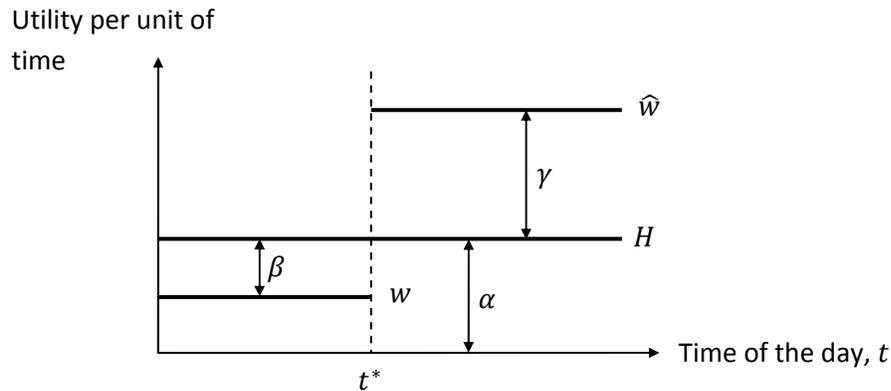
To model scheduling behavior, we apply the model proposed by Vickrey (1973), and later by Tseng and Verhoef (2008) and Fosgerau and Engelson (2011), in which, over the course of the morning, travelers derive utility from being at home and being at work. The morning starts for everyone at a common time  $t_S$ , chosen arbitrarily other than such that everyone will depart from home to work later than  $t_S$ . At a certain time, a driver departs from home to work and drives at a free-flow speed to the CBD. Different locations of houses within a city imply different distances  $Z$  to the CBD and, therewith, different free-flow travel times  $T_{FF} \cdot Z$ , where  $T_{FF}$  is the time needed to cover one unit of distance ( $Z = 0$  denotes the location of the CBD). A driver arrives at the entrance of the CBD, and, depending on the length of a queue, if any, waits  $T_q$  minutes to pass the bottleneck. After passing the bottleneck, the driver immediately arrives at work. The morning finishes for the drivers at the common time  $t_F$ , which is again chosen such that all drivers are at the workplace at that time. Denote a driver’s departure time from home as  $t_d$ , and the arrival at work as  $t_a$ . Throughout the paper, we measure clock time as time before  $t_F$ , which itself we set at  $t_F = 0$ . A higher time value thus indicates an earlier moment; i.e.,  $t_S \geq t_d \geq t_a \geq t_F = 0$ . Figure 1 shows an example of a morning schedule.

**Figure 1.** Example of a morning schedule



We explicitly define the marginal utilities of being at home, at work, and in a vehicle, so as to characterize the time pattern of activities in terms of utilities and opportunity costs. This approach is practical here, as it will later allow connecting the monocentric and the conventional bottleneck models in a structured manner. This conventional bottleneck model has a constant value of travel delays ( $\alpha$ ), and constant values of schedule delay early ( $\beta$ ) and late ( $\gamma$ ). Tseng and Verhoef (2010) show that such preferences emerge when it is assumed that a driver  $i$ 's marginal utility of being at home,  $H_i$ , is constant over the period considered, and the marginal utility of being at workplace  $W(t)$  is piecewise constant with an upward jump at time  $t^*$ .<sup>2</sup> We assume that this is the case, with identical preferences over drivers. More precisely, we assume that  $W(t) = w < H_i$  before  $t^*$  and  $W(t) = \hat{w} > H_i$  after  $t^*$ . The utility of being in a vehicle,  $V$ , is set to zero without loss of generality.<sup>3</sup> This utility structure results in several time-invariant values of opportunity costs that are relevant for the further analysis. In particular,

**Figure 2.** Utility structure and opportunity costs of a driver in the Vickrey's dynamic bottleneck model



Based on Tseng and Verhoef (2008), Figure 2.

<sup>2</sup> We make  $H_i$  driver specific in anticipation of its endogeneity later on in the exposition.  $W$  will remain uniform across drivers throughout the paper.

<sup>3</sup> Strictly speaking,  $H$ ,  $W$  and  $V$  are the values of the marginal willingness to pay for being at home, at work and in a car. For brevity we call them utilities.

the opportunity cost of being in a vehicle is  $H_i - V = \alpha_i$ , which can be interpreted as a value of travel time in the standard linear scheduling model. Opportunity cost of being at work before  $t^*$  is  $H_i - w = \beta_i$ , while being at home after  $t^*$  is  $\widehat{w} - H_i = \gamma_i$ . The former is equivalent to a unit shadow price of schedule delay early, the latter to that of schedule delay late. We assume  $V < W(t)$  to rule out a situation in which a driver prefers to stay in the car after arriving at the CBD. Figure 2 depicts this utility structure.<sup>4</sup>

It is instructive to present briefly the basic derivations of a standard (spaceless) bottleneck model, in which  $Z = 0$  and  $H_i = H$  (for an in-depth discussion see Arnott, de Palma, and Lindsey 1993). A peak period of duration  $\frac{n}{s}$  is needed for all  $n$  drivers to pass the bottleneck. The very first and very last driver face no queuing delay and therefore have travel cost of, respectively,  $\beta(t^* - t_q) > 0$  and  $\gamma(t_{q'} - t^*) > 0$ , where  $t_q$  and  $t_{q'}$  are the beginning and the end of the peak period, and  $\frac{n}{s} = |t_{q'} - t_q|$ . In a dynamic equilibrium, all homogeneous drivers should incur the same generalized travel costs, irrespective of their arrival time  $t_a$ . That implies that the peak period starts at time  $t_q = t^* - \frac{\gamma}{\beta + \gamma} \frac{n}{s}$  and finishes at time  $t_{q'} = t^* + \frac{\beta}{\beta + \gamma} \frac{n}{s}$ , while the generalized travel cost is  $\frac{\gamma\beta}{\beta + \gamma} \frac{n}{s}$ . A driver who arrives at work at time  $t^*$ , incurs  $\alpha T_{qt^*}$  waiting time cost, where  $T_{qt^*} = \frac{\gamma\beta}{\beta + \gamma} \frac{n}{s} \frac{1}{\alpha}$ . Between  $t_q$  and  $t^*$  the queue grows linearly over time to equilibrate travel cost, while it shrinks linearly over time between  $t^*$  and  $t_{q'}$ . Thus, the bottleneck model highlights the trade off drivers make between waiting in a queue (travel delay cost) and arriving at the inconvenient time (schedule delay cost). This trade off remains present in our spatial model.

We simplify the derivations considerably by setting schedule delay late  $\gamma_i$  prohibitively high. That is, no driver would choose to arrive at work after the preferred arrival time  $t^*$ . One might think of a typical beginning of the business hours at 9:00 o'clock, which employees cannot miss without facing heavy costs. This allows us to set  $t^*$  equal to the end of the morning  $t_F$ , which simplifies subsequent derivations at the moderate cost of pre-defining the start of the peak period, i.e., the peak period starts at time  $\frac{n}{s}$  and finishes at time  $t^* = t_F = 0$ . One can relax this assumption, as it does not affect the mechanics of the model. We will consider qualitative outcomes of an extended model with late arrivals when discussing the results in Section 4.

A driver  $i$  derives Cobb-Douglas utility  $U$  from the way time is spent in the morning,  $X$ , and from consuming a numeraire good  $Q$ :

---

<sup>4</sup> Alternative functional specifications of marginal utilities are possible, for instance, time-varying downward sloping  $H$  and upward sloping  $W$  functions (see Figure 1 in Tseng and Verhoef (2008)), which result in smooth time allocation. For example, Fosgerau and de Palma (2011) show that under that specification marginal utility of the last minute of being at home is equal to the marginal utility of the first minute of being at work. While our approach allows various specifications, we chose to stay as close as possible with the conventional linear (" $\alpha$ - $\beta$ - $\gamma$ ") scheduling model, which is widely used in the literature.

$$(1) U_i = X_i^b Q_i^{1-b}$$

where  $X$  is the aggregated utility of time used at different locations, as it can be derived from the scheduling model used:

$$(2) X_i = (t_s - T_{FF}Z_i - T_q(t_{ai}) - t_{ai})H_i + t_{ai}w$$

The first term on the right hand side of equation (2) is the utility one derives from being at home in the morning, while the second term is the utility from being at work before  $t^*$ . All city inhabitants consume some amount of the numeraire good  $Q$  at a price 1, under the budget constraint:

$$(3) m = Q_i + r(Z_i)L(Z_i)$$

where  $L(Z_i)$  is the amount of land (a proxy for housing) that an inhabitant consumes at location  $Z_i$ , with the rent  $r(Z_i)$  that an inhabitant pays to an absentee landlord. The key assumption of our model is that consumption of land  $L(Z_i)$  affects the marginal utility of being at home  $H_i$  in a positive way. That is: a bigger house is assumed to bring higher utility, *ceteris paribus*, but we assume that this higher utility rises with the amount of time spent in the house. The probably simplest formulation incorporating this would be:

$$(4) H_i = w + \varepsilon + \varphi L(Z_i)$$

where  $\varepsilon$  is a parameter that keeps marginal utility of being at home above  $w$  even if  $L = 0$ , and  $\varphi$  is a parameter that “converts” amount of land consumed into marginal utility values for spending time at home.

