Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Part A 41 (2007) 734–749

TRANSPORTATION
RESEARCH
PART A

www.elsevier.com/locate/tra

Measuring marginal congestion costs of urban transportation: Do networks matter?

Elena Safirova [a,*], Kenneth Gillingham [b,1], Sébastien Houde [a,2]

[a] *Resources for the Future, 1616 P Street, NW, Washington, DC 20036, United States*
[b] *Stanford University, P.O. Box 16336, Stanford, CA 94309, United States*

Received 24 January 2004; received in revised form 17 October 2006; accepted 12 December 2006

Abstract

In determining the marginal cost of congestion, economists have traditionally relied upon directly measuring traffic congestion on network links, disregarding any "network effects," since the latter are difficult to estimate. While for simple networks the comparison of the network-based congestion costs with the link-based ones can be done within a theoretical framework, it is important to know whether such network effects in real large-scale networks are quantitatively significant.

In this paper we use a strategic transportation planning model (START) to compare marginal congestion costs computed link-by-link with measures taking into account network effects. We find that while in aggregate network effects are not significant, congestion measured on a single link is a poor predictor of total congestion costs imposed by travel on that link. Also, we analyze the congestion proliferation effect on the network to see how congestion is distributed within an urban area.

Keywords: Marginal congestion costs; Congestion pricing; Urban networks

---

1. Introduction

The principle of marginal cost pricing of urban transportation infrastructure has become increasingly politically acceptable. Recent studies have made great strides forward towards developing more detailed and realistic urban transportation network models and more accurate empirical estimates of marginal congestion costs (MCC). Precise estimation of congestion costs is important and policy-relevant for several reasons. First, congestion costs serve as status indicators that describe the current state and trends of congestion. Second, they provide a basis for cost-benefit analysis that assesses whether individual projects and programs are worthwhile investments. Finally and possibly most importantly, obtaining accurate marginal congestion costs is crucial for designing efficient transportation infrastructure pricing schemes (Lee, 1997).

---

[*] Corresponding author. Tel.: +1 202 328 5120; fax: +1 202 939 3460.
E-mail addresses: safirova@rff.org (E. Safirova), kgilling@stanford.edu (K. Gillingham), houde@rff.org (S. Houde).

[1] Tel.: +1 650 353 6578; fax: +1 650 723 1614.

[2] Tel.: +1 202 328 5015; fax: +1 202 939 3460.

0965-8564/$ - see front matter
doi:10.1016/j.tra.2006.12.002

Traditionally, people have thought about costs of transportation in terms of average costs and therefore disregarded the negative externality each traveler imposes on others on a congested road. Using average costs tends to underestimate the true costs of congestion since this externality is not taken into account. More recently, a popular way to define the costs of urban congestion in the US is based on costs of delay, i.e. the difference in travel costs computed on the basis of the difference between actual speeds and free-flow speeds (Schrank and Lomax, 2005). While the costs of delay are an acceptable benchmark, especially for comparing congestion levels across different metro areas, they are not particularly instructive since bringing all traffic to free-flow speeds at all times would constitute an inefficient overprovision of road space and therefore cannot serve as a meaningful policy goal.[3] In the United Kingdom, a long-standing discussion of the nature and numerical estimates of congestion has led to a notion of "total costs of congestion" that suffers from the same problems as the costs of delay estimates (Graham and Glaister, 2006). Only marginal costs of congestion capture the magnitude of congestion externalities and thus should be examined most closely in the design of transportation policies (Newberry, 1990).

Currently, there exist several competing approaches to computing marginal congestion costs of urban transportation that go beyond one-link static models. One approach emphasizes the connections between transportation and other sectors in the economy and focuses on analyzing the total impact of traffic congestion. Although presenting a rich picture of inter-sectoral relationships, such models tend to feature a relatively simple representation of the transportation system. For example, Mayeres and Proost (2001) evaluate the efficiency effects of transportation charges in Belgium by computing the marginal welfare cost of public funds for a number of tax instruments. Parry and Bento (2001) emphasize the interaction between congestion and labor supply and point out that congestion-pricing schemes are sensitive to the allocation of the revenues. While these general-equilibrium effects are important, the simple treatment of the transportation system in this class of models prevents them from addressing complex network modeling issues.

