Congestion pricing on a road network:
A study using the dynamic equilibrium simulator METROPOLIS

André de Palma
Université de Cergy-Pontoise and ENPC
Senior member of the Institut Universitaire de France
THEMA, 33 Bd du Port
F-95011, Cergy-Pontoise
FRANCE
andre.depalma@eco.u-cergy.fr

Moez Kilani
ISG Sousse
Université du Centre de Sousse
TUNISIA
m.kilani@planet.tn

Robin Lindsey
Department of Economics
University of Alberta
Edmonton, Alberta
CANADA T6G 2H4
rlindsey@ualberta.ca

First version: November 1, 2003

Revised version: August 27, 2004

ABSTRACT

The gradually accelerating pace at which tolls are being implemented or actively considered around the world suggests that road pricing is an idea whose time may finally have come. Nevertheless, various design considerations must be addressed and public acceptability hurdles overcome if road pricing is to become widespread. And to measure the efficiency gains and welfare distributional effects of road pricing accurately, models need to account for potential behavioural responses to tolls – and in particular adjustments in trip timing.

This paper analyzes some road pricing schemes using the dynamic network simulator METROPOLIS: a tool that treats endogenously departure-time decisions as well as mode and route choices of individual travelers. Simulations are conducted for a stylized urban road network that consists of radial arterials and circumferential ring roads, and with trip origins and destinations that are distributed throughout the network. Six types of link-tolling schemes are analyzed: (1) the system optimum which can be supported approximately by imposing time-varying step tolls that eliminate queuing, (2) a set of comprehensive flat (time-independent) tolls, (3, 4) second-best flat and step tolls for a toll cordon, and (5, 6) second-best flat and step tolls within a charge area.

Two results show the superiority of step tolls over flat tolls. First, step tolls easily outperform flat tolls in terms of welfare gains while inducing a smaller shift of trips from auto to transit. Second, step tolls generate smaller revenues than do flat tolls, and consequently have more favourable distributional impacts on travelers.

1 INTRODUCTION

The theory of road pricing dates back to Pigou (1920) and the literature is growing rapidly as practice starts to catch up.[1] Pigouvian formulae for first-best congestion tolls are straightforward to derive for deterministic user equilibrium in static models (Beckmann *et al*., 1956; Dafermos, 1973; Yang and Huang, 1998) and the formulae also apply to stochastic user equilibrium (Yang, 1999). However, several complications arise in designing and evaluating real-world road-pricing schemes.

First, standard Pigouvian formulae apply only if first-best conditions hold; *i.e*. only if pricing of externalities is ubiquitous and unconstrained. Yet comprehensive congestion pricing is unlikely to be implemented in the near future, and in the interim tolling will probably be confined to the most heavily traveled roads and congested areas. Furthermore, congestion and other external traffic costs vary by road type, vehicle and driver characteristics, but limitations on tolling technology and legal restrictions currently prevent sufficient toll differentiation to price all trips at marginal social cost. Second-best rules for congestion pricing when parts of a network cannot be tolled or other distortions exist are difficult to derive, and they require information on demand and cost elasticities that is usually not readily available.[2] And empirical applications have shown that second-best tolls can differ markedly in size from their first-best Pigouvian counterparts, and can even be negative.

---

[1] Reviews of the literature on road pricing and surveys of current practice are found in Small and Gómez-Ibáñez (1998), Lindsey and Verhoef (2001) and Transportation Research Board (2003).

[2] Insights into second-best pricing using models that feature two or three routes or travel modes are found in Verhoef *et al*. (1996), Mayet and Hansen (2000), Arnott and Yan (2000), Small and Yan (2001) and Verhoef and Small (2004) *inter alios*. More general networks are featured in Yang and Lam (1996), Parry and Bento (2000) and Verhoef (2002).

Second, in order to predict the effects of road pricing schemes accurately it is necessary to account for trip-timing decisions. Optimal time-varying tolls for a single road link have been solved with endogenous trip timing based on schedule delay costs (the costs of arriving early or late) and for various specifications of the congestion function (Vickrey, 1969; Henderson, 1974; Arnott *et al.*, 1990; Laih, 1994; Yang and Huang, 1997; Arnott and Kraus, 1998). Tolling one or both of two routes connecting a single O-D pair is studied using the bottleneck model by Braid (1996) and De Palma and Lindsey (2000). And De Palma, Lindsey and Quinet (2004) have recently applied the bottleneck model to the tolling of subsets of links on a simple hub-and-spoke network.

Yet time-of-day pricing has yet to be analyzed on medium- or large-scale networks using a fully dynamic model. Tolling of general networks is analysed in Carey and Srinivasan (1993) and Yang and Bell (1997), and time-of-day pricing is simulated on a road network in Hong Kong by Hau *et al*. (2003). However, trip timing is exogenous in these studies. Viti *et al*. (2003) treat departure-time choice endogenously, but their objective function includes travel time costs but not schedule delay costs. Safirova *et al*. (2004) study congestion pricing for Washington, D.C., using a discrete-choice demand structure with multiple behavioural dimensions and a relatively detailed network specification. But their application includes only three broad time periods, and the tolls they consider are the same across time periods and freeways.

A third complication in the design of practical road pricing schemes is that acceptability barriers pose a significant barrier to tolling generally, and congestion charging specifically. It appears to be practically difficult – if not impossible – to design a tolling scheme *cum* revenue use plan (or indeed any policy package) in which all individuals end up better off (Nakamura and Kockelman, 2002). An assessment of who gains and who loses, and by how much, is therefore warranted as part of the evaluation process. The welfare-distributional effects of road pricing have been studied using the traditional one-link supply and demand-curve approach of Walters (1961) for many years.[3] But because the politics of road pricing tends to play out at a local level, it is important to examine its geographical impacts across regions. Some recent studies have done so (Yang and Zhang, 2002; Safirova *et al*., 2004; Santos and Rojey, 2004) but not with time-varying tolls in a fully dynamic model.

In summary, existing studies of road pricing all lack one or more of the following elements: (1) modeling of trip-timing decisions at a fine level of temporal resolution, (2) accounting for trip-timing preferences in the welfare assessment, (3) tolling by time of day, and (4) road networks of practical interest. A simulation approach is required to encompass all these dimensions. This paper studies time-of-day road pricing using the dynamic network equilibrium simulator METROPOLIS. METROPOLIS is a fully dynamic model that treats endogenously mode, departure time and route choices, and it is operational on large-scale networks. Because METROPOLIS tracks the movements of individual vehicles, it permits welfare analysis at the level of individual travelers. METROPOLIS has recently been used to analyze road pricing in Ile de France (De Palma, Lindsey and Kilani, 2003). However, given the size of the Ile de France network it was only feasible to simulate a few pricing regimes that were not fully optimized, and the welfare-distributional impacts were assessed only at an aggregate level.

