CONGESTION TOLLING
WITH
AGGLOMERATION EXTERNALITIES*

Richard Arnott [+]

March 2007

Preliminary draft: Please do not cite or quote without the permission of the author.

Abstract

Consider an urban economy with two types of externalities, negative traffic congestion externalities and positive agglomeration externalities deriving from non-market interaction. Suppose that urban travel can be tolled, that non-market interaction cannot be subsidized, and that non-market interaction is stimulated by a reduction in travel costs. Then the optimal toll is below the congestion externality cost. This paper explores this line of reasoning.

Keywords: congestion, congestion toll, agglomeration, externalities

JEL codes: D60, H20, R40

[+] *Tel.*: 617-552-3674. *Address*: Boston College, Department of Economics, Chestnut Hill, MA 02467, USA. *E-mail Address*: richard.arnott@bc.edu.

Congestion Tolling with Agglomeration Economies

1. Introduction

This paper explores an idea that has been in the air for some time but has not to my knowledge been formally examined. Many believe that non-market interaction is a dominant, and perhaps the principal, cause of urban agglomeration. Non-market interaction, it is argued, generates uninternalized positive externalities on both the consumption and production sides, which results in its being undersupplied in an otherwise efficient economy. One feature of "an otherwise efficient economy" is Pigouvian tolling of urban traffic congestion, which entails setting the toll equal to the congestion externality cost. If the toll is marginally lowered from the Pigouvian level, there will be more travel. If more travel generates more non-market interaction, then the small reduction in the toll below the Pigouvian level generates a first-order welfare gain with respect to non-market interaction and only a second-order welfare loss with respect to congestion. If, therefore, it is infeasible, for whatever reason, to directly internalize uninternalized urban agglomeration externalities due to non-market interaction, then the optimal congestion toll is lower than the Pigouvian toll.

What does the validity of this intuitive argument depend on? And if or when the argument is valid, what determines how much lower the optimal toll is than the Pigouvian toll. Under reasonable parameter values, might this effect be so strong as to warrant excluding urban travel from congestion tolling? If current empirical knowledge

is insufficient to estimate the optimal toll, what additional empirical information is needed to do so?

This paper addresses these questions through analysis of a pair of models and discussion of some extensions. Section 2 starts by briefly considering a variant of the monocentric city model, with traffic congestion and external economies of scale *à la* Chipman-Henderson (Chipman [10], Henderson [15]), and then adapts the model to include labor-leisure choice *à la* Parry-Bento [27] and external scale economies *à la* Henderson [16]. Section 3 addresses the same questions but in the context of a model of intra-day traffic dynamics and labor productivity, and considers a variety of extensions. Section 4 concludes.

The literature has identified many sources of agglomeration economies. The paper considers only one source of agglomeration economies, productivity-improving workplace interaction, and assumes that this interaction is not perfectly mediated through the market so that the corresponding externalities are completely uninternalized. How the qualitative results of the paper generalize to other sources of agglomeration economies is left for future research.

2. A Two-island Model: Productivity Depends on Aggregate Workdays

When asked about how traffic congestion and agglomeration economies interplay, a well-trained urban economist would think immediately of the interplay in the context of the closed, monocentric city model, in which the city's population is fixed. As it turns out, in the context of this model the interplay between traffic congestion and agglomeration economies is uninteresting, but presenting the monocentric city model will set the stage for other urban models in which the interplay is potentially important.

The standard treatment of traffic congestion in the monocentric city model (e.g., Strotz [35], Mills and de Ferranti [24], Solow [34], Kanemoto [17], and Arnott [1]) assumes that traffic is uniform over the day, with each resident taking a single return trip from home to the CBD (central business district), that traffic congestion uses up physical resources rather than time, and that the private cost to the driver (in terms of the composite commodity) of traversing a section of road between distances y and y + dy away from the CBD is c(Q(y),w(y))dy, where Q(y) is the number of travelers on the road at y per day, which equals the number of residents living beyond y, and w(y) is the capacity or width of the road at y. The social cost of the driver traversing the section of road is [c + Q(∂c/∂Q)]dy. Thus, there is the familiar congestion externality, which can be internalized by imposing a Pigouvian tax, in this context a congestion toll, equal to $\tau =$ Q(∂c/∂Q) per unit distance.

The treatment of external economies of scale in production follows Henderson [15]. Letting E denote the aggregate number of employees in the CBD, and e_j the number of employees at firm j, the firm's output is $G(E)ke_j$, where $G' > 0$ and k is a constant. Each firm takes $G(E)$ as fixed, and hence views itself as facing constant returns to scale in production. At the level of the urban economy, however, there are increasing returns to scale. The scale economies are therefore external to the individual firm. In the urban economics context, these Marshallian production externalities are considered under the rubric of urban agglomeration economies[1]. The private marginal product of labor is Gk but the social marginal product of labor is $Gk + EG'k$. When a firm hires a worker, it causes the marginal product of labor at all other firms to rise, generating a positive externality. This urban agglomeration externality would be internalized by introducing a Pigouvian wage subsidy of EG'k.

An essential feature of the model is that urban production depends on the *number* of urban residents. In a closed, urban economy, the number of residents is fixed. Since congestion tolling has no effect on the number of residents, it has no effect on urban output. Thus, there is no interplay between the agglomeration and traffic congestion externalities, so that the optimal toll coincides with the Pigouvian toll.

The interplay between the two externalities in an open monocentric model is more complex. How the combined externalities distort allocation then depends on how they

---

[1] There is a large and relatively recent literature on urban agglomeration economies. Fujita and Thisse [13] and Puga [11] review the theoretical literature. Rosenthal and Strange [30] reviews the empirical literature, and Moretti [25] the theoretical and empirical literatures on human-capital-based agglomeration economies.

affect the distribution of the economy's population across cities, an issue which is treated in section 12.3 of Pines and Papageorgiou [28]. This paper sidesteps this issue by focusing throughout on a single, closed city.

Assuming that a closed city's output depends only on its population is an evident simplification. It would be a good assumption if all households have a single wage-earner, and if all wage-earners have a workday of the same length and work the same number of days per week. That might have been a good approximation to reality at certain times and places in the past, but employees now have more flexibility with respect to when and how much they work and labor force participation rates show considerable variation. The next section will present a dynamic model that solves for the pattern of employment and congestion over the course of the day. This section considers a simpler, stationary-state model in which each individual decides how many fixed-length workdays per week to work and productivity depends on aggregate work hours. The interplay between congestion and productivity will be mediated by the labor-leisure choice.