Another approach assumes that marginal congestion costs obtained for each link on the network could be used as substitutes for the true system-wide marginal congestion costs. For instance, Anderson and Mohring (1997) compute marginal congestion costs on the road network of the Twin Cities area using a link-by-link method and draw on obtained results to simulate a marginal congestion pricing policy. Ozbay et al. (2001) use speed–flow relationships on each link to compute marginal costs along routes for a full network. They thus assume that any additional flow in the system does not disturb the existing flow patterns on the network. The latter authors recognize that this is an approximation: "We are aware of the fact that the resulting value will not be the same as the true system-wide marginal cost. This value can only be obtained by performing a new traffic equilibrium assignment, which will reflect the change in flow patterns due to the addition of an extra unit of demand. However, compared to the overall demand, because the additional demand is relatively small… we can assume that the resulting costs will be reasonable approximations of actual costs." (p. 85). These changes in flow patterns due to the addition of an extra unit of demand are commonly referred to as *network effects* and a full accounting of the impacts of these effects on the true system-wide marginal congestion costs remains an open research topic.

Because of the intrinsic network nature of road transportation, it is often hard to separate localized impacts of a policy (*link effects*) from the global impact (*network effects*). Although most transportation policies in practice (widening a road, pedestrianization of a city center, cordon toll) are local, their influence often can spread out well beyond the small area of the initial impact. Responding to policies, the traffic will shift from the impacted part of the network to other areas, and the intensity of the shift will depend on several factors, such as strength of the impact, demand structure and network configuration. The issues of connectivity of transportation networks are also important in application to network security and vulnerability issues (Jenelius et al., 2006) and dynamics of investment pattern in parts of the network (Zhang and Levinson, 2006).

A branch of the literature has approached the issue of network effects in the context of optimal and second-best congestion tolls. In the traffic assignment literature it is well-known that marginal-cost tolls are optimal on a network with fixed demands (see, e.g., Sheffi, 1985). Yang and Huang (1998) set up an optimization problem to determine optimal congestion tolls for each link on a general congested network. The solution suggests

---

[3] For one thing, reducing congestion to zero at all times is likely to be prohibitively costly and not achievable on limited budgets of metropolitan planning organizations.

that network interactions do matter, but it is impossible to quantify such effects using a theoretical model. Hearn and Yildirim (2002) develop an algorithm for finding congestion tolls on a network with elastic demands and use this algorithm for determining the congestion tolls for a small network. They numerically solve for optimal congestion tolls for a theoretical 9-node network with elastic demands and conjecture that the same algorithm can successfully be applied to solving modest-sized urban networks. However, elastic-demand optimal toll problem has yet to be solved for a realistic transportation network.

Although elastic demands and route choice pose significant problems for transportation researchers, real networks feature an array of other complicating features – mode choice, different times of day, and heterogeneous agents. Each of these factors has the ability to significantly complicate the overall network equilibrium. While in the literature mode choice, time of the day differentiation, and agent heterogeneity deserve and receive separate treatment (e.g., Verhoef, 2002; Verhoef and Small, 2004; Liu and McDonald, 1998; Arnott and Yan, 2000; Armelius, 2005), in this paper we analyze their *composite* effect on a real network. In particular, we attempt to determine whether in the presence of all the complicating factors, network effects can be significant enough to render calculations of marginal congestion costs using the common link-by link method inaccurate.

In order to achieve this research goal, we employ a strategic transportation planning model, START, calibrated to the Washington, DC metro area as an example of a sufficiently large network featuring mode choice, time periods, agent heterogeneity, and a realistic distribution of demand. While we do not claim that the results of this paper are general and applicable to other metropolitan areas, we intend our work to shed some light on the relationship between marginal congestion costs measured on individual links isolated from the network and marginal congestion costs measured on links in an integrated network.