The current paper builds on De Palma, Lindsey and Kilani (2004) by considering a stylized urban road network. Six types of link-tolling schemes are analyzed: (1) the system optimum which can be supported approximately by imposing time-varying step tolls that eliminate queuing on every link, (2) a comprehensive set of second-best flat (time-independent) tolls, (3, 4)

---

[3] Literature reviews are found in Richardson and Bae (1998) and Santos and Rojey (2004).

second-best flat and step tolls for a toll cordon, and (5, 6) second-best flat and step tolls within a charge area.

Each of these tolling regimes is of interest. The first-best optimum (Regime 1) is a standard that serves as a benchmark for other tolling regimes and establishes an upper bound on the potential welfare gains from tolling. Comprehensive flat tolls (Regime 2) are of interest primarily to assess how much time-of-day tolling boosts welfare gains and how it affects the burden of tolls borne by travelers. The other four regimes involve tolling only part of the road network which, as noted above, will continue to be the rule for some time to come. Cordon tolls are currently in use in Norway, and have been proposed elsewhere.[4] Although time-of-day pricing of cordons is technologically straightforward, practical and acceptability constraints may limit the degree of time variation that will eventually be adopted. It is therefore of interest to consider both flat tolls and step tolls with a limited degree of time variation.

Area-based schemes differ from cordon tolls in that vehicles are charged for moving within an area rather than just for entering it.[5] London's congestion pricing scheme entails a flat charge of £5 between 7:00 and 18:30 on weekdays. A comparison of flat and step tolls for area schemes is motivated by much the same practical considerations as for cordon tolls. And a parallel comparison of flat and step tolls for cordons and area schemes is also of interest to assess their relative merits. To facilitate comparison, the locations of the cordon toll and the perimeter of the area charge are assumed to be given and coincident. The study therefore bypasses the question of where tolling points should be located.

---

[4] Toll cordons have been widely studied; recent contributions include May *et al*. (2002), Shepherd and Sumalee (2002), Mun *et al*. (2003), Santos and Rojey (2004) and Zhang and Yang (2004).

[5] Area-based pricing schemes are reviewed in Transportation Research Board (2003). Hyman and Mayhew (2002) use a simulation model to compare area-based charging with destination-based and distance-based schemes.

Section 2 following provides an overview of METROPOLIS, describes the network and demand specification used for the simulations, and discusses the challenges of computing optimal time-varying tolls. Section 3 briefly describes the equilibrium in the absence of tolls. Section 4 assesses in some detail the welfare gains and distributional impacts of the tolling schemes. Concluding remarks are provided in Section 5.

2 THE MODEL

2.1 Travel demand in METROPOLIS

The dynamic network equilibrium simulator METROPOLIS[6] treats endogenously mode, departure-time and route choices on a road network at the level of individual travelers.[7] A two-stage nested logit model is used as shown in Figure 1. Mode choice is described at the outer nest. The generalized systematic cost of public transport, which is assumed to be exogenous and independent of time of day, is given by

$$C_P = C_{P0} + \alpha_P \mathrm{T}_P \ ,$$

where T_P denotes travel time by transit, α_P is the unit cost of transit travel time, and C_{p0} is a penalty that includes the fare, and the costs of walking time, waiting time, average schedule delay, physical discomfort, loss of privacy, *etc*.

Unlike for transit, trip-timing preferences are central for the auto mode. Following Vickrey (1969) travelers are assumed to have individual-specific preferred arrival time windows at their destination, and incur a schedule delay cost if they arrive earlier or later. The generalized

---

[6] A detailed description of the system architecture and previous applications of METROPOLIS are found in de Palma and Marchal (2002a, 2002b).

[7] Numbers of trips, destination and vehicle occupancy (fixed at one) are exogenous, and there is no trip chaining.

systematic cost of an auto trip is additively separable in travel time costs, schedule delay costs and toll payment[8], and is given by the function

$$C_A(t) = \alpha_A \mathrm{T}_A(t) + \beta Max\left[0,\ t^* - \Delta - t - \mathrm{T}_A(t)\right] + \gamma Max\left[0,\ t + \mathrm{T}_A(t) - t^* - \Delta\right] + \tau,$$

where t denotes departure time, T_A auto travel time, α_A the unit cost or value of auto travel time (VOT), t^* the midpoint of the on-time arrival window, Δ the half-width of the window, β the unit cost of arriving early, and γ the unit cost of arriving late. The amount paid in tolls, τ, depends on the choice of departure time and route.

Departure time choice is described in the inner nest shown in Figure 1 by a standard continuous logit specification. In the simulations featured in the paper the permitted departure-time interval is assumed to be 6:00 AM to 12:00 AM as shown. Route choice is governed by a heuristic based on a generalization of Wardrop's minimum-cost principle, which closely approximates the minimum generalized cost. Route choice decisions within a day are revised at road network intersections. A day-to-day adjustment process with exponential learning by drivers governs changes in mode choice, departure time and route choice, and guides the system towards a stationary state. Details about the learning process are found in de Palma, Marchal and Nesterov (1997). Simulation experiments that were performed with various networks show that various adjustment processes lead to the same stationary solution.

The road network and travel-demand specifications used in the simulations are described in Sections 2.2-2.6. Sections 2.7 and 2.8 explain the procedures that were employed to compute social surplus and to optimize tolls.

---

[8] Vehicle operating costs are ignored because their effects can (approximately) be subsumed either in the value of auto time or the fixed penalty for taking public transportation.

2.2 The road network

The laboratory road network, shown in Figure 2, is defined for a circular city with a radius of 16 km. There are four concentric ring roads: Rings 1-4, spaced 4 km. apart. For ease of reference, the city centre is called Ring 0. Eight arterial roads, uniformly orientated around the compass, run in both directions between Ring 0 and Ring 4. Arterial links that connect Ring x to Ring x-1 are called "*Inx*" links, and arterial links that connect Ring x-1 to Ring x are called "*Outx*" links, x=1,…,4. Except for the *In1* and *Out1* links, both inbound and outbound links comprise two lanes, each lane with a flow capacity of 2,000 vehicles/hr and a free-flow speed of 70 km/hr. To capture elevated levels of traffic congestion in the inner city, *In1* and *Out1* links have a lower total flow capacity of 3,000 vehicles/hr, and a lower speed limit of 50 km/hr. All links on the ring roads have a flow capacity of 2,000 vehicles/hr and a speed limit of 50 km/hr. Congestion on all links takes the form of queuing. Thus, if the arrival rate of vehicles at a link exceeds its flow capacity, a queue develops.[9]

2.3 The transit network

Public transportation is provided on a network topologically congruent with the road network at a constant (usage-independent) speed of 40 km/hr.