**Assumption 1.** *Ceteris paribus*, the larger the house one consumes, the higher is one’s marginal utility of spending additional unit of time in it.

Assumption 1 is consistent with the observed preferences for larger housing, as shown by hedonic price studies (see, e.g., Bajari and Kahn 2008). We add to that a time aspect, by assuming that an individual may derive more utility from a house by spending more time in it. While a house might yield utility via various channels – e.g., as a storage place, a status symbol, or an investment tool – we believe it is safe to assume that at least some part of the utility from housing depends on the amount of time an individual spends there. For our purposes, this is the only component of utility from housing that we need to (and will) specify.

We allow for endogenous variation in consumption of land. Rents are endogenously determined, to insure that inhabitants are indifferent between various locations within a city. That is, utility should be constant over space for a spatial equilibrium to be fulfilled. To close the model, we normalize the agricultural rent at the city fringe at 1.

The usual monocentric city model has an increasing consumption of land the further one lives from the CBD. Consequently, population density, which is the inverse of land consumption decreases.<sup>5</sup> Endogenous rents are also decreasing with distance, to compensate for the increase of the transportation cost. Note that with these patterns, Assumption 1 brings variation in scheduling preferences across drivers. Equation (4) links location  $Z_i$  with the marginal utility of being at home  $H_i$ , and hence with the value of travel time  $\alpha_i$  and schedule delay early  $\beta_i$ . Depending on the size of the house, a driver chooses a certain departure time from home and vice versa. The model thus produces endogenous preferences in terms of scheduling (dis)utilities, and hence an endogenous ordering of travelers over time during the peak, and a dynamic equilibrium that accounts for this (endogenous) heterogeneity.

In our model, the arrival time at work  $t_a$ , the location of housing  $Z$ , and the consumption of land  $L$  given  $Z$ , are the three margins of behavior, with the rents endogenously determined.<sup>6</sup> A driver  $i$  then maximizes utility  $U$ , which we may rewrite from equation (1) as:

$$(5) \quad U_i = \left( (t_S - T_{FF}Z_i - T_q(t_{ai}) - t_{ai})(w + \varepsilon + \varphi L(Z_i)) + t_{ai}w \right)^b (m - r(Z_i)L(Z_i))^{1-b}$$

---

<sup>5</sup> Inclusion of the travel related costs (both free-flow travel time and congestion costs) into the utility in terms of time in our model is a departure from a standard static monocentric model specification, where pecuniary travel costs enter directly into a budget constraint. Nevertheless, the qualitative results concerning land consumption and the shape of the bid rent curve are the same, as we show later.

<sup>6</sup> To maintain the focus of the paper on the relationship between time allocation vs. spatial behavior we keep (implicit) wages fixed and homogeneous in order to prevent too many factors affecting equilibrium. Our model shows the effect of these two “margins” (dynamic and space) while keeping other possible variations “fixed”.

### 3. Solution of the model

#### 3.1 Solution in the nutshell

In our model, both dynamic and spatial equilibria conditions apply, which require that (i) no traveler can gain from unilaterally changing departure time, and (ii) utility within the city boundaries is constant. In the modeling of endogenous choices of location, land consumption, and arrival times, we consider utility maximization as a sequence of a short-run, a medium-run and a long-run optimization problem. In the short-run equilibrium, in which consumption of land over space and rents are fixed, an individual maximizes utility by choosing the time of arrival at work. This short-run setup seems reasonable, as usually travel time adjustments can be made at shorter notice than changes in house size and location. We analytically derive an arrival pattern, and corresponding generalized congestion cost, that fulfills the dynamic equilibrium condition, and that applies given the spatial equilibrium. In the medium run, assuming fixed rents, but taking into account that the dynamic commuting timing equilibrium will adjust with the spatial equilibrium, we analytically derive the spatial pattern of land consumption. It satisfies the land market condition, which implies that everything else kept constant, one cannot gain in utility by changing land consumption. Finally, to determine a fully endogenous long-run equilibrium, we compute rents that satisfy the spatial equilibrium condition, given the assumed land consumption and a corresponding short-run equilibrium arrival pattern. Then, we plug in these rents into an optimal land consumption function, as defined by the medium-run solution, which, in turn, affects the equilibrium arrival pattern. The long-run equilibrium is then found by repeatedly substituting the spatial pattern of land consumption, and computing the arrivals and rents, until convergence is reached. The convergence criterion is the equality of the initially assumed city population to the overall number of inhabitants living within the endogenously defined city boundaries, as calculated by the spatial integral of the population density over the entire city, where population density is the inverse of a land consumption. As an ultimate check of reaching an equilibrium, we compute equilibrium utility levels over space, which should be constant within the city.

#### 3.2 Short-run equilibrium

In the short-run equilibrium, times of arrival at work are determined. Consumption of land and rents are fixed, meaning that non-time component of the utility function is unaffected by changes in times of arrival. For a person to be indifferent for a marginal change in arrival time  $t_a$ , we must have that at the moment of arriving,  $\frac{\partial X}{\partial t_a} = 0$ . From equation (5) it follows that:

$$(6) \quad \frac{\partial X}{\partial t_a} = -\frac{\partial T_q}{\partial t_a} (w + \varepsilon + \varphi L) - (w + \varepsilon + \varphi L) + w = 0$$

From (6), we derive the time derivative of the equilibrium growth rate of the travel delays as:

$$(7) \frac{\partial T_q}{\partial t_a} = \frac{\varepsilon + \varphi L}{w + \varepsilon + \varphi L}$$

Equation (7) has a close correspondence with the conventional equilibrium in the standard bottleneck model, where the growth rate of travel delays during early arrivals at work is  $\frac{\beta}{\alpha}$ .

It is easy to show that with heterogeneous preferences, the equilibrium queue growth cannot decrease over time during the period of early arrivals (e.g., Small and Verhoef 2007, p. 132). We can interpret equation (7) as the slope of an isocost line for individuals traveling at that instant. Drivers with heterogeneous values of travel time and scheduling delays would then arrive at the moments which bring them to the lowest achievable isocost level. The queuing time function  $T_q(t_a)$  should then be convex (meaning a non-decreasing queuing growth), to make sure drivers are unable to achieve a lower cost level by rescheduling.

It thus follows from equation (7), that arrivals at work will be in order of increasing land consumption. The “first-in first-out” queuing discipline then implies the same ordering of arrivals at the back of the queue. The order of departure from home, however, also depends on the free-flow travel time.