The rest of the paper is structured as follows. In Section 2, we provide an overview of the START Model and key facts about the Washington, DC modeling region. In Section 3, three methods of computing congestion costs are described and compared. Section 4 presents major results of this research and outlines the implications of these results for congestion pricing. Section 5 concludes and provides direction for future study.

2. START model overview

The START (Strategic and Regional Transport) modeling suite was developed by MVA Consultancy and has been applied to a range of urban centers in the United Kingdom, including Birmingham, Edinburgh, Bristol and South England (Croombe et al., 1997; May et al., 1992).

Unlike traditional transportation models, START is designed to predict the outcomes of different transportation strategies, where strategies refer to the combinations of different transport elements, which in broad terms encompass changes in capacity (e.g., new infrastructure), operating conditions, and prices. Therefore while most of the components of the model are conventional, the suite features a limited number of zones and an aggregated representation of the supply side (transportation network and provision of public transportation) coupled with a rich and detailed demand side. An important advantage of the model is its relatively fast runtime, which provides the opportunity to conduct a number of simulations to better understand probable policy consequences.

Traffic congestion on highway links in START is modeled via speed/flow·distance curves specified for each highway link.[4] Unlike some other popular transportation models, congestion arising at intersections is not explicitly represented. Explicitly defined routes between each origin and destination pair determine the quantity of traffic and the distance traveled by that traffic on each of the START links. Thus, when the START demand model forecasts the number of trips between each origin and destination and assigns these onto specific routes and consequently onto links, the "flow·distance" information allows the link speed to be calculated. The choice of route is determined endogenously and is influenced by the congestion on each link.

---

[4] START differs from most traffic assignment models since it utilizes speed/flow·distance curves instead of speed/flow curves. This feature is required because in this modeling suite routes can contain portions of road links instead of entire links. The speed is in the units of miles per hour and flow·distance is in the units of PCU-miles, where PCU stands for passenger-car-unit (to account for buses and trucks requiring more road space than a typical passenger car).

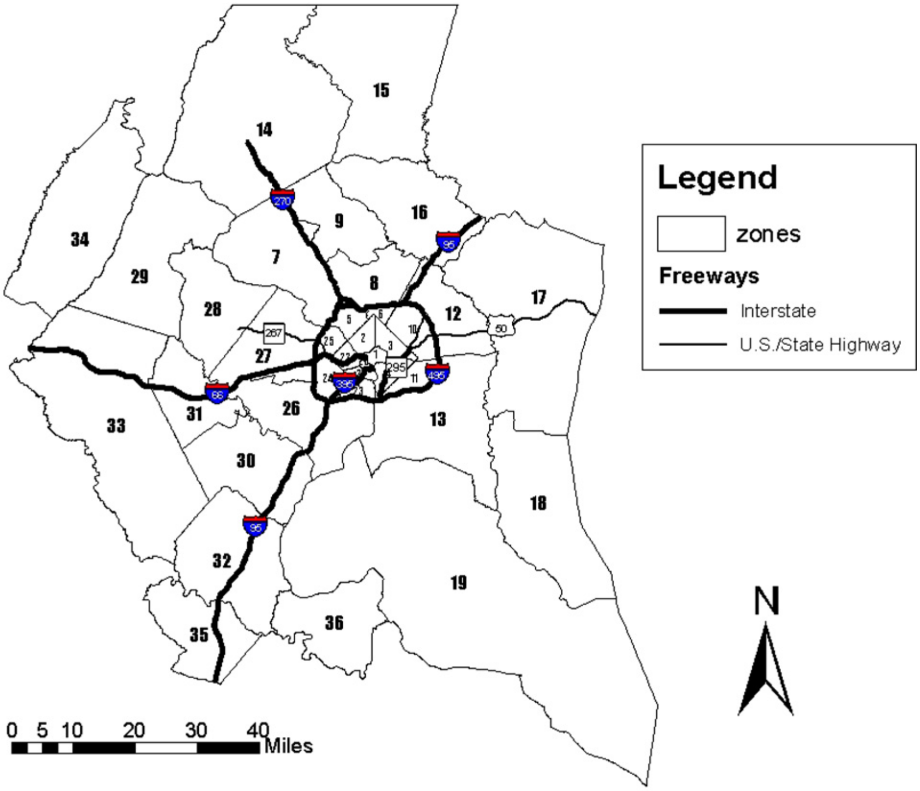

Fig. 1. START modeling region with all special links.