2.4 Origin-destination matrix

Trips originate and terminate at zones that are joined to each of the 33 network nodes by short congestion-free links (connectors). A static O-D matrix[10] describes trips that are interpreted to be morning home-to-work trips. Each zone generates 8,000 trips; the total number of trips is therefore 264,000. The number of trips from a zone to each of the other 32 zones is an exponentially decreasing function of the free-flow automobile travel time.[11] Trip destinations are therefore concentrated towards the centre because of its greater average accessibility. The number of trips per O-D pair has a mean value of 250, and a range of 123 to 660.

2.5 Travel demand parameters

Travel-demand parameter values are listed in Table 1. As explained in Appendix A, parameter values were chosen to be representative of developed countries and to yield plausible results for the no-toll equilibrium.

2.6 Computational considerations

METROPOLIS employs a day-to-day adjustment process with learning by individuals. Depending on the initial conditions specified, about 50 iterations (days) are required to approach a stationary state. Because individual mode and departure-time choices are reviewed on an ongoing basis, travel conditions continue to oscillate thereafter with consequent fluctuations in social surplus and other summary statistics.[12] To make the welfare analysis more accurate, a further 75 iterations were run and average values for social surplus and toll revenue were

---

[9] METROPOLIS can model horizontal queuing and spillback, but this feature was not enabled for the simulations.

[10] The matrix is static because trip timing is endogenous.

[11] Exponential decay is at a rate –0.07/min based on estimates for home-based work trips in Krishnamurthy and Kockelman (2003, Table 3). Conceptually, it would be preferable to base demand on the equilibrium generalized travel cost (or, even better, traveler's surplus) between nodes. But since equilibrium depends on the O-D matrix, it would be necessary to adjust the O-D matrix iteratively. Furthermore, the O-D matrix would slowly adjust after the introduction of road pricing, and the adjustment would differ for each tolling regime. Because of these complications the O-D matrix is treated as exogenous. In support of this Eliasson and Mattsson (2001) conclude that the effects of tolls on location choices are secondary compared to their effects on traffic volumes, modal splits and trip distances.

[12] It is an open question whether the day-to-day oscillations generated by METROPOLIS are representative of real-world behaviour.

computed over these iterations.[13] These values will be referred to as "equilibrium" values even though an equilibrium is never reached exactly.

With 264,000 travelers, a simulation of 125 iterations takes about an hour on a 2 GHz Pentium 4 personal computer. Given the large number of simulations required to identify optimal tolls for some of the tolling regimes (see Section 2.8), this computation time would be excessive. Fortunately, computation time can be reduced by exploiting a property of the model. In the single bottleneck model (Vickrey, 1969) user costs are homogeneous of degree zero in demand and capacity; *i.e.* changing the number of travelers and the capacity of the bottleneck by the same proportion leaves travel costs unchanged. This is also a property of METROPOLIS: changing the level of trip demand for each O-D pair and the capacities of each link by the same proportion leaves travel costs approximately[14] unchanged. By experimentation it was determined that demand and capacities could be reduced to one tenth of their "actual" size with very little change in congestion, social surplus or other summary statistics. Computation time per simulation was correspondingly reduced by about a factor of 10.

2.7 Tolls and tolling regimes

In METROPOLIS vehicles pay tolls when they enter links. The toll on each link can vary in steps as frequently as every minute. The toll is flat if it is constant throughout the permitted travel period. Negative tolls are ruled out as inadmissible because of the possibility that drivers

[13] Social surplus is the sum of net consumer's surplus and toll revenue. Consumers' surplus is computed as *per* the nested logit model, with the inner logsum evaluated over departure times and the outer logsum over auto and public transportation. The average values of social surplus and toll revenue were relatively insensitive to the number of iterations used for averaging. To test this, averages were computed for each of the six optimized tolling regimes for 10, 25, 50, 75, 100 and 125 iterations. For the welfare gain the range of the computed values varied from 0.5% to 1.7% of the mean values with 75 iterations. Percentages were even smaller for revenue.

[14] Homogeneity holds only approximately for two reasons. First, because trip demands must be integer-valued for each O-D pair, demands need to be rounded to the nearest integer after re-scaling. Second, mode and departure-

with low unit costs of travel time and/or schedule delay choose to cycle around the network in order to make money.[15]

A rather large number of link-tolling regimes can be envisaged on the laboratory road network including single and multiple cordons, area charges, tolling of arterials or ring roads, destination/parking pricing, and combinations of schemes.[16] Computation time and space constraints preclude an exhaustive examination, and attention is restricted here to six particular schemes. The first is the system optimum in which time-varying tolls are levied on all links that are susceptible to congestion. The second regime applies flat (time-independent) tolls to the same set of links. It is second best in terms of its lack of time variation. Tolling regimes 3 and 4 apply flat and step tolls to a cordon defined by Ring 2 of the network.[17] Tolls are charged for crossing inside Ring 2 by tolling each of the eight *In2* links using the same toll or toll schedule. (No charge is levied for driving around Ring 2 itself.) For reasons discussed in Section 2.8, the step toll is applied in four half-hour intervals over the period [6:30, 8:30].

Tolling regimes 5 and 6 apply flat and step tolls to the area within Ring 2. These schemes are implemented by applying the same toll to the eight *In2* links and the connectors joining the nine residential zones on Rings 0 and 1 to the road network. Residents living on Rings 0 and 1 who drive use the connectors only once, and never use the *In2* links. Residents living on Rings 2-4

---

time choices are invariant to scale only in a large-numbers sense because individual-specific preferences are randomly assigned and therefore do not scale exactly.

[15] The non-negativity constraint is restrictive inasmuch as negative tolls can be second-best efficient (Braid, 1996; Yang and Huang, 2003; de Palma, Lindsey and Quinet, 2004). To see this, suppose A and B are two routes connecting the same O-D pair. If B is more congested than A, and B cannot be tolled, then a negative toll on A is warranted if the benefits from reduced congestion on B outweigh the losses from increased congestion on A. This is more likely if a time-varying toll is applied on A because the real costs of a given number of trips on A are then reduced.

[16] A preliminary assessment and comparison of flat tolling versions of many of these schemes was carried out in an earlier version of this paper.

[17] Ring 2 yields a higher welfare gain than do cordons defined by the other three ring roads, and was selected for this reason.

never use the tolled connectors, and need to drive on an *In2* link at most once. Therefore, no one pays the area charge more than once. The step area charge is applied over the same four half-hour intervals as the step cordon toll.