With the order of passing the bottleneck being determined, we can then express generalized congestion cost (including travel delay cost and schedule delay cost, but excluding free-flow travel time cost) in a way similar to van den Berg and Verhoef (2011). To this end, observe that the peak period starts at  $\frac{n}{s}$ , and that in case of homogeneous drivers with identical scheduling preferences, the average (per user) congestion cost would be equal to  $\frac{n}{s}(\varepsilon + \varphi L)$ . This corresponds to the generalized congestion cost from the standard bottleneck model with early arrivals only,  $\frac{n}{s}\beta$ . Equation (8) defines a short-run equilibrium congestion cost  $C_i$  for a driver  $i$  when drivers have heterogeneous scheduling preferences (in particular when these preferences depend on the amount of land consumed,  $L_i$ ):

$$(8) C_i = \frac{\theta_i}{s}(\varepsilon + \varphi L_i)$$

where  $\theta_i$  is the (weighted) number of drivers that would bring a driver  $i$ , with given scheduling preferences, at the same isocost level as where she is now, if all other drivers had the same scheduling preferences as driver  $i$  has herself:

$$(9) \theta(L_i) = \int_{\underline{L}}^{L_i} \frac{w + \varepsilon + \varphi L_i}{\varepsilon + \varphi L_i} \frac{w + \varepsilon + \varphi l}{\varepsilon + \varphi l} \frac{1}{l} dl + \int_{L_i}^{\bar{L}} \frac{1}{l} dl$$

where  $\underline{L}$  and  $\bar{L}$  are the minimum and maximum levels of land consumption within the city. Equation (9) shows that a driver who arrives closer to  $t^*$  gains most from heterogeneity in scheduling preferences. The reason is that due to the convexity of the travel delay function  $T_q(t_a)$ , the negative congestion externality that earlier drivers impose on her is smaller than the negative externality that would result from drivers with the same scheduling preferences (who would make the travel delay function linear instead of convex).

Given the short-run trip timing equilibrium, we can rewrite equation (2) as follows:

$$(10) X_i = (t_S - T_{FF}Z_i)(w + \varepsilon + \varphi L_i) - \frac{\theta(L_i)}{s} (\varepsilon + \varphi L_i)$$

The first term on the right hand side of equation (10) is equal to an “ideal” utility level, which a driver would reach over the course of the morning, had she arrived at the CBD at time  $t^*$  without facing traffic jam. The second term is the congestion cost, given the trip timing equilibrium just described.

### 3.3 Medium-run equilibrium

For the medium-run problem, in which an inhabitant maximizes utility by adjusting land consumption, given the pattern of land rent, prices and accounting for the effect of land use on trip timing, we assume that inhabitants cannot change their locations. We also assume that an equilibrium congestion cost derived from the short-run equilibrium problem holds, but we treat  $L$  as a variable argument in that function, to reflect that individuals know they will adjust travel moments when adjusting land consumption. Then, we can specify the utility function and a budget constraint as follows:

$$(11) U_i = \left( (t_S - T_{FF}Z_i)(w + \varepsilon + \varphi L_i) - \frac{\theta(L_i)}{s} (\varepsilon + \varphi L_i) \right)^b Q_i^{1-b}$$

$$(12) m = Q + r_i L_i$$

We maximize utility by defining the Lagrangian and solving the usual first-order conditions. After solving for the optimum land consumption, we get:

$$(13) L_i = \frac{(b-1)(\theta\varepsilon - s(t_S - T_{FF}Z_i)(w+\varepsilon))}{\varphi(\theta - s(t_S - T_{FF}Z_i))} + \frac{bm}{r_i}$$

Note that when congestion costs are zero, i.e.,  $\theta = 0$  (that would correspond to a model with space-varying free-flow travel time, but without a bottleneck), the optimal consumption of land would be:

$$(14) L_i = \frac{(b-1)(w+\varepsilon)}{\varphi} + \frac{bm}{r_i}$$

The second term on the right hand side of both equations (13) and (14) is the regular conditional demand for land, for a standard Cobb-Douglas utility specification in a static monocentric city model. The first term on the right hand side of equation (14) is a negative correction term, which accounts for the fact that not all utility from housing comes from the size of the house. The stronger the size effect is (a small  $(w + \varepsilon)/\varphi$  ratio), the smaller (in absolute terms) the term is. But, if the size of the house adds little to housing utility, the term gets larger (in absolute terms), and  $L$  is eventually bounded by the non-negativity constraint.

The negative correction term in equation (13) differs from that of equation (14) because in our model, some part of utility from housing is forgone due to the congestion. That is, housing now yields lower utility than with free-flow travel time only, because a driver spends time in a queue and/or early at work, instead of spending time at home. It is easy to show that the correction term is, not surprisingly, larger (in absolute terms) when congestion is stronger, resulting in a weaker demand for land.

### 3.4 Long-run equilibrium

With the long-run equilibrium, we refer to the complete solution of the model in which land rents have adjusted over space in such a way that in the resulting medium- and short-term equilibrium, utility is constant over space within the city. We use numerical methods to compute this long-run equilibrium, as an insightful analytical solution seems unattainable. Before turning to the calibration of our numerical model, we define several functions which we need for the long-run analysis.

First, note that a pattern of land consumption over space results in a population density,  $(\frac{1}{L})$ , that, in turn, determines an endogenous city boundary,  $\bar{Z}$ , because for a closed city it must be true that:

$$(15) n = \int_0^{\bar{Z}} \frac{1}{L(z)} dz$$

Equation (15) will serve as a convergence criterion for our numerical computation, which we define below.

We restrict attention to the empirically relevant case with increasing land consumption, at greater distances from the CBD. Indeed, it is conceivable that the model might support also the opposite spatial pattern of land consumption, where land consumption decreases with distance if those who live closest to the CBD are the last to depart from home in the morning and therefore consume more housing than those near the fringe. But we do not consider this potential type of equilibrium worthwhile for analysis. Thus, as the larger houses are located closer to the city fringe, the order of arrival can be defined in function of both land consumption and location, and we can change the variable of integration in equation (9) from  $l$  to  $z$  as follows:

$$(16) \theta_i(Z_i) = \int_0^{Z_i} \frac{w + \varepsilon + \varphi L(Z_i)}{\varepsilon + \varphi L(Z_i)} \frac{w + \varepsilon + \varphi l(z)}{\varepsilon + \varphi l(z)} \frac{1}{l(z)} dz + \int_{Z_i}^{\bar{Z}} \frac{1}{l(z)} dz$$

Equation (16) implies that each particular location may and will have its own optimal consumption of land, and hence scheduling preferences. We then use equations (11) and (12) to rewrite utility from equation (5) as a function of location  $Z$  only:

$$(17) U(Z_i) = \left( (t_S - T_{FF}Z_i)(w + \varepsilon + \varphi L(Z_i)) - \frac{\theta(Z_i)}{s}(\varepsilon + \varphi L(Z_i)) \right)^b (m - r(Z_i)L(Z_i))^{1-b}$$

In the calibration we assume that the morning period lasts for 30 units of time, i.e.,  $t_S = 30$ ; the free-flow travel time needed to cover one unit of distance is normalized to 1, i.e.,  $T_{FF} = 1$ ; the marginal utility of being at work before  $t^*$  is normalized to 1, i.e.,  $w = 1$ ; the parameter that keeps marginal utility of being at home above  $w$  is 0.1, i.e.,  $\varepsilon = 0.1$ ; the parameter that “converts” amount of consumed land into marginal utility of being at home is equal to 2, i.e.,  $\varphi = 2$ ; the number of people in the city is 10, i.e.,  $n = 10$ ; the capacity of the bottleneck is normalized to 1, i.e.,  $s = 1$ ; the exponent for ordinary goods in the Cobb-Douglas utility is 0.6, i.e.,  $b = 0.4$ ; and the wage is 4, i.e.,  $m = 4$ . These parameter values were chosen such that they yield what we believe are reasonable equilibrium outcomes. For instance, empirical values for the scheduling parameters suggest that their relative values approximately satisfy  $\alpha = 2\beta$  (see Small 1982). While our model results in the distribution of those values, we try to stay around that ratio for the average values. A bit more than two fifth of the morning period is characterized with a positive commuting flows (around 2.5 hours, if the morning lasts from 6 to 12). Sensitivity analyses have demonstrated the robustness of the main results with respect to these parameter values.