The interplay between congestion and the labor-leisure choice has been considered in a series of papers in the literature (e.g., Parry and Bento [27]; Calthrop, Proost, and van Dender [8]). The substitution effect of income taxation distorts the labor-leisure choice, encouraging leisure. If all travel is for commuting purposes, reducing the toll below the Pigouvian level increases the net wage and stimulates labor. With distortionary income taxation, the optimal toll therefore falls short of the Pigouvian toll, and at high rates of income taxation could even be negative, if this were institutionally feasible.

Here, income taxation will not be treated, but the basic idea is the same. Start with a situation where the Pigouvian congestion toll is applied. External economies of production result in workers being paid their average product of labor, while efficiency calls for them to be paid their marginal product, which exceeds the average product. Thus, the wage rate is set inefficiently low, encouraging leisure, and labor can be stimulated by lowering the toll below the Pigouvian level.

The essential differences between the model presented below and the closed, monocentric city model discussed earlier are that, first, productivity is an increasing function not of the number of workers but of the aggregate number of work hours[2], and, second, tolling reduces the net wage. The model of this section has a simpler spatial structure than the monocentric city, with two islands connected by a congestible causeway, one for residence, the other for work.

2.1. Model description

The economy is in stationary state, and the unit of time is a day.

---

[2] Henderson [16] employed a similar assumption in the context of a model of staggered work hours, assuming that a worker's productivity at a point in time is a function of the number of workers at work at that point in time. Subsequent works that employ Henderson's assumption include Mun and Yokekawa [26], and Yoshimura and Okumura [41], and Arnott, Rave, and Schöb [6]. Empirical support for Henderson's assumption is provided in Wilson [40]. Using the results of a 1975 survey for Singapore, Wilson found that, after controlling for measured differences between workers, the daily wage is on average twice as high for workers with a peak work start-time (when more workers are at work) than for those with an off-peak work start-time. Intuition suggests that this difference is too large to be explained by intra-day productivity effects alone, and that sorting of workers across start times, on the basis of ability attributes observable to employers but not to the empirical researcher, must play an important role too. No empirical work has been done that attempts to distinguish between the two effects, though Ross and Fu [31] examine the analogous issue for spatial rather than temporal agglomeration.

Residents: The city is closed, with N identical residents. A resident's tastes are described by a strictly increasing and strictly concave utility function, defined over[3] other goods (the numéraire) c and leisure ℓ: u = u(c,ℓ). The workday, L, is of fixed length, and residents decide on the proportion of days to work, x.

Congestion: Production takes place on one island, residence on another. Between the two islands is a causeway of fixed capacity, and round-trip travel time is increasing in the daily flow, t = t(f). The only travel is for commuting purposes, and travel uses up no physical resources.[4] In the numerical examples, the form of the congestion function will be assumed[5] to be $t(f) = t_0 + t_1 f^\beta$, where t_0, t_1, and β are positive constants.

Production: Where H denotes aggregate workdays, aggregate daily output is F(H), with F' > 0 and F" > 0; hence, production exhibits increasing returns to scale. These scale economies are external to the individual competitive firm, so that a firm employing h workdays of labor views itself as facing the production function [F(H)/H]h. In the numerical examples, the form of the production function will be assumed to be F(H) = kH^α, where k and α positive constants, with α - 1 being the degree of increasing returns to scale. A firm employing h workdays of labor views itself as facing the production function Kh, with K = $kH^{\alpha-1}$. The average product of labor is $kH^{\alpha-1}$ and the marginal product $\alpha kH^{\alpha-1}$.

---

[3] The model is consistent with the assumption that all residents reside on identical lots on the residence island.

[4] This assumption is retained throughout the paper. Introducing money costs of travel would be straightforward, though somewhat tedious, and would not alter any of the qualitative results.

[5] This is the most commonly employed form of the congestion function, and is termed variously the Bureau of Public Roads congestion function or the Vickrey congestion function. It has the neat property that the ratio of the congestion externality to private congestion (the time loss by the individual driver due to congestion) is β.

The competitive equilibrium is solved first and then the social optimum.

2.2. Competitive equilibrium

Because the returns to scale are external to the individual competitive firm, the workday wage, w, equals the average product of labor over the workday. A congestion toll is imposed of τ per round trip, and toll revenues are distributed as an equal, lump-sum subsidy to all residents, T. A resident's daily budget constraint is

$$(w - \tau)x + T - c = 0 , \tag{1}$$

and his time constraint, expressed in averages per day, is

$$1 - (L + t)x - \ell = 0 . \tag{2}$$

After substituting these constraints into the utility function, the individual's maximization problem is

$$\max_{x} u((w - \tau)x + T, 1 - (L + t)x) . \tag{3}$$

An interior maximum is assumed. The corresponding first-order condition is

$$(w - \tau)u_c - (L + t)u_\ell = 0 . \tag{4}$$

A resident views his opportunity cost of leisure or the private value of time as $(w - \tau)/(L + t)$. Letting AP $(= w = K = F/H)$ denote the average product of a workday and PT $(= L + t)$ the private time cost of a workday, this may be written as $(AP - \tau)/PT$.

An alternative solution procedure is to combine (1) and (2) by substituting out x, and then have the individual maximize $u(c,\ell)$ subject to this combined constraint

$$1 - \frac{(L + t)(T - c)}{w - \tau} - \ell = 0 . \tag{5}$$

2.3. Social optimum

Since daily output is F(H) = F(Nx), the resource constraint for the economy is

$$F(Nx) - Nc = 0 \ . \tag{6}$$

The economy-wide time constraint from the planner's perspective is

$$1 - (L + t(Nx))x - \ell = 0 \ , \tag{7}$$

since the flow rate on the causeway is Nx. After substituting these constraints into the utility function, the planner's maximization problem is

$$\max_{x} u(\frac{F(Nx)}{N}, 1 - (L + t(Nx)x)) \ . \tag{8}$$

On the assumption that the global optimum is interior, the corresponding first-order condition is

$$F'u_c - (L + t + Nt'x)u_\ell = 0 \ . \tag{9}$$

Letting MP (= F') denote the marginal product of a workday and ST (= L + t + Nt'x) the social time cost of a workday, the social value of time (which may be termed alternatively the social opportunity cost or shadow price of leisure) may be written as MP/ST.