2.8 Toll optimization

For all six tolling regimes tolls were chosen to maximize social surplus. Since congestion is the only external cost of auto travel, and since there are neither economies nor diseconomies of scale in public transportation, the only role for tolls is to price congestion. Nevertheless, solving for optimal tolls is complicated by several considerations. First, there is no automated procedure for computing or setting tolls in METROPOLIS, and toll values have to be entered manually for each simulation. Second, unlike with static models simple Pigouvian formula cannot be used to set tolls in a dynamic model. And third, optimal tolls generally differ for each type of link. To tackle these difficulties, different procedures were used for the system optimum, the flat-tolling regimes and the step-tolling regimes. Each procedure is described in turn.

System optimum (*Regime 1*)

In the bottleneck model with a single link it is optimal to eliminate queuing, and this can be accomplished by applying a continuously time-varying toll (the ”fine” toll). First-best analysis suggests that if all links can be tolled, a ”no-queue” tolling policy is also optimal for each link on a network. However, three qualifications apply. First, it may in fact not be optimal to eliminate queuing everywhere on a network.[18] Second, it would probably be impractical to vary tolls

[18] De Palma and Jehiel (1994) demonstrate this by presenting examples in which schedule delay is costly relative to travel time. To minimize total costs in such a case it may be optimal to load some travelers onto the network quickly and “store” them in queues in order to clear the way so that other travelers can complete their trips on time.

continuously since it is unlikely that drivers could time their trips with corresponding precision.[19] Third, due to daily fluctuations in demand and capacity, link inflows vary from day to day in practice, and it would be possible to eliminate queuing consistently only by setting tolls at a high level that would result in underutilization of capacity on most days. Indeed, due to the oscillations of METROPOLIS this complication is manifest in the simulations.

These three caveats notwithstanding, no-queue tolling is a simple and intuitive policy that may in practice yield most of the potential welfare gains from tolls. The procedure that was used involves an iterative adjustment of the toll on each link using the formula[20]:

$$\tau_t^{i+1} = Max\left[0,\ \tau_t^i + \lambda \cdot \alpha_A \cdot tt_{Free} \frac{Occ_t^i - Occ_{\max}}{Occ_{\max}}\right], \qquad (1)$$

where Occ_t^i is occupancy (*i.e.* number of vehicles) on the link in time-of-day interval t on the last iteration of simulation i, $Occ_{\max}$ is the maximum occupancy of the link above which queuing occurs[21], tt_{Free} is free-flow travel time on the link, and $\lambda \in (0,1)$ is an adjustment factor. According to eqn. (1) the toll for interval t is raised for simulation i+1 if the occupancy for simulation i exceeded capacity (*i.e.* a queue was present), and lowered if there was excess capacity. Since negative tolls are ruled out, a minimum toll of zero is imposed.

The tolls were set in five-minute intervals – which is the smallest time interval at which information on occupancies can be extracted from METROPOLIS after a simulation. Occupancies in the no-toll equilibrium were used to establish initial conditions, although any

---

[19] Also, as noted below the algorithm used for no-queue tolling uses information that is available from METROPOLIS output only in five-minute intervals.

[20] The formula is inspired by Vickrey (1967, p.127). It can be employed on any set of links. De Palma, Kilani and Lindsey (2004) apply it to the tolling of individual links and cordons on the network in Figure 2, and Hau *et al*. (2003) use a similar procedure to toll a tunnel in Hong Kong.

[21] The formula for $Occ_{\max}$ is provided in Appendix B.

starting values could be used. Successive simulations using eqn. (1) were begun with the adjustment parameter set at $\lambda = 1$ and continued until a decrease in social surplus occurred. A second set of iterations was then conducted with $\lambda = 0.5$ and continued until social surplus fell again.

The no-queue tolling algorithm is a heuristic that is (approximately) optimal only for the system optimum. The other five tolling regimes considered are second best in terms of limited time variation and (except for Regime 2) also network coverage, and other optimization procedures were used.

Flat cordon tolls and area tolls (Regimes 3 and 5)

The flat-cordon and flat-area tolling regimes each employ a single toll level. To derive it several initial simulations were run for a wide range of toll values. A quadratic function or response surface[22] was then estimated of the form

$$G_i = a\tau_i + b\tau_i^2 + \varepsilon_i, \quad i = 1,2,... \tag{2}$$

where G_i denotes the welfare gain relative to the no-toll equilibrium for simulation *i*, τ_i is the toll, *a* and *b* are coefficients, and ε_i is an error term. Given coefficient estimates $\hat{a}$ and $\hat{b}$, the fitted optimal toll is $\hat{\tau} = -\hat{a}/(2\hat{b})$ and the estimated maximum welfare gain is $\hat{G} = -\hat{a}^2/\left(4\hat{b}\right)$. Further simulations were then performed using tolls set at, or close to, $\hat{\tau}$. Simulations were stopped when neither $\hat{\tau}$ nor $\hat{G}$ changed appreciably between successive simulations.

Step cordon tolls and area tolls (Regimes 4 and 6)

Step tolls are defined by the number of steps, the timing of the steps and the level of the toll

[22] May *et al*. (2001) provide guidelines for using response surfaces. A quadratic function worked more quickly and reliably than did splines.

for each step. For a given number of steps the toll schedule can be optimized in three ways: (a) by optimizing both the timing of steps and the toll levels, (b) by fixing pre-specified toll levels and optimizing only the timing, and (c) by fixing the timing of the steps and optimizing only the levels. Approach (a) was tested for the cordon toll with a single non-zero toll interval. Although this toll is defined by just three parameters (time on, time off, and level) over two-dozen simulations were required to find an approximate optimum and the quadratic response surface did not fit well. Approach (a) was therefore abandoned. Approach (c) was chosen over Approach (b) on the grounds that where variable pricing has been implemented, tolls are usually changed at times that are easy to remember; *i.e.* on the hour or half past the hour.[23]

Following Approach (c), the step toll was levied in four half-hour steps over the period [6:30, 8:30] during which most trips take place. An initial set of simulations was performed using alternative combinations of the four toll levels that were selected using a simple fractional factorial design. A quadratic response surface was then fitted of the form

$$G_i = \sum_{t=1}^{4}\left(a_t\tau_i(t) + b_t\tau_i^2(t)\right) + \sum_{s=1}^{3}\sum_{t=s+1}^{4} c_{st}\tau_i(s)\tau_i(t) + \varepsilon_i, \quad i = 1,2,... \tag{3}$$

where $\tau_i(t)$ is the toll for step *t* in simulation *i*, the a_t, b_t and c_{st} are coefficients, and other variables are defined as for eqn. (2). Formulae for the optimal tolls and maximum surplus were then derived from the coefficient estimates using the symbolic software MAPLE©. The rest of the procedure was analogous to that used for the flat tolls.