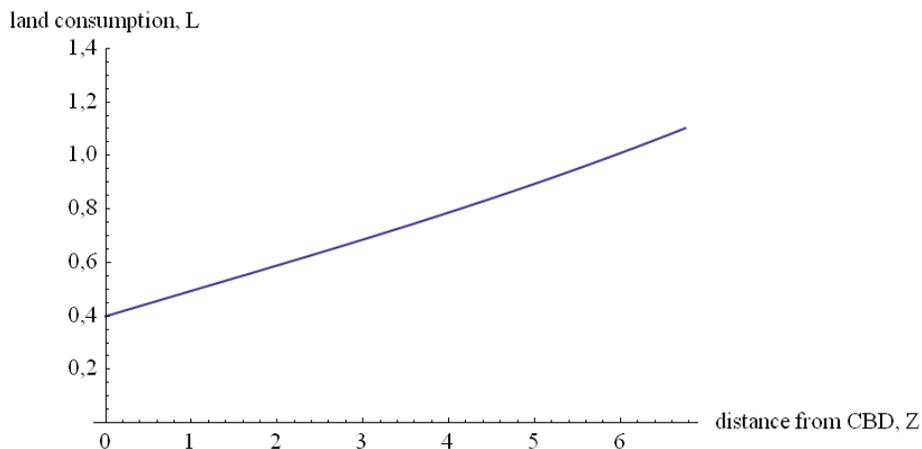
The numerical model achieves convergence by starting with some arbitrary pattern of land consumption (in our base case that is  $L = 1 + \frac{1}{100}Z$ ), and next finding a trip-timing (short-run) equilibrium. Equation (15) defines a distance from the CBD to the city fringe ( $\bar{Z} = 10.52$ ). We then plug in that value into equation (16) to get a short-run trip timing equilibrium. Next, we take partial derivative of the utility function from equation (17) with respect to location and equate it to zero by varying land rents. That derivative yields a differential equation for the equilibrium rent function, which we solve numerically while applying the equilibrium boundary condition  $r(\bar{Z}) = 1$ . After solving for the equilibrium rent, we plug these rents values into equation (13) to get an equilibrium land consumption. We then calculate equation (15) again to check whether under newly found land consumption, the city boundary is within the convergence criterion from the value at the start of the iteration. If it is not, then we calculate a new boundary, and repeat the computations. The convergence is reached when the difference of the boundary values between two iterations is less than  $10^{-8}$ . This corresponds to less than  $10^{-8}$  percent of the equilibrium city size, and in an equilibrium in which the population  $n$  is less than  $10^{-9}$  percent away from the target value of 10.

#### 4. Unpriced market equilibrium

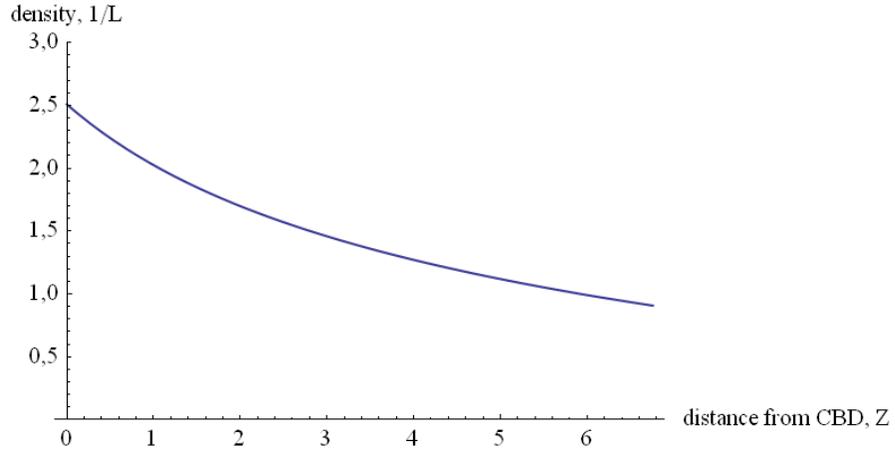
Figures 3 – 5 show equilibrium base case patterns of, respectively, land consumption, population density and rents over space, for the base calibration of the numerical model. At the city fringe the rent is 1, and the integral of population over the city territory is 10, which is consistent with our initial assumptions. The upward sloping consumption of land function, and downward sloping rent function, are of course consistent with patterns which are usually found for the static monocentric city models and found in empirical literature (e.g., Glaeser 2008).

Figure 6 shows the dynamic equilibrium travel time patterns by residential location. We give an intuitive description of the morning commute first, and then turn to the details of Figure 6.

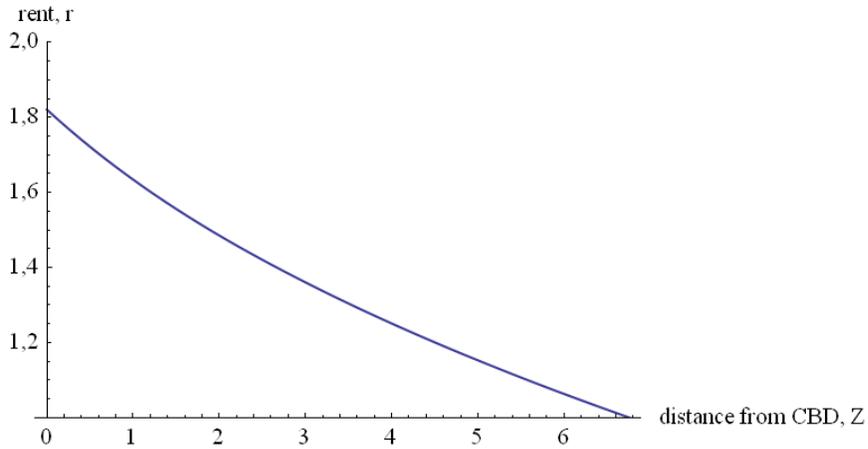
**Figure 3.** Consumption of land by residential location (base equilibrium)



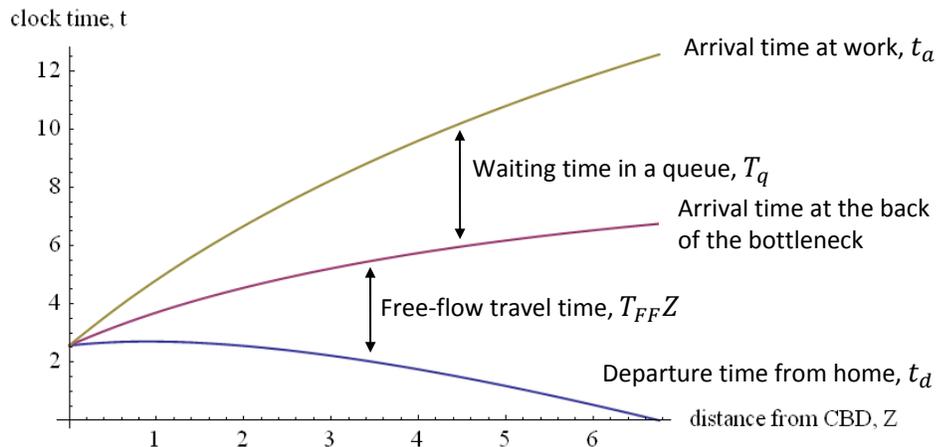
**Figure 4.** Population density by residential location (base equilibrium)



**Figure 5.** Rents by residential location (base equilibrium)



**Figure 6.** Travel patterns by residential location in an unpriced market equilibrium (base equilibrium)



The morning commute in our base case is as follows. Although arriving last, just before  $t^*$ , the inhabitant who lives at the city fringe is the first to depart from home. With the progress of clock time, drivers depart from home in order of distance from the city fringe. This dynamic pattern of departure moments depends on the free-flow travel speed with which drivers cover the distance between their residence and the CBD, as in a market equilibrium they pass the bottleneck in order of the relative size of their houses. That is, the driver with the smallest house (living right next to the CBD) is the first to pass the bottleneck, and the one with the largest house (at the fringe) is the last one. All drivers, except the first to pass the bottleneck, incur some positive waiting time. The waiting time is reflected in Figure 6 by the difference between times of arrival at the back of the queue, and at work. The queuing delay grows monotonically during the peak, reaching a maximum for the last arrival, at  $t^*$ . We assume that the entire morning lasts 30 units of time, and our model results in an endogenous period of 12.57 units of time with a positive commuting flow (i.e., during the other 17.43 units of time everyone is at home). The traffic jam itself lasts 10 units of time, as  $\frac{n}{s} = 10$ . The endogenous city boundary is at  $\bar{Z} = 6.76$ , and the utility level across the city equals 9.58.

Keeping in mind the spatial order of arrivals at work – from locations closest to the CBD to the ones further away – it is straightforward to derive the spatial pattern of arrival times  $t_a(Z)$ . This is a matter of taking the spatial integral of the population density over bottleneck capacity. We compute an equilibrium spatial pattern of waiting time  $T_q(Z)$  by rewriting equation (7):

$$(18) \quad \frac{\partial T_q}{\partial Z} = \frac{\varepsilon + \varphi L}{w + \varepsilon + \varphi L} / \frac{\partial Z}{\partial t_a}$$

We solve numerically this differential equation by computing an inverse function  $Z(t_a)$ , and setting  $T_q(0) = 0$ , as we know that a driver who lives right next to the CBD is the first to pass the bottleneck and she will face no waiting time. Knowing the arrival times  $t_a(Z)$  and the waiting times  $T_q(Z)$ , it is straightforward to calculate the departure times  $t_d(Z)$ , as shown in Figure 1.