2.4. Competitive decentralization of the social optimum

It is assumed that the government can impose a toll but not a wage subsidy.   Since there is only one margin of choice, x, and since toll revenues are fully redistributed, then it should be possible to set the toll at such a level that the social optimum is decentralized. This is shown in Figure 1. It plots the transformation frontier and indifference curves in

ℓ-c space.   Since it has been assumed that the global maximum is interior, the social optimum occurs at a point of tangency between an indifference curve and the transformation frontier, the point Ω in the Figure.   The social optimum can be decentralized by having each resident maximize his utility subject to a linear budget constraint that is tangent to the indifference curve at the social optimum. The resident chooses the socially optimal level of ℓ, and hence via (2) of x. Furthermore, since toll revenue is fully redistributed, the allocation of the decentralized economy lies on the transformation frontier, so that the socially optimal level of c is achieved as well.
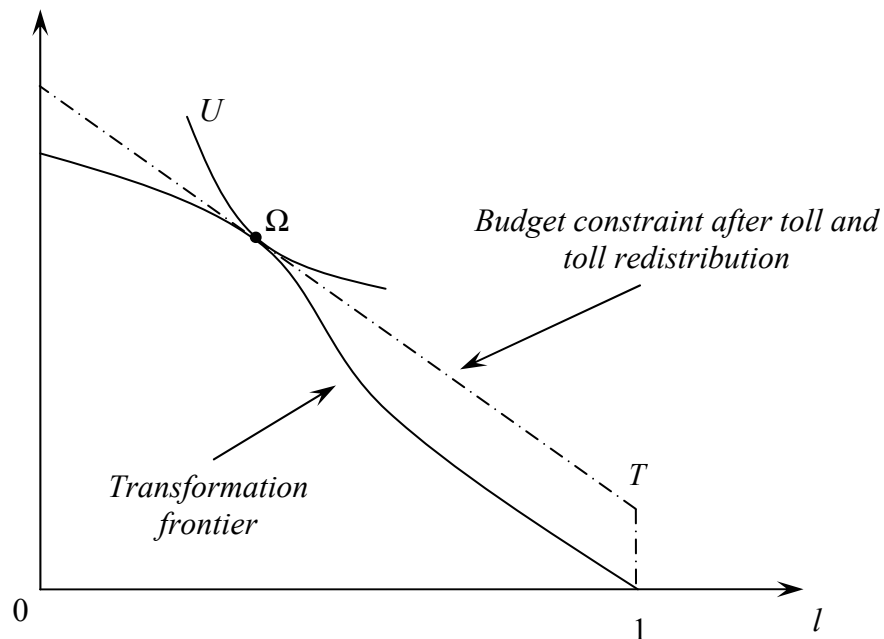
Figure 1: Decentralization of the optimum

The toll should be set to equalize the private and social opportunity cost of leisure. Letting τ^* denote the optimal congestion toll, from (4) and (9) one obtains

$$\tau^* = -(MP - AP) + \frac{u_\ell}{u_c}(ST - PT) \ . \tag{10}$$

As expected, the optimal toll equals minus the agglomeration externality in terms of output, plus the congestion externality in terms of time multiplied by the value of time. Because the model incorporates only one margin of choice, application of the optimal toll results in attainment of the social optimum.

Some numerical examples are now considered that use the specific functional forms listed earlier. With these functional forms, (10) reduces to

$$\frac{\tau^*}{w} = 1 - \frac{\alpha PT}{ST} \ . \tag{11}$$

Now, $PT = L + t_0 + PC$, where t_0 is free-flow travel time and PC is private congestion, the per-trip time loss incurred by a driver due to congestion, and $ST = L + t_0 + PC + CE$, where CE is the per-trip congestion externality. With the form of the congestion function assumed, $CE = \beta PC$, so that $ST = L + t_0 + (\beta + 1)PC$. Inserting these results into (11) gives

$$\frac{\tau^*}{w} = 1 - \frac{\alpha(L + t_0 + PC)}{L + t_0 + (\beta + 1)PC} \ . \tag{12}$$

The literature attempting to measure urban agglomeration economies is reviewed in Rosenthal and Strange [30]. A variety of different empirical approaches have been employed. Measures of α vary across industries, but estimates of the order of 1.05 are typical. A range from 1.0 to 1.08 is considered here, with 1.0 corresponding to constant returns to scale and hence no agglomeration externality. There is more disagreement

concerning the magnitude[6] of β. Estimates obtained from observing the relationship between flow and velocity at a point on a major traffic artery or freeway find β to be around 3.0 in moderately congested conditions and even higher with heavy congestion. More recent studies that look at long sections of road or freeway obtain estimates closer to 1.0. Here values of β between 1.0 and 3.0 are considered. The workday is assumed to be 8.0 hours and one-way free-flow travel time to be 20 minutes. Furthermore, the congestion function is varied across examples such that with optimal tolling[7] one-way travel time is 30 minutes (thus, in all the examples. PT = 9.0 hours and ST = 8.666 + (β + 1)0.333 hours). Table 1 gives the values of τ^*/w for different pairs of α and β, which are calculated by direct application of (11).

α β	1.0	1.02	1.05	1.08
1.0	0.0357	0.0164	- 0.0125	- 0.04143
2.0	0.0690	0.0503	0.0224	- 0.0055
3.0	0.100	0.0820	0.0550	0.0280

Table 1: Values of τ^*/w for different values of α and β, free-flow return-trip travel time of 0.666 hours, rush-hour return-trip travel time of 1.0 hours, and a workday of 8.0 hours.

---

[6] See the discussion relating to Figure 3.6 of Small [32] and in Chapter 5 of Arnott, Rave, and Schöb [6]. Arnott, de Palma, and Lindsey [5] show that with identical individuals the elasticity is 1.0 for the simple bottleneck model of traffic congestion, which endogenizes departure time and assumes that congestion takes the form of a queue behind a single bottleneck with fixed discharge rates. Cassidy [9] reviews results of his work with Carlos Daganzo, based on very detailed traffic count data, that finds that the discharge rate from bottlenecks on freeways is more or less fixed, which supports the bottleneck model. But more recently, the same authors have argued, in informal communication with the author, that the elasticity of travel time with respect to flow may be considerably higher when spillbacks and merges occur.
[7] The numerical analysis would be considerably more complicated if the form of the congestion function or rush-hour travel time in the absence of congestion tolling were held constant across the numerical examples. The results would then depend on the form of the utility function.

When α = 1.0, external scale economies are absent and the optimal toll coincides with the Pigouvian toll. The optimal toll is largest when the congestion externality, measured by β, is large, and when the external scale economies, measured by α, are small. When β = 3.0 and α = 1.0, the optimal toll equals 10.0% of the daily wage. With a workday wage of $200, the round-trip congestion toll would be $20.00. The magnitude of the optimal toll is lowest when the congestion externality is small and external scale economies are large. When β = 1.0 and α = 1.08, the optimal toll equals minus 4.14% of the daily wage, corresponding to a subsidy to a round trip of around $8.29.