Therefore, the model is very well positioned to address the question of the extent of congestion spillovers in the network.

The Washington-START model has 40 travel zones with three stylized transportation links in each zone (inbound, outbound and circumferential) and a number of other "special" links that represent freeway segments and bridges. Six main corridors – I-270, I-95, and US-50 in Maryland and I-66, I-95, and US-267 in Northern Virginia connect the outer suburbs to the central region within the circular road I-495/I-95 known as the Beltway (see Fig. 1). The rail network combines the Washington Metro-rail system and suburban light rail systems (MARC, VRE). Bus travel is represented by a highly stylized route network, with bus accessibility in any zone determined by the density of stops, frequency of service, and reported bus travel times. Transit crowding costs and parking search costs are explicitly included in the model. We also account for existing high occupancy vehicle (HOV) lanes on I-95, I-395, I-66, and VA-267 in Northern Virginia, as well as I-70 and US-50 in Maryland. Moreover, we recently have made several improvements to transit modeling, such as incorporation of park-and-ride facilities for rail trips, placing buses on links used by other on-road vehicles (so that buses are affected by and contribute to road congestion), and more detailed treatment of rail network.[5]

This rather aggregated supply-side representation is combined with a detailed demand-side structure. The model features 8 household types differentiated by income and vehicle ownership levels. There are six trip purposes: home-based work (HBW), home-based shopping (HBS), home-based other (HBO), non-home-based work (NHBW), non-home-based other (NHBO), and freight. Home-based trips either originate or terminate at home. The model distinguishes four travel modes: single occupancy vehicle (SOV), high-occupancy vehicle (HOV), bus/rail, and non-motorized (walk/bike). It also contains three times of day: morning peak, afternoon peak, and off-peak (weekend travel is excluded). Table 2 contains an overview of the breakdown in travel demand in DC by purpose and time of day.

---

[5] See Nelson et al. (2006) for more details on transit modeling improvements.

Table 1
Beta coefficient values

	Home based work	Home based shopping	Home based other	Non-home based work	Non-home based other
Trip generation	−0.0045	−0.005	−0.0045	−0.0045	−0.0045
Destination choice	−0.02	−0.05	−0.02	−0.02	−0.02
Mode choice	−0.05	−0.05	−0.05	−0.05	−0.05
Time choice	−0.05	−0.1	−0.09	−0.1	−0.1
Route choice	−0.185	−0.185	−0.185	−0.185	−0.185

START takes the distribution of households by demographic segment and residential location as exogenous. Travel decision-making is modeled as a nested logit tree. The utility functions at each nest are linear in generalized cost (the combined time and monetary costs of travel). The value of time is a fraction of the traveler's wage rate and this fraction varies according to trip purpose and mode.[6] In successive nests, households choose whether to take the trip, then destination, mode, time of day, and route.

These choices can be seen as a sequential process. First, the decision whether to make a trip at all is made. Then, conditional on that choice, a destination is chosen. After that, conditional on the choices made previously, mode is selected, and so forth.

Therefore, the probability that a consumer makes a trip i to destination j by mode m during period t using route r is

$$P_{ijmtr} = P_i P_{j|i} P_{m|ij} P_{t|ijm} P_{r|ijmt}. \tag{1}$$

The five choice levels are described by logit models. For example, the route choice (the lowest nest) is given by a logit demand form:

$$P_{r|ijmt} = \frac{\exp(A_{ijmtr} - \beta^r p_{ijmtr})}{\sum_{l=1}^{R_{ijmt}} \exp(A_{ijmtl} - \beta^r p_{ijmtl})}, \quad r = 1, \ldots, R_{ijmt}, \tag{2}$$

where

$P_{r\|ijmt}$	the probability that a route r is chosen conditional on choice of generation, destination, mode and time of day
p_{ijmtr}	the generalized costs of route r
A_{ijmtr}	the constant term which includes all aspects of attractiveness of a travel option except the cost
β^r	the route choice elasticity parameter
R_{ijmt}	the number of routes in the nest