Comprehensive flat tolls (*Regime 2*)

Regime 2 involves setting flat tolls on all the congested links. As noted in Section 4, this turned out to involve all 12 link types except *In4*, *Out4* and *Ring 4*. A quadratic response surface

[23] Setting tolls in round values, such as multiples of \$0.25 or \$1.00, hastens payment at conventional toll booths, but not with electronic tolling.

approach analogous to the one used for Regimes 4 and 6 was used. Unlike for eqn. (3), computation constraints militated against fitting a response surface with all two-way interaction terms. The interaction terms were restricted to pairs of links for which traffic flows were judged to be strongly interdependent, such as link types *In1* and *In2*.

3 NO-TOLL EQUILIBRIUM

Some properties of the no-toll equilibrium are listed in Table 2. About 70% of trips are made by automobile, and the elasticity of auto trips with respect to generalized travel cost is –0.32 which is representative of medium-run elasticities for morning peaks. The average auto trip is 15 km. at a speed of about 40 km/hr. and lasts 22 minutes. The congestion index – defined to be the ratio of queuing delay to free-flow travel time on the same route – indicates that travel time is nearly 40% higher than under free-flow conditions. Average generalized trip cost is $5.16, of which queuing delay accounts for $1.03 and schedule delay $1.47. Consumer's surplus is negative because only the costs of travel are incorporated.

About 29% of drivers arrive on time and 57% arrive early; only 14% are late due to the relatively high unit cost of late arrival. Aggregate auto travel distance (for the 1/10 scale network) in the last row of the table is a surrogate measure for emissions, noise and accidents, and is reported for comparison with the tolling regimes.

Figures 3-5 display the evolution of the congestion index by type of link over the first three hours (6:00-9:00) of the simulation period when congestion is manifest.[24] Congestion is most severe on the innermost (*In1*) links. At 7:15 the congestion index reaches 3.5; given a free-flow

[24] Despite symmetry of the network, flows and congestion indexes vary for links of the same type on account of idiosyncratic traveler preferences and oscillations about the steady state. The indexes shown are averaged over links of the same type.

travel time of 4.8 minutes on the *In1* links this implies a delay of nearly 17 minutes. Moving outwards towards the periphery, congestion falls progressively and peaks earlier because jobs are concentrated in the centre and traffic is primarily inbound. Congestion on the outbound links (Figure 4) is much lower than on the inbound links, and peaks a few minutes later. The ring roads display an intermediate level of congestion (Figure 5) and a similar time pattern.[25]

4 TOLLING REGIME RESULTS

4.1 Regime 1: System optimum

Results for the system optimum (SO) are reported in Table 3 with selected statistics for the no-toll equilibrium (NTE) listed in the column to the left for ease of comparison. In the SO all links are tolled except for the outermost arterials and ring road (link types *In4*, *Out4* and *Ring 4*). Tolling reduces the auto share of trips by only 2% relative to the NTE and total travel distance is virtually unchanged, but the overall congestion index drops to less than 13% of the NTE level.

To provide a sense of magnitudes the welfare gain and toll revenue are reported on an annual per capita basis assuming 250 trips per year. Interestingly, the welfare gain nearly matches the toll revenue (see the row "Welfare gain/revenue"). This result contrasts with the findings of most static models, but it is consistent with the single-link bottleneck model in which welfare gain and toll revenue are equal. Since the SO by definition achieves the highest feasible level of welfare[26] it scores 100% in terms of relative welfare gain.

As noted earlier the welfare gain and toll revenue values are averaged over the last 75 iterations of each simulation. The plus/minus values reported in Table 3 are the standard

[25] As noted in Section 2.2, spillback was not activated. If the density of queued vehicles is assumed to be 250/km. per lane, spillback would occur intermittently on the *In1* links between 7:05 and 8:00.

deviation and (in brackets) the standard deviation of the estimate assuming serial independence between iterations.[27] The standard deviation is considerably higher for the welfare gain than for revenue – a pattern also found for the other tolling regimes.

The last four rows of Table 3 provide several measures of acceptability:

1. *Gainers no rebate*: the fraction of travelers who gain before usage of revenue is accounted for;
2. *Gainers 100% rebate*: the fraction of travelers who gain if all revenue is rebated as a uniform per capita lump sum;
3. *Rebate*: *50% gain*: the fraction of revenue that needs to be rebated as a uniform per capita lump sum in order for a majority of travelers to gain;
4. The standard deviation of change in consumer's surplus.

Each of these measures is calculated for all travelers, including those who take public transportation. The first three measures assess changes in absolute well-being, while the fourth measure gauges changes in relative well-being or inequity. In the SO more than a quarter of travelers gain before accounting for use of revenue, and the proportion rises to over 4/5 if all revenue is rebated.

Table 4 provides a more detailed assessment of the welfare impacts by disaggregating travelers by origin and destination ring.[28] Each cell of the table refers to an origin-destination ring pair. To economize on space and also to facilitate comparison, the results for tolling Regime 2 are included in the two right-hand entries of each cell; these should be ignored for the moment.

---

[26] Subject to the provisos noted in Section 2.8.

[27] To the extent that the oscillations are representative of real-world behaviour, the standard deviation can be interpreted as a way to measure fluctuations in welfare gain and toll revenue that would actually be experienced.

[28] This segmentation is imperfect since travel distance between rings varies with the compass orientations of the origin and destination zones. Disaggregation to the level of zone pairs would be impractical.

As the key in the top-left cell of the table indicates, entry (a) is the percentage of travelers who gain consumer's surplus, and entry (b) is the mean change in consumer's surplus per trip. The final row of Table 3 reports results aggregated over destinations, the final column aggregates over origins, and the bottom-right cell refers to all travelers.

Beginning with origin Ring 0 and destination Ring 1, Table 3 indicates that while only 39% of the travelers between these rings gain from the SO toll, their average change in consumer's surplus is \$0.33 per trip. A similar pattern applies for residents of Ring 0 traveling to the other rings further out. Thus, on balance only a minority of city-centre residents gain, but their benefits are skewed sufficiently to the right that average consumer's surplus rises.

From the remaining columns of Table 4 it is evident that a majority of travelers in every cell lose from the toll, but average losses are never large and in some cells there is a small gain. Overall, there is rather little geographical variation in gains and losses, and this might suggest that network effects are not a major factor in determining the welfare impacts of tolls. However, this is not true of the other tolling regimes.

4.2 Regime 2: Comprehensive flat tolls

Similar to the SO, flat tolls in Regime 2 are levied on all links except the outermost arterials and ring road. As Table 3 shows, the toll on the *In2* links (\$2.15) exceeds the toll on the *In1* links (\$0.73) despite the fact that congestion delays in the NTE are much longer on the *In1* links.[29] The flat tolls reduce auto travel by more than do the SO tolls because queuing is suppressed less, and auto travel is correspondingly less efficient.[30] The flat tolls also impose a significantly

[29] This is possible because a majority of traffic on the *In1* links also uses the *In2* links, and tolling one link in a series alleviates congestion on the other as well.
[30] This is consistent with analytical results for the single-link bottleneck model (Braid, 1989, Arnott *et al*., 1993).

greater toll revenue burden on travelers[31] but they yield less than half the welfare gain per-capita. As a consequence, the welfare gain per dollar of revenue is much lower and all four measures of acceptability are less favourable.