The numerical results in Table 1 are based on a very simple model. The only margin of individual choice is the proportion of fixed-hour workdays to work; there is only the single employment location and the single residence location; at the level of the economy, work and commuting are uniformly distributed over the course of the day; land use is fixed; the assumption of a fixed-hour workday, while empirically reasonable, is *ad hoc*; the agglomeration economies are a black box; and it was simply assumed that agglomeration economies are uninternalized. The reader is therefore cautioned not to attach much significance to particular numerical values. The purpose of the example was rather to indicate that, if tolling is the only instrument available to mitigate the distortions associated with congestion and agglomeration externalities, then, within the range of empirical estimates of α and β, and for a particular model, the optimal toll might not only

be substantially lower than the Pigouvian toll, but might even be negative, entailing a subsidy to urban travel.[8]

It is appropriate at this juncture, before moving on to other models, to pause and reflect. Back in the 1970's, when applied competitive microeconomic theory was making the transition from partial to general equilibrium analysis, there was a small literature on piecemeal second-best policy (e.g., Boadway and Harris [7]). It asked whether there is any presumption that the policy prescriptions for correcting distortions derived from partial equilibrium analysis carry over to a general equilibrium setting. In the context of externalities, the question would be: If the government has detailed information related to some externalities but not to others, is there a presumption that applying the partial equilibrium rules for internalizing those externalities about which it is well informed would improve efficiency? The literature on urban congestion tolling has implicitly assumed the answer to be positive. This paper challenges this implicit assumption. As is well known, urban agglomeration comes about from the balancing of positive, attractive forces and negative, repulsive forces. Congestion is one of the negative forces and scale economies one of the positive forces. Transactions costs, difficulties in defining property rights, and asymmetric information give rise to incomplete markets and incomplete contracts, which result in the market outcome incompletely incorporating these forces – uninternalized externalities. There seems to be no sound basis for believing that the

---

[8] One could augment the model in reasonable ways so that the optimal toll is even lower. As noted at the beginning of the section, the income tax discourages work; incorporation of income taxation would therefore lead to lower optimal tolls. In the model, the private value of travel time equals the (net-of-toll) daily wage divided by the sum of hours worked and commuting time. Empirical estimates (Small, Winston, and Yan [33] is the state-of-the-art study) however find that the value of travel time is about one-half the wage rate. This result could be incorporated into the above model by assuming that workers enjoy travel time, though less than leisure, and if this were done the optimal congestion toll would be even lower.

market does a better job of incorporating the positive than the negative forces, nor therefore that the government's intervening to internalize only the negative externalities would be beneficial.

The above analysis made this point in the context of a specific model in which there is only one negative externality, traffic congestion, and one positive externality, external economies of scale. It assumed that congestion tolling is feasible but that a wage subsidy is not, and demonstrated that imposition of even a low toll -- and hence *a fortiori* the Pigouvian congestion toll -- might reduce efficiency. The analysis could just have well have addressed the mirror question. It could have assumed that a wage subsidy is feasible and that congestion tolling is not, and demonstrated that application of a small wage subsidy could reduce efficiency.

A model is just a model, and has no policy implications *per se*. But to the extent that the above model captures essential features of real urban economies, its message is conservative: Think twice before applying piecemeal second-best policy in the urban context. That is not to say that the government should not attempt to intervene to correct externalities but rather than it should consider the interplay between externalities in doing so.

The rest of the paper looks at the interplay between congestion and external economies of scale in the context of more sophisticated and realistic models of the urban economy.

3. Intra-day Dynamics

3.1. Introduction

The previous section assumed the urban economy to be in a stationary state, with uniform traffic flow and an equal number of workers at work at all times. In fact, while the distribution of work start-times is not as concentrated as it used to be, it remains highly concentrated. Since workers traveling at peak hours experience considerably more congestion than workers traveling in off-peak hours, they must be compensated in some way for the higher congestion they experience. One form of compensation is reduced schedule delay costs; working normal hours allows them to better synchronize their non-work activities with others. Another is higher wages. If wages are indeed higher for a normal workday, it must be because an individual who works a normal workday is more productive, presumably because of expanded scope for interaction. These observations suggest that a persuasive analysis of uninternalized agglomeration externalities[9] should consider the intra-day dynamics of work productivity and congestion.

Henderson [16] is the seminal paper along these lines. His paper assumes that all a city's residents work at the same location, the CBD, and reside at a common but different location, that commuting entails congestion, and that the rate at which a worker produces

---

[9] The literature has evolved in the context of studying the positive and normative aspects of flextime and staggered work hours. A central question has been whether the government should encourage the staggering of work start-times. It can do this either by regulating the distribution of private firms' work start-times or by staggering the work start-times of government workers. Works that examine issues related to flextime and staggered work hours include Henderson [16], Giuliano and Golob [14], Mun and Yonekawa [26], Yoshimura and Okumura [41], and Arnott, Rave, and Schöb, Ch.4 [6].

the composite good at a point in time is an increasing function of the number of workers then at work. Since each worker is uncompensated for the benefit his presence at work provides other workers then at work, there is a positive uninternalized externality, that Yoshimura and Okumura [41] refer to as a *temporal* agglomeration externality. Having specified the congestion technology as well, one can solve for the equilibrium distribution of work start times and the equilibrium evolution of congestion over the course of the day. Henderson's model can be criticized in two respects. First, the interaction technology is unduly restrictive.[10] For example, it may be important only that there be some period during the workday when everyone on a team can assemble for a joint meeting, and it may be beneficial for each member to have some quiet time. Second, the specification of the congestion technology assumes that a worker's commute time depends on the number of workers who leave home at the same time as he does.[11] This leads to the out-of-equilibrium possibility that a worker who leaves home later than another arrives at work earlier, which is inconsistent with the laws of physics. At first glance, the simple bottleneck model (Vickrey [37], Arnott, de Palma, and Lindsey [5]), in which rush-hour traffic congestion is modeled as a queue behind a single bottleneck of fixed flow capacity, would appear to be an attractive alternative since it is tractable and its physics sound. Unfortunately, the bottleneck model results in a uniform arrival rate at work and, if one assumes a workday of fixed length, no congestion during the evening

---

[10] Where $\varphi(t)$ is a worker's productivity at time t and W(t) the number of workers at work at time t, Henderson assumes that $\varphi(t) = \hat{\varphi}(W(t))$. A general specification would be that the output over the entire workday of a worker who arrives at time t, $\Phi(t)$, is a function of the number of workers at work throughout his shift, i.e. $\Phi(t) = \hat{\Phi}(\langle W(t)\rangle)$, where $\langle W(t)\rangle$ denotes the entire time path of W(t), and is hence a functional.