The parameter $0 \leqslant \beta^r < 1$ represents the elasticity of travel choices at the nest level. When $\beta^r \to 0$, the choice between routes is price-inelastic, and is hardly affected when generalized costs change. On the other hand, when $\beta^r \to 1$, travel choices are very sensitive to prices. Values of parameters used in the model are presented in Table 1.

To calibrate the model, we use data on how many trips occur on different links within zones by consolidating output from the Metropolitan Washington Council of Governments (COG) Version 1 transportation planning model (which disaggregates over 2100 travel zones). Using data from the Census Transportation Planning Package and 1994 Travel Survey, we estimate how many households from different demographic groups live and work in different zones, and from this we are able to allocate total trips on any given link to different household groups. Data on wages and price indices were obtained from the Census and Bureau of Labor Statistics. Trip times on each link are validated against estimates of rush-hour speeds developed from analysis of aerial photography (Council of Governments, 1999). The model in its present form has also been used to conduct policy simulations of gasoline taxes (Nelson et al., 2003), HOT lanes (Safirova et al., 2003),

---

[6] Values of time in current application vary between \$2.70 and \$18.80 in 2000 dollars.

Table 2
DC region trips demand by purpose and time period

	Home based work	Home based shopping	Home based other	Non-home based work	Non-home based other
Morning peak	1455	252	2054	191	399
Afternoon peak	1515	1158	3228	347	1535
Off-peak	1820	1504	3607	1959	2589

Units in thousands of trips.

and congestion pricing (Safirova et al., 2004, 2005), as well as to evaluate the benefits of public transit (Nelson et al., 2006). Moreover, recently it has been integrated with a model of land use and regional economy RELU and within that package used to test long-term effects of congestion pricing (Safirova et al., 2006a,b).

3. Congestion measurement

As has been discussed in the introduction, one method of measuring marginal congestion costs is on a link-by-link basis, which makes the implicit assumption that proliferation of congestion on the network is relatively insignificant. Undoubtedly, it is the simplest method of measuring marginal congestion costs and this simplicity is quite appealing. In this paper we will call this approach method 1.

In order to take into account the redistribution of traffic flow over the network, we also develop methods 2 and 3. The attractiveness of method 2, like method 1, lies in its relative simplicity in application to our model. In particular, we are able to compute the MCC on all network links via method 2 by running the START model only once. On the other hand, method 3 seems to present the most theoretically correct results, but requires the highest level of effort – a separate run of the model is required for the computation of MCC on each link. Therefore, we apply method 3 only to a limited, but representative set of links.[7]

3.1. Method 1

Method 1 simply utilizes the exogenous speed/flow·distance relationship governing congestion on each individual link. Suppose a speed/flow·distance relationship is denoted by $S_k = S(\mathrm{FD}_k)$, where S_k is the speed on the link and FD_k is the flow·distance on that link. Then the marginal congestion costs per mile on link k would be equal to

$$\mathrm{MCC}_k^1 = \left(\frac{1}{S_{k1}} - \frac{1}{S_{k0}}\right) \times \mathrm{FD}_k \times \mathrm{VOT}_k, \tag{3}$$

where S_{k0} and S_{k1} are correspondingly the initial and resulting speed levels on the link k after adding one unit of flow·distance to the link. Therefore, marginal congestion costs imposed by one extra vehicle mile traveled on a link k comes to an increase in travel time $\left(\frac{1}{S_{k1}} - \frac{1}{S_{k0}}\right)$ experienced by all other link k users FD_k multiplied by the average value of time VOT_k.