The outcomes of the model extended to include late arrivals are qualitatively similar to the ones shown here. The order of arrival at work after  $t^*$  would follow the same spatial pattern as the one before  $t^*$ . That is, an inhabitant with the smallest house is the first to arrive at work after  $t^*$ , the one with the largest house is the last. In other words, the arrival pattern for late arrivals is not a mirror image of the pattern for early arrivals. To see this, observe that the inhabitants with the small houses have a large schedule delay cost of being late ( $\gamma$ ), as they value time spent at home at a lower rate and thus each minute of being late means a relatively large opportunity cost (to remind,  $\gamma_i = \hat{w} - H_i$ ). In contrast, those with the large houses incur relatively low schedule delay cost of being late. In equilibrium, the reduction rate of queuing time should be decreasing for late arrivals to accommodate drivers with heterogeneous scheduling preferences. The moment of arrival that brings an individual with the smallest house to the lowest cost level is the moment right after  $t^*$ ; hence, the pattern of arrivals repeats itself.

Note that at each given location, there will then be two moments of departure, one for the early arrival, and the one for later. The share of drivers at each location that arrive early versus late in equilibrium should be such that the utility levels of a person who arrives after  $t^*$  from a given location is equal to the utility level of the one who arrives before  $t^*$  from the same location. Computationally, this appears to be a rather complex problem, not in the least place because the timing of the peak will become endogenous, whereas we do not foresee any large implications for the qualitative outcomes of the model. We therefore refrained from this more elaborate setting, and chose to use the model with early arrivals only.

## 5. City with the first-best road pricing

### 5.1 Model with the road pricing

The literature on the standard Vickrey dynamic bottleneck model shows that waiting time in a queue, and thus waiting time costs  $\alpha T_q$ , is a pure welfare loss, that a first-best time-dependent road toll fully eliminates. In the case of homogeneous drivers this is achieved by substituting (at each moment of arrival) the waiting time cost from the unpriced market equilibrium with an equally large road toll (see, e.g., Vickrey 1969; Arnott, de Palma, and Lindsey 1993). Although the generalized travel price stays the same as in a no-toll equilibrium, from the social point of view a road toll is not a cost, but a welfare neutral monetary transfer from the drivers to the government. Thus, road pricing reduces an overall welfare loss due to congestion. The schedule delay cost remaining in the optimum,, and it is not affected by road pricing if drivers are homogeneous. Arnott, de Palma, and Lindsey (1993) offer further discussion of the standard Vickrey's model with pricing.

We are interested how the results of our spatial model change under first-best road pricing, compared to the no-toll equilibrium outcomes. To this end, we repeat the analysis presented above while accounting for first-best time-dependent toll  $\tau(t_a)$ . The setup of the model slightly changes, as the drivers under first-best road pricing no longer face waiting time, i.e.,  $T_q = 0$ , so the time of arrival at the back of the bottleneck is identical to the time of arrival at work. We can rewrite the utility function from equation (5) and the budget constraint from equation (3) as follows to account for pricing:

$$(19) U_i = ((t_S - T_{FF}Z_i)(w + \varepsilon + \varphi L_i) - t_{ai}(\varphi L_i + \varepsilon))^b Q_i^{1-b}$$

$$(20) m + \chi = Q_i + r(Z_i)L_i + \tau(t_{ai})$$

where  $\chi$  represents a lump-sum monetary transfer from the government back to the drivers, which is the average (per driver) of the total road toll revenues. Under a no-toll equilibrium, both the road toll  $\tau$  and the lump-sum transfer  $\chi$  are of course zero by definition.

As in the no-toll setting, location, land consumption, rents and arrival times at work are endogenous. We endogenize a road toll as well – it should satisfy the dynamic equilibrium condition that no driver has an incentive to unilaterally change her departure time to gain in utility. Since there is no waiting time, the departure times from home are then derived by subtracting the free-flow travel times from the arrival times at work.

## 5.2 Solution of the model with road pricing

We again distinguish between a short, a medium and a long-run optimization problem. In the short run, in which a driver chooses time of arrival at work while keeping everything else constant, we determine the optimal toll  $\tau(t_a)$  that decentralizes the queue-free optimum as a dynamic equilibrium. In contrast to the no-toll case, a change in the time of arrival now affects the monetary component of the utility function, as different moments of arrival correspond to different road tolls, and hence to different remaining (after-toll) budgets. An equilibrium road toll reflects the trade-off that individuals make between the opportunity cost of an early arrival versus the money one has to pay, and therewith, the forgone consumption of the numeraire good – as the consumption of land is fixed in the short-run. We can rewrite the necessary marginal condition for an individual's optimized arrival time as follows:

$$(21) \frac{\partial U}{\partial t_a} = - \frac{\partial U}{\partial Q} \frac{\partial Q}{\partial t_a}$$

Since the price of a numeraire good is 1,  $\frac{\partial Q}{\partial t_a}$  directly shows the equilibrium shadow price of a marginal change in arrival time in terms of foregone consumption of the numeraire good, as the budget constraint (20) implies:

$$(22) \frac{\partial \tau}{\partial t_a} = \frac{\partial Q}{\partial t_a}$$

Applying equilibrium condition (21) and substituting (19) into it leads to the following condition on the slope of the toll schedule at the moment of travelling:

$$(23) \frac{\partial \tau}{\partial t_a} = \frac{bQ(\varepsilon + \varphi L)}{(1-b)((A-T_{FFZ})(w + \varepsilon + \varphi L) - t_a(\varepsilon + \varphi L))}$$

or, after simplifying (23):

$$(24) \frac{\partial \tau}{\partial t_a} = (\varepsilon + \varphi L) \frac{b}{(1-b)} \frac{Q}{X}$$

In the standard bottleneck model with homogeneous drivers, the growth rate of the optimal toll for each arrivals is  $\beta$ , which has the equivalent of  $H - w = \varepsilon + \varphi L$  in our model. Intuitively, for each minute of arriving closer to the preferred arrival time  $t^*$ , the scheduling cost decreases by  $\beta$ , so to keep the generalized cost constant over the peak period (as required in the dynamic equilibrium) without queuing, that utility gain is compensated for by an equally large increase in the monetary toll. Our result differs because, unlike in the spaceless basic bottleneck model, in a monocentric city the marginal utility of income varies over space when the income elasticity of land demand is non-zero, as shown, for example, by Mirrlees (1972) and Wildasin (1986). The closer to  $t^*$  a driver arrives at work, the more toll she has to pay and thus the more valuable money is (under positive income elasticity of demand for land). The optimal toll schedule implied by (24) accounts for this.

Analogously to the no-toll scenario, a toll function  $\tau(t_a)$  should be convex to support an equilibrium under heterogeneous scheduling preferences. This is achieved if the drivers arrive at the CBD in order of increasing land consumption, as in the no-toll case. Given this order of arrival, we may rewrite equations (8) and (9) and derive the first-best generalized price for road use  $C_i^{FB}$ :

$$(25) C_i^{FB} = \frac{\theta_i^{FB}}{s} (\varepsilon + \varphi L_i) \frac{b}{(1-b)} \frac{Q}{X}$$

where  $\theta_i^{FB}$  is the (weighted) number of drivers that would bring a driver  $i$ , with given scheduling preferences, at the same isoprice level where she is now, if all other drivers had the same scheduling preferences as a driver  $i$  herself. Following van den Berg and Verhoef (2011),  $\theta_i^{FB}$  is:

$$(26) \theta_i^{FB} = \int_{\underline{L}}^{L_i} \left( (\varepsilon + \varphi l) \frac{b}{(1-b)} \frac{Q(l)}{X(l)} \right) / \left( (\varepsilon + \varphi L) \frac{b}{(1-b)} \frac{Q(L)}{X(L)} \right) \frac{1}{l} dl + \int_{L_i}^{\bar{L}} \frac{1}{l} dl$$

We apply a geometric argument and derive the optimal road toll<sup>7</sup>:

$$(27) \tau_i = \left( \frac{\theta_i^{FB}}{s} - t_a \right) (\varepsilon + \varphi L) \frac{b}{(1-b)} \frac{Q}{X}$$

The lump-sum transfer from the government to the drivers is then the average (per driver) of the overall toll revenues collected:

$$(28) \chi = \frac{1}{n} \int_0^{\bar{Z}} \frac{\tau}{L(z)} dz$$

In the medium-run we solve for the optimal land consumption, while taking rents and location as given, but accounting for the effect that the size of one's house will affect one's position in the order of passing the bottleneck. That is, the larger the house, *ceteris paribus*, the closer to the preferred arrival time one will reach the CBD in the short-run equilibrium. We specify a Lagrangian using equations (19) and (20). Solving the first-order conditions, we get closed-form solutions for the optimal consumption of land  $L^*(r(Z), Z)$  and numeraire good  $Q^*(r(Z), Z)$ .