[11] The Henderson treatment of congestion dynamics does however have the virtue of tractability, and for this reason has been employed in a number of studies, including Mun and Yokenama [26], Yoshimura and Okumura [41], Lindsey [21], and Arnott, Rave, and Schöb ([6], Ch.4).

rush hour[12] (Arnott [2]). Another possibility is to incorporate non-stationary flow congestion, which entails combining the equation of continuity with some function relating travel speed and density (Lighthill and Whitham [20]), but determining equilibrium then entails solving a partial differential equation whose properties are difficult to analyze.

The rest of the paper works with a general formulation of the intra-day dynamics of productivity and congestion, drawing on a general treatment of congestion formulated by Marvin Kraus [18]. The only restriction it imposes on either the dynamics of traffic congestion and of intra-day productivity is that time be discrete.

The city is closed with population N, and as in the model of the previous section there is a single residence location and a single employment location, which are connected by a single congestible road. All individuals are *ex ante* identical, with the common utility function $u(c,\ell)$ defined over other goods and leisure. The day is divided up into an arbitrarily large, but finite, number, I, of discrete time increments of equal length, indexed by i. The distribution of work start-times is given by the vector $\mathbf{n}$, where n_i is the number of individuals who start work at time i. It is assumed that the length of the workday is exogenous and the same for all workers, that a single numéraire good is produced, that the output of an individual worker can be identified and is a function of $\mathbf{n}$, and that each worker is paid his output. Let $z_i(\mathbf{n})$ be the output of an individual worker

---

[12] Suppose that the bottleneck's flow capacity is b. Then during the morning rush hour residents arrive at the CBD at the rate b per unit time. If all of them start work as soon as they arrive, then they start work at the rate b per unit time. And if they all have the same length of workday, they leave work at the rate b per unit time. But leaving work at this rate results in no queue behind the bottleneck and therefore no congestion.

who starts work at time i, and refer to **z(n)** as the private product function. It is furthermore assumed that the travel time of an individual who starts work at time i is implied by **n**, so that one may write $t_i(\mathbf{n})$, with **t(n)** being referred to as the congestion function.

3.2. A simple variant of the model

Since the model is unfamiliar and rather abstract, a simple variant of the model in which each worker works every day (x = 1) is presented. The utility of an individual who starts work at time i is then $u(z_i(\mathbf{n}) - \tau_i + T, 1 - L - t_i(\mathbf{n}))$, where τ_i is the toll payable by a worker who starts work at time i, T is the equal lump sum received by each individual from redistribution of toll revenue, and L as before is the fraction of a day taken up by a workday.

A competitive equilibrium conditional on $\boldsymbol{\tau}$ is defined to be a triple $(\mathbf{n},T,\underline{u})$ satisfying the following three conditions:

1. **n** is determined as the outcome of decentralized choice. Letting $\mathcal{W}$ denote the set of times at which workers start work, this condition can be written as

$$u(z_i(\mathbf{n}) - \tau_i + T, 1 - L - t_i(\mathbf{n})) = \underline{u} \qquad \text{for all } i \in \mathcal{W}$$

$$u(z_i(\mathbf{n}) - \tau_i + T, 1 - L - t_i(\mathbf{n})) \leq \underline{u} \qquad \text{for all } i \notin \mathcal{W} \ ,$$

(13)

where $\underline{u}$ is the equilibrium level of utility. Eq. (13) implies that no individual can improve his utility by shifting to a different work start-time.

2. All toll revenues must be redistributed.

$$\sum_i n_i \tau_i - NT = 0 \ .$$

(14)

3. The entire population must work.

$$\sum_i n_i - N = 0 \ . \tag{15}$$

Since τ and T enter the government revenue constraint and the utility function only as τ_i - T, there is an element of indeterminacy of the equilibrium. Everyone takes one trip per day, so, if all tolls are raised by a common amount, the redistributed toll revenue rises by the same amount, and utility remains unchanged.

The efficient competitive equilibrium will be defined to be the competitive equilibrium that maximizes the common utility level. The efficient competitive equilibrium will be obtained as the solution to a planning problem. The most natural specification of the planning problem is to maximize the common utility level with respect to $\underline{u}$, $\mathbf{n}$, τ, and T subject to (13), (14), and (15). Since, however, the solution will be more intuitive if the Lagrange multipliers are expressed in monetary units, an alternative specification will be analyzed, in which the government is assumed to maximize its surplus, defined as toll revenue less lump-sum transfers, subject to the utility and population constraints.

This maximization problem in Lagrangian form is

$$\max_{\mathbf{n},\tau,T} \ \sum_i n_i\tau_i - NT + \sum_i \lambda_i[u(z_i(\mathbf{n}) - \tau_i + T, 1 - L - t_i(\mathbf{n})) - \underline{u}^*] + \phi[N - \sum_i n_i] \ , \tag{16}$$

where a * on a variable indicates its value at the efficient competitive equilibrium, λ_i is the Lagrange multiplier on (13) for time i, and ϕ that on (15). The first-order conditions[13] are

---

[13] The first-order conditions should properly be written in Kuhn-Tucker form. Doing so would not alter the economic results.

$$n_j: \quad \tau_j - \phi + \Sigma_i \lambda_i \left(\frac{\partial u_i}{\partial c_i} \frac{\partial z_i}{\partial n_j} - \frac{\partial u_i}{\partial \ell_i} \frac{\partial t_i}{\partial n_j} \right) = 0 \tag{17}$$

$$\tau_i: \quad n_i - \lambda_i \frac{\partial u_i}{\partial c_i} = 0 \tag{18}$$

$$T: \quad -N + \Sigma_i \lambda_i \frac{\partial u_i}{\partial c_i} = 0 \quad . \tag{19}$$

Note that (19) is obtained by summing (18) over i, reflecting the indeterminacy noted earlier. Define

$$\Gamma_{ij} = -\frac{\partial z_i}{\partial n_j} + \frac{\partial u_i / \partial \ell_i}{\partial u_i / \partial c_i} \frac{\partial t_i}{\partial n_j} \quad . \tag{20}$$

Substituting (18) and (20) into (17) yields

$$\tau_j = \phi + \Sigma_i n_i \Gamma_{ij} \quad . \tag{21}$$

Now, Γ_{ij} is the net externality cost an individual with work start-time j imposes on an individual with work start-time i, which equals minus the positive agglomeration externality benefit plus the congestion externality cost. Thus, $\Sigma_i n_i \Gamma_{ij}$ is the net externality cost an individual with work start-time j imposes on all other individuals, and hence (21) states that the optimal toll to apply to an individual with work start-time j equals the net externality cost he imposes, plus a constant. What is the role of the constant ϕ? Recall that the first-order conditions do not determine the level of T. Setting T such that NT = $\Sigma_i n_i^* \Gamma_{ij}^*$, results in $\phi = 0$ at the optimum. Finally, note that the optimum to this problem with T set in this way satisfies the definition of a competitive equilibrium, and is the efficient competitive equilibrium.