Returning to Table 4 and inspecting entries (c) and (d) it is apparent that the welfare-distributional impacts of the flat tolls differ from those of the SO toll. First, travelers in almost all of the cells lose out, and in several cases the reduction in consumer's surplus approaches $2. Second, the welfare loss tends to rise with the average distance between origin and destination ring, thereby indicating that the effects of the flat toll are qualitatively similar to those of a distance toll.

The inferior performance of the flat tolls relative to the SO tolls in terms of welfare gain and acceptability clearly derives from their inability to target queuing by time of day. However, contrary to what might be thought the flat tolls do have a significant impact on the distribution of departure times. This is demonstrated in Figure 6 by plotting changes in the normalized cumulative distributions of departure times for the flat and SO tolling regimes. The SO tolls induce a small acceleration of departures between 6:00 and 6:45, and postponement thereafter with a maximum shift of about 8% at 7:30. By contrast, the flat tolls delay departures throughout the travel period with a maximum shift of over 11%. Although the flat tolls do not provide a direct incentive for drivers to retime their trips, they do reduce congestion appreciably and they also cause a larger shift of traffic to transit than do the SO tolls.

4.3 Regimes 3 and 4: Flat and step cordon tolls

The cordon tolls in Regimes 3 and 4 are paid on the *In2* links. The optimal flat toll is about $5 (see Table 3), which is more than twice the toll paid on the same *In2* links in Regime 2.

---

[31] Averaged over the period 6:00 AM – 9:00 AM the SO tolls are lower than the flat tolls on all links except *In1*.

Coincidentally, it is about equal to the sum of the tolls charged in Regime 2 on the *In1*, *In2* and *In3* links – thereby showing how second-best tolls are adjusted upward when complementary links are untolled. The flat cordon toll generates only about half the welfare gain and toll revenue from the comprehensive flat tolls, but it has rather more favourable welfare-distributional effects.

The step cordon toll is higher than the flat cordon toll during the 7:00-7:30 interval, and lower at other times.[32] It yields nearly twice the welfare gain while raising less revenue and, consequently, imposes a lesser burden on travelers. The welfare gain is comparable with Regime 2, which indicates that the disadvantage of the step cordon toll in terms of limited network coverage is largely offset by its time variation.

Table 5 summarizes the disaggregate welfare impacts of the cordon tolls using the same format as Table 4. The main winners for both flat and step tolls are residents of Ring 1, who benefit from congestion relief without having to pay the toll. Residents of Rings 2-4 all lose from the toll, with the biggest losses incurred by residents of Ring 2 who are more likely to work inside the cordon and therefore to pay the toll. Losses also predominate for travelers who work on Ring 1, many of whom live on or outside Ring 2.

The pattern of gains/losses for residents of Rings 1-4 is consistent with intuition: those who pay the cordon tolls are worse off, while those who do not pay gain. And in every cell travelers fare better with the step toll than with the flat toll. Intriguingly, this pattern does not apply to residents of Ring 0. Although these residents do not pay tolls, they benefit only slightly from the flat toll since they drive exclusively on the outbound links that are not very congested in the

[32] Braid (1989) shows that in the single-link bottleneck model the second-best flat toll can be higher than the peak toll if auto travel demand is sufficiently inelastic.

NTE. And, surprisingly, they actually lose a bit from the step toll.[33] This appears to be because the step toll discourages other groups from traveling into the centre less than the flat toll does.

Differences between the effects of the flat and step cordon tolls on departure times are shown in the first two plots of Figure 7 (marked by hollow symbols). Similar to Regime 2 in Figure 6, the flat cordon toll shifts the distribution towards later departures throughout the travel period, whereas the step toll causes an initial acceleration.

4.4 Regimes 5 and 6: Flat and step area charges

The area charge differs from the cordon toll in that it is paid for all travel within Ring 2 rather than just for crossing inside the ring. Because trips initiated within the ring by residents of Rings 0 and 1 tend to be shorter than other trips, and to impose less congestion, the optimal flat and step area charges are a bit lower than their cordon-toll counterparts (Table 3). But since the area charges are paid on more trips, they have larger impacts on the auto share and travel costs. Also, as shown in Figure 7 the area charges cause bigger shifts in departure times. However, the welfare gain per dollar of revenue is lower than for the cordon tolls and travelers fare correspondingly less well. And unlike with the cordon tolls, residents of all rings are worse off on average (Table 6). Indeed, the greatest costs are borne by residents of Ring 0 who pay the toll if they drive at all while benefiting less from congestion relief because of their shorter, outbound, trips. Still, with the sole exception of the O-D ring pair 4-2, the step area charge is less burdensome than the flat charge and it actually leaves workers in Ring 0 slightly better off on average.

[33] Safirova *et al*. (2004) report a similar result for tolling of the Washington, D.C., road network.

5 CONCLUDING REMARKS

Various design considerations must be addressed if road pricing is to become widespread. And to measure reliably the welfare gains and distributional effects of candidate schemes, models need to have several capabilities: (1) specification of trip-timing decisions at a fine level of temporal resolution, (2) accounting for trip-timing preferences in the welfare analysis, (3) tolling by time of day, and (4) road networks of practical interest.

This paper analyzes some road pricing schemes using the dynamic network simulator METROPOLIS: a tool that treats endogenously departure-time decisions as well as mode and route choices of individual travelers. Simulations are conducted for a stylized urban road network. Amongst the results of the simulations, two stand out. First, for the parameter values used step tolls generate approximately twice the welfare gains from flat tolls while inducing a smaller shift of trips from auto to transit. Second, step tolls generate smaller revenues than do flat tolls, and consequently have more favourable distributional impacts on travelers.

These results are consistent with the findings of earlier analytical studies using the bottleneck model on one or two links. The present study goes further by incorporating network effects and analyzing welfare-distributional impacts on individuals according to trip origin and destination. Naturally, the findings are provisional since the simulations were performed on one network with a single set of parameter values, while ignoring the costs of setting up and operating tolling systems. Nevertheless, the findings are illustrative. And the analysis demonstrates some of the challenges encountered when computing and assessing first-best and second-best optimal time-of-day tolling schemes using dynamic models.