In the long-run equilibrium, in which the equilibrium rents are determined, we follow the heuristic numerical procedure as described for the no-toll case. Again, we only consider the empirically relevant case of  $\frac{\partial L}{\partial Z} > 0$ . We start the heuristic by defining an arbitrary consumption pattern of land (in our calibration model:  $L = 1 + \frac{1}{100} Z$ ) and numeraire good ( $Q = 3 + \frac{1}{100} Z$ ). Using equation (15), we next find a distance to the city fringe. Arrival time at work for a given driver is then equal to the moment of the start of the peak period (i.e.,  $\frac{n}{s}$ ) minus the time needed for all drivers who live closer to the CBD to pass the bottleneck:

$$(27) t_{ai} = \frac{n}{s} - \frac{1}{s} \int_0^{Z_i} \frac{1}{L(z)} dz$$

To solve the long-run equilibrium, we change the variable of integration in equation (26) from  $L$  to  $Z$  (as we did in the no-toll case), and find the first-best road toll implied by the short-run equilibrium using

---

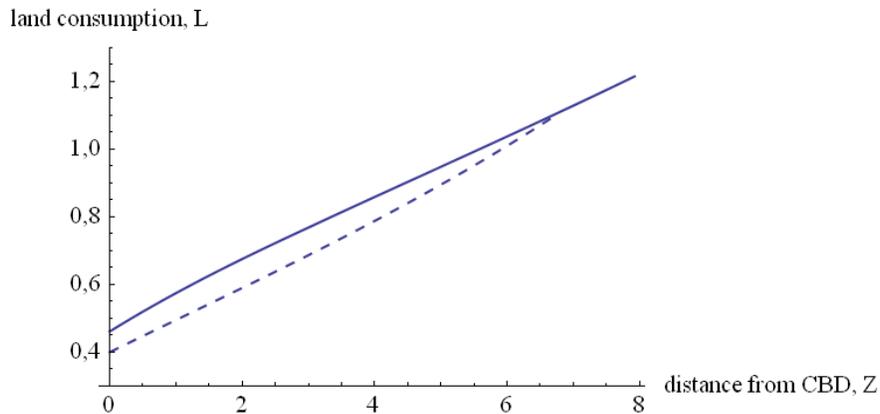
<sup>7</sup> The road toll at each moment of arrival can be found by multiplying the slope of the isocost function as given in (24) by the base (shown in the left bracket of (27)) which is the time difference between given arrival time and the start of the peak period, if all drivers would have the same scheduling preferences as an individual who arrives at the given moment.

equation (27) and a lump-sum transfer from equation (28). We next take the partial derivative of the indirect utility function with respect to distance and set it to zero. From the resulting differential equation, we numerically find an equilibrium rent function  $r^*$  which satisfies the boundary condition  $r(\bar{Z}) = 1$ . We then plug  $r^*$  into the functions of  $L^*$  and  $Q^*$ , to find the optimal consumption of land and of the numeraire good. We check for the distance to the city fringe, and if  $\bar{Z}$  is outside the convergence criterion (as described in subsection 3.4) from the value at the start of iteration, we repeat the computations, until convergence is reached.

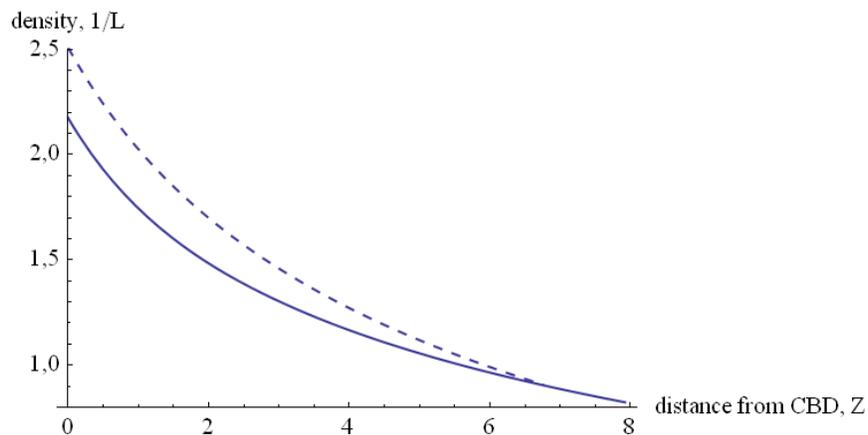
### 5.3 Equilibrium with road pricing

Figures 7 – 9 show (with solid lines) the spatial patterns of land consumption, population density and rents for the reported calibration parameters, when the first-best time-dependent road toll is levied in

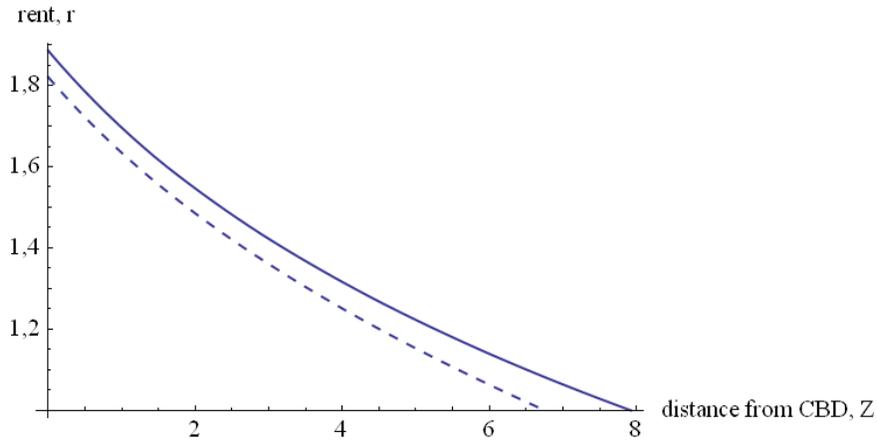
**Figure 7.** Consumption of land by residential location [with pricing (solid) and without (dashed)]



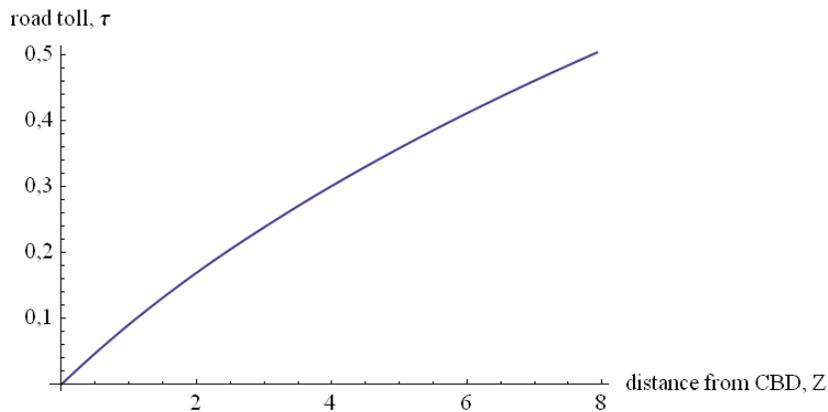
**Figure 8.** Population density by residential location [with pricing (solid) and without (dashed)]