The second-order conditions are not written down since the focus is on the global maximum for which they will hold.

If the government is able to charge a time-varying road toll but not any form of wage subsidy, the above result indicates that, when each individual is paid his private product, the Pareto efficient allocation can be achieved by the government imposing a time-varying road toll, such that the toll paid by each individual equals the congestion externality cost he imposes on others in his travel to and from work, less the production benefit he confers on others over his workday, and redistributing the toll revenue collected as a uniform, lump-sum subsidy.[14]

This result is what one would expect. Since both externalities are specific to a work start-time, setting the toll at each start-time at a level that internalizes the sum of the externalities associated with that work start-time should permit decentralization of the Pareto efficient allocation – and it does.[15]

---

[14] The model assumed that the utility function has the form $u(c,\ell)$. However, individuals might have preferences over the work start-time *per se* for scheduling reasons. This can easily be incorporated by assuming instead that the utility has the form $u_i(c,\ell)$. It is easily checked that the introduction of work start-time preferences, which loom large in the literature on the bottleneck model in the form of schedule delay costs, do not alter the above result.

[15] A word of caution is in order. Decentralization of an efficient allocation entails the central planner making the efficient allocation and then setting prices that support it -- given that allocation and those prices, no one has an incentive to alter his behavior. This does not imply that if the government simply sets the prices, the economy will necessarily achieve the efficient allocation.

Here it has been shown that, if the government chooses the distribution of work start-times, sets the time-varying toll according to (14) with $\phi = 0$, and redistributes the toll revenue as an equal lump sum, no one has an incentive to change his behavior. It has not been shown that the economy will necessarily achieve the efficient allocation if the government simply sets the time-varying toll according to (14) with $\phi = 0$ and redistributes the toll revenue as an equal lump sum.

In the problem under consideration, because of the nonconvexities generated by the external economies of scale, there may be multiple local optima. See Mas-Colell, Whinston, and Green [22], Ch. 11, for a discussion.

3.3. Extensions

This subsection considers a variety of extensions of the simple model variant analyzed in the previous subsection. An earlier version of the paper explicitly analyzed some of these extensions, but here all the extensions will simply be sketched. The extensions will be introduced one at a time rather than in combination.

- *Heterogeneous individuals*

Individuals may differ in tastes, in how they contribute to congestion, and in how their presence at work affects the productivity of others. The analysis of the simple model variant may be extended to treat such heterogeneity by adding a subscript to index individual type, so that n_{ik} would be the number of individuals of (exogenous) type k who start work at time i. Since individuals differ by type, there is not a single efficient allocation but rather a set of Pareto efficient allocations. Associated with a particular Pareto efficient allocation would be a set of utility levels $\{u_k*\}$. Decentralization of this Pareto efficient allocation can be achieved by setting Pigouvian tolls distinguished by type, $\{\tau_{ik}\}$, and making the lump-sum transfers type-specific as well, $\{T_k\}$, such that $\sum_k N_k T_k = \sum_i \sum_k n_{ik} \tau_{ik}$. But in most contexts, the government is unable to completely observe relevant differences in individuals' types, and hence unable to perfectly differentiate tolls and transfers on the basis of them. How does this alter the nature of the problem? Generally, first-best allocations cannot be achieved and determination of optimal tolls becomes an exercise in the theory of the second best.[16] These second-best optimal tolls

---

[16] Determination of the second-best set of congestion tolls, $\{\tau_i\}$, entails solving a fairly conventional optimal tax problem. There is a substantial body of literature that investigates second-best issues in urban transportation: Lévy-Lambert [19] investigates second-best mass transit pricing with underpriced auto

reflect both the congestion and agglomeration externalities. There is, however, one circumstance in which Pareto efficiency is still achieved – when both the congestion externality[17] and the agglomeration externality are *anonymous* with respect to individual type k. For this to be the case, for each i, the n_{ik}'s must enter the **t** and **z** functions additively, i.e. for all i and k, $t_{ik} = t_i = t_i(\sum_{k'} n_{1k'}, \sum_{k'} n_{2k'}, ---)$. Since driving behavior is probably much the same across unobservable types, this is a reasonable assumption for the congestion function. But intuition and the limited empirical evidence available (Wilson [40]) suggest that the agglomeration externality varies considerably across education level and occupation.

The paper has explicitly analyzed two models, that in section 2 and that in section 3.2. In both of these two models, it was possible to achieve full efficiency with only a road toll, even though there were both congestion and agglomeration economies. The discussion of the previous paragraph indicates that this result is not general and derives from the simplicity of the two models. Nevertheless, the central insight of the paper remains intact: (second-best) optimal road tolls should be set taking into account uninternalized agglomeration externalities.

---

congestion; Wheaton [38] and Wilson [39] consider second-best highway capacity with underpriced auto congestion; Mayeres and Proost [23] examine optimal tax and investment rules with congestion externalities; Arnott and Yan [3] consider the second-best, two-mode problem when auto congestion is underpriced; and Verhoef [36] examines second-best congestion tolling when only a subset of the road network can be tolled.

[17] Arnott and Kraus [4] prove essentially the same result, though the model in that paper does not incorporate the agglomeration externality and treats time as continuous.

- *Heterogeneous firms*

Suppose now that individuals are identical but that firms differ with respect to their private product functions, so that $z_{im} = z_{im}(\{n_{im}\})$, where m indexes firm type. Perhaps the firms are in the same industry but differ in the production process they employ; perhaps they are in different industries. Since individuals are identical, there is a single Pareto efficient allocation. Decentralization of this Pareto efficient allocation can be achieved by setting Pigouvian tolls distinguished by firm type, $\{\tau_{im}\}$, and providing the same lump-sum transfer of toll revenue to all individuals. But realistically the government would be unable to perfectly differentiate tolls by firm, in which case the Pareto efficient allocation cannot in general be achieved and the calculation of optimal tolls becomes an exercise in the theory of the second best.

- *Incorporating labor-leisure choice*

Now return to the simple model variant. Labor-leisure choice can be incorporated into the model in the same way as was done in the model of section 2. The results of the model of section 3.2 carry over. The efficient competitive equilibrium can still be decentralized via application of the optimal time-varying toll. With the toll imposed, individuals face the correct marginal prices, and make efficient decisions on both the work start-time and labor-leisure margins. One property of the efficient competitive equilibrium is worthy of remark. Consider the case where individuals have no preference for work start-time *per se*, *viz.* the utility function has form $u(c,\ell)$ rather than $u_i(c,\ell)$. Figure 1 continues to apply. Since all individuals have the same equilibrium indifference curve in ℓ-c space and receive the same lump-sum toll revenue payment, they must all have the same budget

constraint *in this space,* and must all consume the same amount of leisure and composite good and have the same value of time (u_ℓ/u_c), independent of work start-time. This implies that net-of-toll income and the sum of time spent in travel and at work must be the same for all individuals, independent of work start-time. Those who travel at the peak of the rush hour go into work less frequently than those who travel in the shoulders, and the equilibrium distribution of work start-times is such that their workday wage is higher by just enough that net-of-toll income is independent of work start-time. If, in contrast, individuals have a preference for work start-time *per se*, the corresponding properties of the efficient competitive equilibrium are more complicated, and depend *inter alia* on whether a more convenient work start-time is more complementary to leisure or consumption.