Several extensions are worth pursuing. First, while travelers here differ with respect to origins, destinations, preferred arrival times, and idiosyncratic preferences, other dimensions of heterogeneity are important to consider; in particular values of travel time and unit costs of schedule delay. Second, the analysis was restricted to home-to-work trips. If trips for other purposes and at other times of day were included, the estimated daily gains from road pricing would be higher although tolls would need to be re-optimized for different trip matrixes. Third, in assessing welfare impacts attention was restricted to rebates of toll revenue. Revenues can also be used to expand road capacity and public transportation service. And it may be possible to improve acceptability of road pricing by granting discounts or exemptions to the most adversely affected (or politically active) groups.

6 ACKNOWLEDGMENTS

We are grateful to Michael Bell and three anonymous referees for very helpful comments. Thanks are also due to the European Commission for partial financial support under Project MC-ICAM (Implementation of Marginal Cost Pricing in Transport – Integrated Conceptual and Applied Model Analysis), and to the Social Sciences and Humanities Research Council of Canada for support of the project "Road pricing in urban areas". Finally, we would like to thank Faisal Khan for able research assistance and Fabrice Marchal for helpful technical advice.

7 APPENDIXES

7.1 Appendix A

This appendix explains the parameter values listed in Table 1.

Unit costs of auto travel time

Estimates of VOT for commuting vary widely. They are sensitive to what dimensions of choice are included in the estimation procedure. And they are typically higher when estimated using stated preference data than revealed preference data, and higher for congested than uncongested conditions. Recent U.S. estimates range from about \$4/hr. to \$30/hr. A value near the geometric mid-point of this range of α_A=\$10/hr. was chosen.

Desired arrival times

The distribution of work-start times varies across cities and across employment sites within cities. Wilson (1988, p.15) reports a standard deviation of 46 mins. for work start times in Singapore, and Chu (1999, Figure 1) displays a distribution with a standard deviation of about 35 mins. for commuters in the San Francisco Bay Area. In contrast, some analytical studies (Vickrey, 1969; Cohen, 1987) assume a uniform distribution with a range of one hour, which has a standard deviation of only 17 mins. A standard deviation of 20 mins. was selected here on the grounds that the duration of the congested period in the model is much shorter than that in Wilson (1988) and Chu (1999). The mean work-start time does not affect results of interest, and was set at 8:00.

Schedule delay costs

Similar to VOT, estimates of unit schedule delay costs vary widely. The benchmark study is Small (1982) who estimates marginal rates of substitution between schedule delay and travel time. The estimates for his basic linear model (Model 1, Table 2) are β / a_A=0.61 and γ / a_A=2.38. Small also reports that many commuters have some flexibility in when they arrive at work, but incur a discrete penalty if they arrive late by more than a threshold. Since discrete penalties are not yet featured in METROPOLIS, these factors were accommodated roughly by setting the width of the on-time arrival window at Δ=10 mins., and rounding up the value of γ / a_A to 2.5. Given α_A=\$10/hr. the unit values for schedule delay costs were β=\$6/hr. (rounded down from \$6.1/hr.) and γ=\$25/hr.

Scale parameter for departure-time choice

This parameter was set at the METROPOLIS default value of μ_t=\$2. Smaller values were avoided because of a tendency for simulations to oscillate without converging towards a stationary state.

Public transportation

The logit scale parameter for mode choice was set at the METROPOLIS default value of μ_m=\$5. The fixed transit penalty, C_{p0}, and unit cost of in-vehicle transit time, α_P, were chosen by trial and error to yield a relatively high auto share of trips and a reasonable elasticity of auto demand with respect to generalized cost in the no-toll equilibrium.

7.2 Appendix B

Maximum link occupancy that can be sustained under free-flow conditions is $Occ_{\max} = s \cdot L / v$, where *s* denotes flow capacity of the link, *L* its length, and *v* free-flow speed. As an example, for a type *In1* link *s*=3,000 vehicles/hr., *L*=4 km. and *v*=50 km/hr. and hence $Occ_{\max} = 3{,}000 \cdot 4 / 50$ =240 vehicles.

8 REFERENCES

Arnott, R., A. de Palma and R. Lindsey (1990), "Economics of a Bottleneck", *Journal of Urban Economics* 27, 111-130.

Arnott, R., A. de Palma and R. Lindsey (1993), "A structural model of peak-period congestion: a traffic bottleneck with elastic demand", *American Economic Review* 83(1), 161-179.

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

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

Beckmann, M., C.B. McGuire and C.B. Winsten (1956), *Studies in the Economics of Transportation*, New Haven: Yale University Press.

Braid, R.M. (1989), "Uniform versus peak-load pricing of a bottleneck with elastic demand", *Journal of Urban Economics* 26, 320-327.

Braid, R.M. (1996), "Peak-load pricing of a transport facility with an unpriced substitute", *Journal of Urban Economics* 40, 179-197.

Carey, M. and A. Srinivasan (1993), "Externalities, average and marginal costs, and tolls on congested networks with time-varying flows", *Operations Research* 41(1), 217-231.

Chu, X. (1999), "Alternative congestion pricing schedules", *Regional Science and Urban Economics* 29, 697-722.

Cohen, Y. (1987), "Commuter welfare under peak-period congestion tolls: Who gains and who loses?", *International Journal of Transport Economics* 14(3), 238-266.

Dafermos, S. (1973), "Toll patterns for multiclass-user transportation networks", *Transportation Science* 7, 211-223.

De Palma, A. and P. Jehiel (1994), "Queuing may be efficient in bottleneck models", Department of Economics, Technical report 1994.9, University of Geneva.

De Palma, A., M. Kilani and R. Lindsey (2004), “A comparison of second-best and third-best tolling schemes on a road network”, paper submitted for presentation at 84th Annual Meeting of the Transportation Research Board, Washington, D.C.

De Palma, A. and R. Lindsey (2000), “Private toll roads: competition under various ownership regimes”, *Annals of Regional Science* 34(1), 13-35.

De Palma, A., R. Lindsey and K. Kilani (2003), “Modelling implementation paths for urban transport pricing, Workpackage 7.2: Paris case study”, Project MC-ICAM (Implementation of Marginal Cost Pricing in Transport – Integrated Conceptual and Applied Model Analysis: http://www.mcicam.net), October 22.

De Palma, A., R. Lindsey and E. Quinet (2004), “Time-varying road pricing and choice of toll locations”, forthcoming in G. Santos (ed.), *Road Pricing: Theory and Evidence*, Elsevier Science.

De Palma, A. and F. Marchal (2002a), “Real cases applications of the fully dynamic METROPOLIS Tool-box: An advocacy for global large-scale mesoscopic transportation systems”, *Networks and Spatial Theory* 2(4), 347-369.

De Palma, A. and F. Marchal (2002b), “Evaluation of road pricing with dynamic network models”, Transportation Research Board 81st Meeting, Washington, D.C., *National Research Council,* Preprint CD-ROM.