**Figure 9.** Rents by residential location [with pricing (solid) and without (dashed)]



**Figure 10.** Road toll by residential location

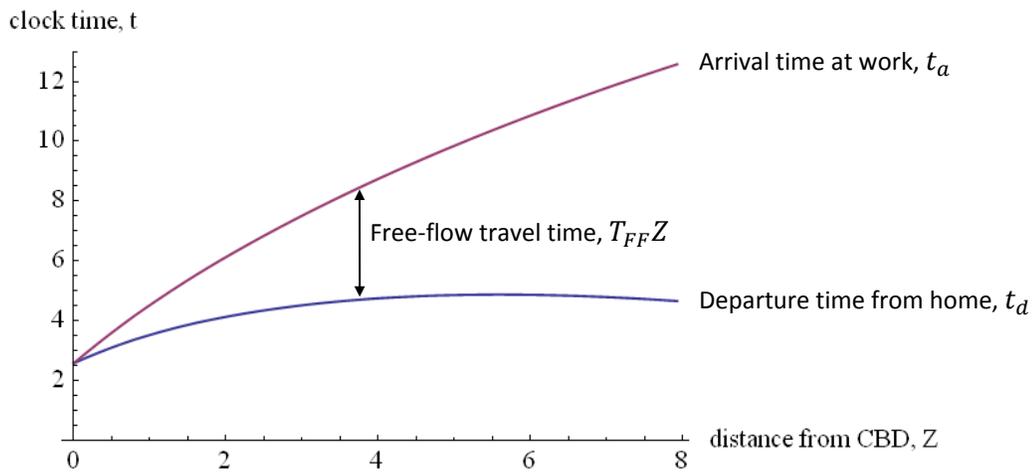


the monocentric city. For comparison, the dashed lines show the corresponding spatial patterns in the no-toll market equilibrium. As in the no-toll case, the model outcomes have been checked to satisfy our initial assumptions of the agricultural rent  $r(\bar{Z}) = 1$  and the total population size  $n = 10$ . Figure 10 shows the first-best road toll plotted against the location within the city, as each particular location corresponds to a certain arrival time. The lump-sum transfer from government is 0.25. The utility level within the city grows from 9.58 under the market equilibrium to 9.96 under road pricing. That corresponds to a 6 percent increase in the city wage.

Arguably, the most striking result of an introduction of the road pricing in the model is the increased demand for land, which results in urban sprawl: the city boundary moves from 6.76 (a no-toll case) to 7.92 (under the road pricing). That result holds even though higher land-rents flow out to the absentee land lord, and even when the lump-sum transfer is not redistributed back to the city inhabitants; although the extent of urban sprawl is then less pronounced. The intuition behind this outcome is,

despite the contrast with effects of road pricing in static models, in fact straightforward: compared to the no-toll case, the drivers spend more time at home, thus they have greater incentives to invest in larger housing. Changes in scheduling behavior can be observed by comparison of Figure 6 versus Figure 11, which shows that the new dynamic equilibrium travel patterns by location are such that all inhabitants, except the one who arrives first at work, depart from home later under optimal road pricing than in the unpriced market equilibrium, and thus spend more time at home. The peak period lasts for 10 units of time (out of a total morning period of 30), and the period of positive commuting flows lasts for exactly 10 units as well (i.e., during the other 20 units of time everyone is at home).

**Figure 11.** Travel patterns by residential location with the first-best road pricing



## 6. Sensitivity analysis and other policies

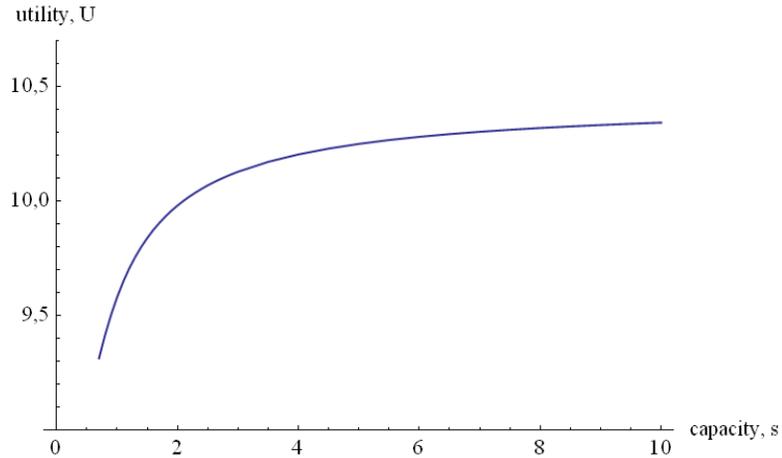
### 6.1 Alternative policy measures

Both land consumption and the utility level of inhabitants increase when first-best time-dependent road pricing eliminates queuing delays in the city. While a road tolling is a potentially Pareto-improving policy, the scarcity of real-life congestion pricing schemes signifies how technically and politically challenging the implementation of the road pricing is (in part, due to the redistribution effects). Noticeable exceptions are congestion charges in London, Singapore, Stockholm, and on some highways in the USA. Alternative policy instruments to improve traffic conditions have been proposed in the literature, and for some of them our model can provide an assessment of their spatial and welfare effects as well. Those policies include bottleneck capacity expansion, a change of the transportation technology, and the promotion of working from home. In our model, those policies might be represented with the changes of parameter values  $s$ ,  $T_{FF}$  and  $\varphi$ . For the sensitivity analysis, we focus on these parameters with a clear policy interpretation. We do not believe that there are other parameters of sufficient interest for a sensitivity analysis.

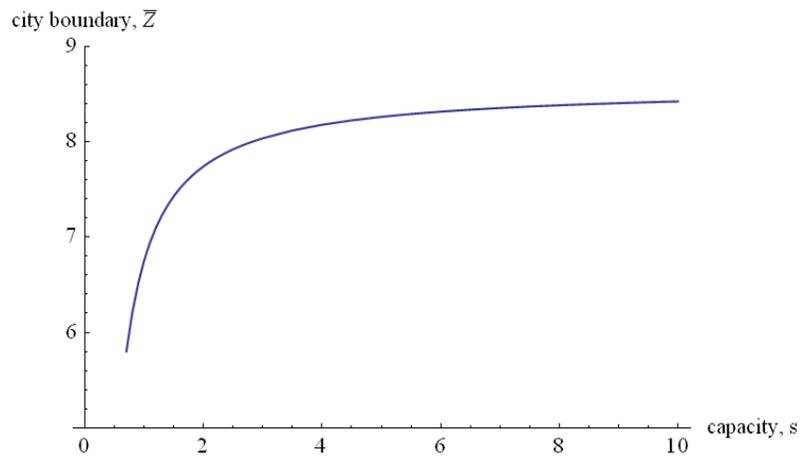
## 6.2 Bottleneck capacity ( $s$ )

Arguably, the most widespread policy response to alleviate congestion is to construct additional road capacity. In our model, we can analyze the role of capacity by varying the value of the parameter  $s$ , the capacity of the bottleneck. Figure 12 illustrates the effect of the capacity change on utility level in the city. The curve in Figure 12 is concave because a variation of  $s$  at lower levels changes peak duration, and therewith utility, to a greater extent than does the further expansion of an already large capacity. Expectedly, the utility increases in the city as the capacity is increased, because the time period of waiting in a car at each moment of arrival decreases, and because we do not consider any cost of capacity. Figure 13 shows that land at greater distance becomes inhabited as the capacity of the bottleneck is increased. The rent curve, not shown, becomes less steep if the capacity increases. This is

**Figure 12.** Utility for various bottleneck capacity levels



**Figure 13.** City size for various bottleneck capacity levels



consistent with the relative benefit of being close to the CBD reduces when queuing – experienced more by more distant residents – becomes less serious.

We also considered a case when the capacity level approached infinity. That case corresponds to a monocentric city model without congestion. All drivers then arrive at the CBD exactly at the preferred arrival time  $t^*$ , and the departure moments are  $t_d = T_{FF}Z$ . The resulting utility level in the city without congestion is 10.43 and the city fringe is located 8.58 units of distance away from the CBD; these are the asymptotes in Figures 12 and 13.

### 6.3 Free-flow speed ( $T_{FF}$ )

Another parameter of interest would be the free-flow speed, increases of which would lower the parameter value  $T_{FF}$ . Some locations that are initially outside the city will become resided when  $T_{FF}$  falls, because commuting time from those locations is no longer prohibitively high. A higher free-flow speed results in our model in larger land consumption, and larger utility level across the city. These effects are in line with those observed empirically after invention of a car (see, for example, Glaeser and Kahn 2004), and also with predictions from the monocentric model with static congestion.