- *Incorporating multiple residential locations and land*

Thus far, it has been assumed that all individuals have a common workplace location and a different but common residential location, and the demand for land/housing has been ignored. The simple model variant can be extended relatively straightforwardly to the discrete version of the monocentric city, in which individuals have a common workplace location but different residential locations and choose lot size. All individuals essentially travel along the same corridor to work and have no route choice. Let n_{im} denote the number of individuals with work start-time i and residential location m. Define **n** so that it includes the extra index m. The private product function continues to depend only on the number of individuals by work start-time: $z_{im} = z_i(\{\sum_m n_{im}\})$. But the travel time function depends on the distribution of individuals by both work start-time and residential

location: $t_{im} = t_{im}(\mathbf{n})$. Solve for the efficient equal-utilities allocation. For this allocation, for all i and m calculate the net externality cost imposed by an individual with work start-time i and residential location m. This allocation can be decentralized by setting tolls $\{\tau_{im}\}$ equal to the corresponding net externality costs, redistributing toll revenues in equal lump-sum fashion, and allowing the market to determine the workday wage according to work-start time, $\{w_i\}$ (so that each worker is paid his private marginal product). Each individual will then face the correct marginal incentives in deciding where to live, how large a lot to consume, and when to start work. Given current technology, it would be infeasible to toll on the basis of work start-time and residential location. To toll on the basis of work start-time would require basing the toll on the time at which an individual arrives at the CBD, but this time provides no information concerning the individual's residential location. However, since equilibrium travel times can be solved for, the same tolling structure can be achieved by differentiating the toll according to departure time and residential location, which is feasible.

- *Multiple employment centers*

Now consider an extension of the simple model variant in which there is a single residential location but multiple possible employment locations, arrayed in a star network around the residential location, each joined to the residential location by a separate transportation link. Let n_{ie} denote the number of individuals with work start-time i and employment location e, t_{ie} denote the corresponding travel time, with $t_{ie} = t_{ie}(\mathbf{n})$ and z_{ie} denote the corresponding private product, with $z_{ie} = z_{ie}(\mathbf{n})$. Solve for the efficient equal-utilities allocation. For this allocation, for each i and e calculate the net externality cost

imposed by an individual with work start-time i at employment location e. This allocation can be decentralized by setting tolls $\{\tau_{ie}\}$ equal to the corresponding net externality costs, redistributing toll revenues in equal, lump-sum fashion, and allowing the market to determine the workday wages, $\{w_{ie}\}$ (so that each worker is paid his private product). Each individual will then face the correct marginal incentives with respect to both work start-time and employment location. Given current technology, this tolling structure would be feasible to implement since it simply requires tolling based on arrival time at each employment location.

- *A general network*

Now extend the simple model variant to incorporate a general travel network described by sets of nodes, links, and routes. Note that a route implies a residential origin and an employment destination. At each node there is a fixed supply of land, and the only land use is residential. Let n_{ir} denote the number of individuals with work-start time i and take route r from home to work, t_{ir} denote the travel time of an individual with work start-time i who takes route r, with $t_{ir} = t_{ir}(\{n_{ir}\})$, and z_{ie} denote the private product of a worker with work start-time i and employment location e (a worker's private product is independent of where he lives and the route he takes to work), with[18] $z_{ie} = z_{ie}(\{\sum_{r \in Re} n_{ir}\})$, where R_e is the set of routes having as their destination employment center e. Solve for the efficient, equal-utilities allocation. For this allocation, for each i and r, calculate the net externality cost imposed by a worker with work start-time i and route r. This allocation can be decentralized by setting tolls $\{\tau_{ir}\}$ equal to the corresponding net externality costs,

---

[18] Rosenthal and Strange [29] and Fu [12] provide estimates of how agglomeration externalities attenuate with distance.

redistributing toll revenues in equal, lump-sum fashion, and allowing the market to determine the workday wages, $\{w_{ie}\}$ (so that each worker is paid his private product) as well as land rents. Each individual will then face the correct marginal incentives with respect to his choices of residential location, employment location, travel route, and lot size.   Note that this toll is differentiated according to work start-time and route (remember that a route defines a residential origin node and an employment destination node). At least given current technology, it is infeasible to differentiate the toll on this basis. The best that can be done is to impose time-varying link tolls, but this entails insufficiently fine differentiation for decentralization of the efficient allocation. The calculation of the optimal time-varying link tolls would then be an exercise in the theory of the second best.

If one were to combine all these extensions, and extend the model in other ways as well (for instance, by treating firms' use of land as an input into production), one would find in general that even optimal time-varying link tolls, combined with optimal lump-sum toll revenue transfers differentiated according to an individual's residential and employment locations, would be insufficient to achieve a Pareto efficient allocation. The second-best optimal tolls would be computed taking into account both the congestion and agglomeration externalities.

4. Conclusion

This paper has made a simple but potentially important point related to urban auto congestion tolling. If it possible to impose road tolls but impossible to internalize positive agglomeration externalities, then the road tolls should be set taking into account the uninternalized agglomeration externalities. Since congestion tolling discourages travel, if less travel exacerbates the distortion associated with agglomeration externalities, which seems likely, then optimal road tolls should be set at levels below the corresponding congestion externality costs.

The cogency of the point hinges on whether it is easier to internalize traffic congestion or agglomeration externalities. Technologically, it is almost certainly easier to internalize congestion externalities since they are now measured with a reasonable degree of accuracy while agglomeration externalities remain amorphous and atmospheric and can be measured only indirectly, by for example spatial variation in rents and wages. However, urban congestion tolling has encountered considerable political opposition, while subsidizing firms with the aim of internalizing agglomeration externalities might be popular.

If imposing urban road tolls turns out to be easier than internalizing agglomeration externalities, determining by how much road tolls should be set below the corresponding congestion externality costs depends on the magnitude of uninternalized agglomeration externalities and on how sensitive the distortion associated with these externalities is to

the level of road tolls. Empirical work has been done measuring agglomeration economies but none attempting to estimate to what extent the agglomeration economies are uninternalized or how the distortion associated with the uninternalized agglomeration economies is affected by the amount of urban travel undertaken or the price of urban travel. Thus, this paper simply demonstrates that optimal urban road tolls *may be* substantially lower than the corresponding traffic congestion externality costs.