De Palma, A., F. Marchal and Y. Nesterov (1997), "METROPOLIS: Modular system for dynamic traffic simulation", *Transportation Research Record* 1607, 178-184.

Eliasson, J. and L.-G. Mattsson (2001), "Transport and location effects of road pricing: A simulation approach", *Journal of Transport Economics and Policy* 35, 417-456.

Hau, T.D, K.I. Wong and S.C. Wong (2003), "Estimating the efficient time-varying toll for parallel tunnels in Hong Kong", Paper presented at the congestion pricing symposium, Imperial College, London, August.

Henderson, J.V. (1974), "Road congestion: A reconsideration of pricing theory", *Journal of Urban Economics* 1, 346-355.

Hyman, G. and L. Mayhew (2002), "Optimizing the benefits of urban road user charging", *Transport Policy* 9(3), 189-207.

Krishnamurthy, S. and K.M. Kockelman (2003), "Propagation of uncertainty in transportation-land use models: An investigation of DRAM-EMPAL and UTPP predictions in Austin, Texas", *Transportation Research Record* 1831, 219-229.

Laih, C-H. (1994), "Queuing at a bottleneck with single- and multi-step tolls", *Transportation Research A* 28A(3), 197-208.

Lindsey, R. and E.T. Verhoef (2001), "Traffic congestion and congestion pricing", in K.J. Button and D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, Oxford: Elsevier Science, 77-105.

May, A.D., R. Liu, S.P. Shepherd and A. Sumalee (2002), "The impact of cordon design on the performance of road pricing schemes", *Transport Policy* 9, 209-220.

May, A.D., S.P. Shepherd, H. Minken, T. Markussen, G. Emberger and P. Pfaffenbichler (2001), "The use of response surfaces in specifying transport strategies", *Transport Policy* 8, 267-278.

Mayet, J. and M. Hansen (2000), "Congestion pricing with continuously distributed values of time", *Journal of Transport Economics and Policy* 34(3), 359-370.

Mun, S., K. Konishi and K.Yoshikawa (2003), “Optimal cordon pricing”, *Journal of Urban Economics* 54, 21-38.

Nakamura, K. and K. Kockelman (2002), "Congestion pricing and roadspace rationing: An application to the San Francisco Bay bridge corridor", *Transportation Research A* 36A(5), 403-417.

Parry, I.W.H. and A. Bento (2000), "Estimating the welfare effect of congestion taxes: The critical importance of other distortions within the transport system", Resources for the Future, Discussion Paper 00-51.

Pigou, A.C. (1920), *Wealth and Welfare*, London: Macmillan.

Richardson, H.W. and C-H C. Bae (1998), "The equity impacts of road congestion pricing", in K.J. Button and E.T. Verhoef (eds.), *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*, Cheltenham, UK: Edward Elgar, 247-262.

Safirova, E., K. Gillingham, I.W.H. Parry, P. Nelson, W. Harrington and D. Mason (2004), “Welfare and distributional effects of HOT lanes and other road pricing policies in Metropolitan Washington DC”, forthcoming in G. Santos (ed.), *Road Pricing: Theory and Evidence*, Elsevier Science.

Santos, G. and L. Rojey (2004), “Distributional impacts of road pricing: The truth behind the myth”, forthcoming in G. Santos (ed.), *Road Pricing: Theory and Evidence*, Elsevier Science.

Shepherd, S.P. and A. Sumalee (2002), “A Genetic algorithm based approach to optimal toll level and location problems”, *Networks and Spatial Economics,* forthcoming.

Small, K.A. (1982), “The scheduling of consumer activities: Work trips”, *American Economic Review* 72(3), 467-479.

Small, K.A. and J.A. Gómez-Ibáñez (1998), "Road pricing for congestion management: the transition from theory to policy", in K.J. Button and E.T. Verhoef (eds.), *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*, Cheltenham, UK: Edward Elgar, 213-246.

Small, K.A. and J. Yan (2001), "The value of "value pricing" of roads: Second-best pricing and product differentiation", *Journal of Urban Economics* 49(2), 310-336.

Transportation Research Board (2003), "Road Value Pricing: Traveler Response to Transport System Changes", *Transit Cooperative Research Program Report* 95; Transportation Research Board (http://gulliver.trb.org/publications/tcrp/tcrp_rpt_95c14.pdf).

Verhoef, E.T. (2002), "Second-best congestion pricing in general static transportation networks with elastic demands", *Regional Science and Urban Economics* 32, 281-310.

Verhoef, E.T., P. Nijkamp and P. Rietveld (1996), “Second-best congestion pricing”, *Journal of Urban Economics* 40(3), 279-302.

Verhoef, E.T. and K.A. Small (2004), "Product differentiation on roads: Constrained congestion pricing with heterogeneous users”, *Journal of Transport Economics and Policy* 38(1), 127-156.

Vickrey, W.S. (1967), "Optimization of traffic and facilities", *Journal of Transport Economics and Policy*, May, 123-136.

Vickrey, W.S. (1969), “Congestion theory and transport investment”, *American Economic Review* (*Papers and Proceedings*) 59(2), 251-260.

Viti, F., S.F. Catalano, M. Li, C. Lindveld and H. van Zuylen (2003), "An optimization problem with dynamic route-departure time choice and pricing", paper presented at 82nd Annual Meeting of the Transportation Research Board.

Walters, A.A. (1961), "The theory and measurement of private and social cost of highway congestion", *Econometrica* 29(4), 676-697.

Wilson, P.W. (1988), "Wage variation resulting from staggered work hours", *Journal of Urban Economics* 24, 9-26.

Yang, H. (1999), "System optimum, stochastic user equilibrium and optimal link tolls", *Transportation Science* 33, 354-360.

Yang, H. and M.G.H. Bell (1997), "Traffic restraint, road pricing and network equilibrium", *Transportation Research B* 31B(4), 303-314.

Yang, H. and H.-J. Huang (1997), "Analysis of time-varying pricing of a bottleneck with elastic demand using optimal control theory", *Transportation Research B* 31B(6), 425-440.

Yang, H. and H.-J. Huang (1998), "Principle of marginal-cost pricing: How does it work in a general road network?", *Transportation Research A* 32A(1), 45-54.

Yang, H. and H.-J. Huang (2003), "The multiclass, multicriteria traffic network equilibrium and system optimum problem", *Transportation Research B*, forthcoming.

Yang, H. and W.H.K. Lam (1996), "Optimal road tolls under conditions of queueing and congestion", *Transportation Research A* 30A(5), 319-332.

Yang, H. and X. Zhang (2002), "The multiclass network toll design problem with social and spatial equity constraints", working paper.

Zhang and Yang (2004), "The optimal cordon-based network congestion pricing problem" *Transportation Research B* 38B(6), 517-537.