In the limiting case of  $T_{FF} = 0$ , where spacious commuting is instantaneous and drivers depart from home and arrive at the back of the bottleneck at the same moment, the model collapses into the spaceless bottleneck model, with homogeneous drivers who derive utility from the land consumption. People pay exactly the agricultural rent for land and do not require a compact “city” (a continuous strip of land that is adjunct to the CBD) to live in. Each inhabitant then consumes 1.12 units of land and the utility derived is equal to 10.71. As in the standard Vickrey’s dynamic bottleneck model, the scheduling behavior then does not affect the spatial allocation of people and vice versa (Arnott 1998).

### 6.4 Lot size effect on the marginal utility of spending an additional time at home ( $\varphi$ )

The parameter  $\varphi$  relates residential land consumption to the marginal utility of spending an additional unit of time at home. Changes in  $\varphi$  might come through changes, among others, in idiosyncratic preferences for spending time at home due to, for example, a marriage (de Palma, Lindsey, and Picard 2010) or the birth of a child. From the policy point of view, (part-day) teleworking could be seen as a measure that might influence  $\varphi$  (Gubins and Verhoef 2011). We interpret teleworking as an out-of-office work arrangement, where an employee can perform some work tasks from home. This possibility would affect the willingness to pay for being at home before  $t^*$ .

An increase in  $\varphi$  leads to a lower density and higher rents, because land yields higher utility per unit of time and per unit of income, and thus induces people to consume more land. That leads to a higher utility level in the city, but the congestion cost becomes larger as well. A steeper travel delay function is required for equilibrium when  $\varphi$  becomes higher and being at home is valued more. Waiting time therefore increases for each moment of arrival, implying a more severe traffic jam.

## 7. Summary and conclusions

We analyzed road congestion in a city where both residential location choice and the choice of the time of departure in the morning peak are the results of optimizing behavior. We show that this setup yields results that are in sharp contrast to those from the more conventional approach of assuming static flow congestion. One of the main results is that there will be a decrease in the population density in a city, and an increase in city size, with first-best time-dependent road pricing (even when the collected road toll revenues are not redistributed back to the city inhabitants). Effectively, road pricing may thus lead to urban sprawl.

The driving force behind our results is an assumption that relates the choice of house size to the choice of the time of departure. An individual who lives in a large house is assumed to value an additional time spent in it at a higher rate than someone who lives in a smaller house. This assumption is justified with (i) the empirical evidences from the hedonic price studies, which indicate a positive effect of a lot size on utility, and (ii) the reasonable observation that, among other channels via which one might derive utility from housing, at least some part of that utility should depend on the amount of time an individual spends in it. We show that drivers in equilibrium, pass the bottleneck in order of increasing house size. In a monocentric city, the smallest houses are located close to the CBD, so drivers who live the closest to the CBD are the first to arrive at work.

The introduction of the first-best time-dependent road pricing eliminates queuing, as it does in the non-spatial bottleneck model, by substituting waiting time costs with a welfare neutral monetary transfer from the drivers to government. The first driver to arrive at work pays no toll, but a driver who arrives closer to the preferred time of arrival pays a toll that counterbalances the lower scheduling cost she faces. Unlike in the standard bottleneck model, the change in toll reflects differences in the marginal utility of income, induces by the positive income elasticity of demand for land. We account for this in our optimal toll function. As in the no-toll case, individuals account for the effect of the house size on the generalized travel cost; and the long-run rents, which we numerically compute, are such that the spatial equilibrium condition is fulfilled. Finally, we determine location specific travel patterns within the city based on the equilibrium land consumption function.

Optimal road pricing induces people to consume more land, as housing yields larger benefits because people spend more time at home. As the departure times at each given location (except the one closest to the CBD) move to a later moment with road pricing, the city territory grows because the commuting time from more distant locations to the CBD is no longer prohibitive. Similar effects apply when the bottleneck capacity expands, or when the free-flow travel time increases. In effect, we show a result which is familiar in the literature, namely, a better transportation technology causes urban sprawl.

Our paper thus implies that the relation between road pricing and city density needs not be as straightforward as suggested by the conventional monocentric model, with static congestion. All that is “needed”, in terms of assumptions, to achieve this, is that spending more time at home provides a greater incentive to use more space. This basic mechanism, represented in an admittedly stylized manner is our model, seems rather plausible. One of the consequences for transportation policy might

be that if some externalities – such as pollution – increase with driving distance, the interaction between congestion pricing and this other externality may turn from favorable – in the static model – a real trade-off – in a dynamic setting.

## References

- Anas, Alex, Richard Arnott, and Kenneth A. Small (1998). "Urban Spatial Structure." *Journal of Economic Literature*, 36, 1426–1464.
- Arnott, Richard (1998). "Congestion Tolling and Urban Spatial Structure." *Journal of Regional Science*, 38, 495–504.
- Arnott, Richard, Andre de Palma, and Robin Lindsey (1993). "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." *American Economic Review*, 83, 161–179.
- Arnott, Richard, and Elijah DePalma (2011). "The Corridor Problem: Preliminary Results on The No-Toll Equilibrium." *Transportation Research Part B: Methodology*, 45, 743–768.
- Bajari, Patrick, and Matthew E. Kahn (2008). "Estimating Hedonic Models of Consumer Demand with an Application to Urban Sprawl." In *Hedonic Methods in Housing Markets*, edited by Andrea Baranzini, Jose Ramirez, Caroline Schaerer, and Philippe Thalmann. Springer.
- de Palma, Andre, Robin Lindsey, and Nathalie Picard (2010). "The Benefits and Costs of Cooperation within Couples: Trip-timing Decisions and Traffic Congestion." Unpublished.
- Fosgerau, Mogens, and Andre de Palma (2012). "Congestion in a City with a Central Bottleneck." *Journal of Urban Economics*, 71, 269–277.
- Fosgerau, Mogens, and Leonid Engelson (2011). "The Value of Travel Time Variance." *Transportation Research Part B: Methodology*, 45, 1–8.
- Glaeser, Edward L. (2008). *Cities, Agglomeration and Spatial Equilibrium*. Oxford University Press.
- Glaeser, Edward L., and Matthew E. Kahn (2004). "Sprawl and Urban Growth." In *Handbook of Regional and Urban Economics*, Vol. 4., edited by J. Vernon Henderson and Jacques-François Thisse. Elsevier.
- Gubins, Sergejs, and Erik T. Verhoef (2011). "Teleworking and Congestion: A Dynamic Bottleneck Analysis." Tinbergen Institute Discussion Paper No. TI 2011-096/3.
- Mirrlees, James A. (1972). "The Optimum Town." *Swedish Journal of Economics*, 74, 114–135.
- Ross, Stephen L., and John Yinger (2000). "Timing Equilibria in an Urban Model with Congestion." *Journal of Urban Economics*, 47, 390–413.

- Small, Kenneth A. (1982). "The Scheduling of Consumer Activities: Work Trips." *American Economic Review*, 72, 467–479.
- Small, Kenneth A., and Erik T. Verhoef (2007). *The Economics of Urban Transportation*. Routledge.
- Solow, Robert M., and William S. Vickrey (1971). "Land Use in a Long Narrow City." *Journal of Economic Theory*, 3, 430–447.
- Tseng, Yin-Yen, and Erik T. Verhoef (2008). "Value of Time by Time of Day: A Stated-Preference Study." *Transportation Research Part B: Methodology*, 42, 607–618.
- van den Berg, Vincent, and Erik T. Verhoef (2011). "Winning or Losing from Dynamic Bottleneck Congestion Pricing? The Distributional Effects of Road Pricing with Heterogeneity in Values of Time and Schedule Delay." *Journal of Public Economics*, 95, 983–992.
- Vickrey, William S. (1969). "Congestion Theory and Transport Investment." *American Economic Review*, 59, 251–260.
- Vickrey, William S. (1973). "Pricing, Metering, and Efficiently Using Urban Transportation Facilities." *Highway Research Record*, 476, 36–48.
- Wildasin, David E. (1986). "Spatial Variation of the Marginal Utility of Income and Unequal Treatment of Equals." *Journal of Urban Economics*, 19, 125–129.