References

[1] Arnott, R., Unpriced transport congestion, Journal of Economic Theory 21 (1979) 294-316.

[2] Arnott, R., Staggered work hours with a dominant employer, Unpublished manuscript, Boston College, Chestnut Hill, MA, 2002.

[3] Arnott, R., A. Yan, The two-mode problem: Second-best pricing and capacity, Review of Urban and Regional Development Studies 12 (2000) 170-199.

[4] Arnott, R., M. Kraus, When are anonymous congestion charges consistent with marginal cost pricing? Journal of Public Economics 67 (1998) 45-64.

[5] Arnott, R., A. de Palma, R. Lindsey, A structural model of peak period congestion: A traffic bottleneck with elastic demand, American Economic Review 83 (1993) 161-179.

[6] Arnott, R., T. Rave, R. Schöb, Alleviating Urban Traffic Congestion, MIT Press, Cambridge, 2005.

[7] Boadway, R., R. Harris, A characterization of piecemeal second best policy, Journal of Public Economic 8 (1977) 169-190.

[8] Calthrop, E., S. Proost, K. van Dender, Optimal road tolls in the presence of a labor tax, Leuven, Center for Economic Studies, 2000.

[9] Cassidy, M., Traffic flow and capacity, in: R. Hall (Ed.), Handbook of Transportation Science, Kluwer Academic Publishers, Norwell, MA, 1999.

[10] Chipman, J.S., External economies of scale and competitive equilibrium, Quarterly Journal of Economics 84 (1970) 347-385.

[11] Duranton, G., D. Puga., Micro-foundations of urban agglomeration economies, in: J.V. Henderson and J-F. Thisse (Eds.), Handbook of Urban and Regional Economics, vol. 4, Elsevier, Amsterdam, 2004.

[12] Fu, Shihe, Smart café cities: testing human capital externalities in the Boston metropolitan area, Journal of Urban Economics 61 (2007) 86-111.

[13] Fujita, M., J-F. Thisse, Economics of Agglomeration: Cities, Industrial Location and Regional Economic Growth, Cambridge University Press, Cambridge, 2002.

[14] Guiliano, G., T. Golob, Staggered work hours for traffic management: A case study, Transportation Research Record 1280 (1990) 46-58.

[15] Henderson, J.V., The sizes and types of cities, American Economic Review 64 (1974) 640-656.

[16] Henderson, J.V., The economics of staggered work hours, Journal of Urban Economics 9 (1981) 349-364.

[17] Kanemoto, Y., Cost-benefit analysis and the second-best land use for transportation, Journal of Urban Economics 4 (1977) 483-503.

[18] Kraus, M., Discrete-time modeling of congestion on a network, Unpublished notes, 1998.

[19] Lévy-Lambert, H., Tarification des services à qualité variable: Applications aux péages de circulation, Econometrica 36 (1968) 564-574.

[20] Lighthill, M., G. Whitham, On kinematic waves II: A theory of traffic flow on long, crowded roads, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 229 (1955) 317-345.

[21] Lindsey, R. Staggered work hours and flextime. Unpublished manuscript. University of Alberta, Edmonton, 2002.

[22] Mas-Colell, A., M. Whinston, J. Green, Microeconomic theory, Oxford University Press, Oxford, 1995.

[23] Mayeres, I., S. Proost, Optimal tax and investment rules for congestion type of externalities, Scandinavian Journal of Economics 99 (1997) 261-279.

[24] Mills, E.S., D.M. de Ferranti, Market choices and optimum city size, American Economic Review Papers and Proceedings 61 (1971) 340-345.

[25] Moretti, E., Human capital externalities in cities, in: J.V. Henderson and J-F. Thisse (Eds.), Handbook of Urban and Regional Economics, vol. 4, Elsevier, Amsterdam, 2004.

[26] Mun, S., M. Yonekawa, Flextime, traffic congestion, and urban productivity, Paper presented at the Second International Symposium on "Structural Change in Transportation and Communications in Knowledge Society: Implications for Theory, Modeling and Data." Kyoto, Japan, 2000.

[27] Parry, I., A. Bento., Revenue recycling and the welfare effects of road pricing, Scandinavian Journal of Economics 103 (2002) 645-671.

[28] Pines, D., Y.Y. Papageorgiou. An Essay on Urban Economic Theory, Kluwer Academic Publishers, Boston, 1998.

[29] Rosenthal, S., W. Strange, Geography, industrial organization, and agglomeration, Review of Economics and Statistics 85 (2003): 377-393.

[30] Rosenthal, S., W. Strange, Evidence on the nature and source of agglomeration economies. in: J.V. Henderson and J-F. Thisse (Eds.), Handbook of Urban and Regional Economics, vol. 4, Elsevier, Amsterdam, 2004.

[31] Ross, S., S. Fu., Wage premium in employment clusters: Agglomeration economies or worker heterogeneity? Unpublished manuscript, Department of Economics, University of Connecticut, 2007.

[32] Small, K., Urban Transportation Economics, Harwood Academic Publisher, Philadelphia, 1992.

[33] Small, K., C. Winston, J. Yan., Uncovering the distribution of motorists' preferences for travel time and reliability: Implications for road pricing, Econometrica 73 (2005) 1367-1382.

[34] Solow, R., Congestion, density, and the use of land in transportation, Swedish Journal of Economics 74 (1972) 161-173.

[35] Strotz, R.H., Urban transportation parables, in: J. Margolis (Ed.), The Public Economy of Urban Communities, Resources for the Future, Washington, DC, 1965, pp. 127-169.

[36] Verhoef, E., Second-best congestion pricing in general networks: Heuristic algorithms for finding second-best optimal tolls and toll points, Transportation Research B 36 (2002) 707-729.

[37] Vickrey, W., Congestion theory and transport investment, American Economic Review Papers and Proceedings 59 (1969) 251-260.

[38] Wheaton, W., Price-induced distortions in urban highway investment, Bell Journal of Economics 9 (1978) 622-632.

[39] Wilson, J., Optimal capacity in the presence of unpriced congestion, Journal of Urban Economics 13 (1983) 337-357.

[40] Wilson, P., Wage variation resulting from staggered work hours, Journal of Urban Economics 24 (1988) 9-26.

[41] Yoshimura, M., M. Okumura., Optimal commuting and work-start-time distribution under flexible work hours system on motor commuting, Proceedings of the Eastern Asian Society for Transportation Studies 2 (2001) 455-469.