Since $S_{k1} = S_{k0} + \frac{\partial S}{\partial \mathrm{FD}_k}$, we can rewrite Eq. (3) as follows:

$$\mathrm{MCC}_k^1 = -\frac{\frac{\partial S}{\partial \mathrm{FD}_k}}{S_{k0}\left(S_{k0} + \frac{\partial S}{\partial \mathrm{FD}_k}\right)} \times \mathrm{FD}_k \times \mathrm{VOT}_k, \tag{3$'$}$$

where VOT_k is the average value of time of travelers on link k (for the sake of simplicity, we do not account for changes in monetary costs of travel due to decreased speeds since those are of much smaller magnitude than direct time losses).

In other words, the marginal congestion cost per mile of travel on a link is equal to the monetary value of time lost by all travelers on link k. This measure assumes that travel times on all other links except the link k are unchanged and there is no traffic reassignment.

---

[7] The links were chosen to represent different link types (inbound, outbound, circumferential, special) as well as different geographical parts of the metro area.

3.2. Method 2

In order to account for the interaction between traffic congestion on different links on the network, one has to perform a new traffic assignment and compute marginal congestion costs on links resulting from changes in speeds on the entire network. In this study, we change (for example, decrease) the demand for travel between all origin-destination pairs by the same small percentage and run the START model with the new initial demand. Then, the marginal congestion costs per mile of travel are computed as follows:

$$\mathrm{MCC}_k^2 = \left(\frac{1}{S_{k1}} - \frac{1}{S_{k0}}\right) \times \frac{(\mathrm{FD}_{k0} + \mathrm{FD}_{k1})}{2} \times \mathrm{VOT}_k \times \left(\frac{1}{\mathrm{FD}_{k1} - \mathrm{FD}_{k0}}\right), \tag{4}$$

where subscripts 0 and 1 denote initial and resulting traffic assignments.

Eq. (4) looks very similar to the Eq. (3) except for the fact that now changes in FD on each link can be much larger than unitary perturbations assumed in method 1. Therefore, we multiply the time losses experienced on link k by the average of the initial and resulting flow·distance on the link $\frac{(\mathrm{FD}_{k0}+\mathrm{FD}_{k1})}{2}$. Likewise, the value of MCC should be prorated to reflect the impact of unitary increase in VMT on the link and therefore the result is multiplied by $\left(\frac{1}{\mathrm{FD}_{k1}-\mathrm{FD}_{k0}}\right)$.

As it stands, in this second method we tacitly attribute changes in a link's congestion level to additional PCUs (passenger-car-units) on the very same link. By adding more PCUs to every link, travel demand is decreased uniformly. However, the flow on different links varies to different degrees depending on redistribution of the traffic in the network.

An important advantage of method 2 is that, unlike method 1, it accounts for the interaction of the speeds and traffic flows on the network. In fact, it accounts for all the network effects and is very cost-effective since obtaining a full set of MCC using this method requires only one model run. Unfortunately, while method 2 still does take into account the effects of congestion on one link on congestion on other links of the network, it is not exact. For example, if we just consider two links on the network – k and n, then the overall network effect would include the changes in speeds on both links due to changes in flow on link k as well as changes in speeds on both links due to changes in flow on link n. However, this method attributes *all* changes in speed on link k to changes in flow on link k, and the same is true for link n. Furthermore, it is not possible to uncouple those effects using method 2. Therefore, to accurately account for all congestion redistribution, we have to turn to method 3.

3.3. Method 3

In order to simulate the overall network effects of a unitary change in flow on a single link, we need to be able to change the demand on the link in question only and to keep all other demands intact. However, since the travel demand is defined by origin-destination pairs, it is impossible to do so. Therefore, instead of changing demand, we simulate the impact of a unitary PCU-mile increase on a link by reducing the capacity on that link by one PCU-mile and rerunning START with the reduced supply. Suppose a network contains a total of N links. Then, the marginal congestion cost per mile on a link k correctly accounting for the full effects on all other links would be

$$\mathrm{MCC}_k^3 = \sum_{n=1}^{N} \left(\frac{1}{S_{n1}} - \frac{1}{S_{n0}}\right) \times \left(\frac{\mathrm{FD}_{n0} + \mathrm{FD}_{n1}}{2}\right) \times \mathrm{VOT}_n, \tag{5}$$

where S_{n1} and FD_{n1} are speeds and flows on a link n resulting from a decrease in road supply on link k by 1 PCU-mile. As we stated in the beginning of this section, although method 3 provides the most accurate results, it is also the costliest of the three since a computation of MCC for each link requires an additional run. However, on average, methods 2 and 3 should produce the same results because both of them include own-link effects as well as network effects, but attribute them to different links. Therefore, since running model 3 for each link of the network is prohibitively costly, we will use the results obtained using method 2 as a proxy for the overall results that would be obtained by method 3.

At the same time, since method 2 is inexact in application to individual links, it is important to see to what extent the results yielded by the three methods differ quantitatively to judge if and under what conditions using methods 1 and 2 achieves satisfactory results.

4. Results

4.1. Comparison of methods

The quantitative results of this study suggest it may be appropriate to use marginal congestion costs obtained using method 1 to approximate average marginal congestion costs accounting for network effects as computed by method 2. To provide a measure of the degree of variability between the two methods, we have computed the average of their differences (thereafter referred as AD) for each peak period. In particular, the average difference is computed as follows:

$$\mathrm{AD} = \frac{\sum_{n=1}^{N} \mathrm{FD}_{n0} \times |\mathrm{MCC}_n^1 - \mathrm{MCC}_n^2|}{\sum_{n=1}^{N} \mathrm{FD}_{n0}}. \tag{6}$$

The AD is an average of the absolute differences between the two calculations of MCC obtained using methods 1 and 2. The differences are computed for each link and then weighted by link flows. In our calculations, these measures come out to be small relatively to the average MCC. For the morning peak, the AD for all links is 1.3 cent, where the average MCC are 6.5 cents and 5.9 cents for methods 1 and 2, respectively. For the afternoon peak, the AD at 4.2 cents is more substantial, but still small relatively to the average MCC, 18.3 (method 1) cents and 14.2 cents (method 2).

The degree of variation between the MCC computed using methods 1 and 2 could also be seen through the comparison between Figs. 2a and b and 3a and b, where congestion costs are averaged by zone for each peak period. We observe that the zonal averages computed by the two methods have very close results. If the policy options are limited to spatially uniform ones, such as a fuel tax, and the goal is to obtain an average marginal congestion cost over the entire region, the results obtained using the two methods are similar.

Note, however, that looking at average measurements can be misleading. On the one hand, average MCC have a great variability; on the other hand, the size of the network effects varies with characteristics of the links. Table 3 shows that although MCC measures are low on average, they are highly variable and positively skewed. As expected, some links have much higher congestion costs. The outbound arterial link of DC Downtown, the most congested link during the evening peak, has MCC of 341.9 cents and 534.1 under methods 1 and 2, respectively. During the morning peak, the inbound arterial link going through East Arlington is the most congested link with MCC of 250.4 cents by method 1 and 228.3 cents by method 2.

Table 4 shows how AD measures vary with the level of congestion. We have calculated the averages for subsets of links corresponding to different congestion levels, for example for the 5% most congested links, 10% and so on. Note that Table 4 also reports the AD measures in relative terms, i.e. as a fraction of the average MCC in the base case. These latter measures correspond to the relative importance of the network effects, as computed by method 2. We observe that in absolute terms the network effects are greater on the most congested links. But in relative terms those measures practically do not vary with the level of congestion, which is particularly true during the evening period. Although we do not suggest generalizing this result, it is worth mentioning that it primarily depends on the structure of alternative routes available to travelers during each time periods.

Before we go on to compare the results of the method 3 with those of two other methods, it is useful to decompose the MCC computed using method 3 into two components:

$$\begin{aligned}\mathrm{MCC}_k^3 = {} & \left(\frac{1}{S_{k1}} - \frac{1}{S_{k0}}\right) \times \left(\frac{\mathrm{FD}_{k0} + \mathrm{FD}_{k1}}{2}\right) \times \mathrm{VOT}_k + \sum_{n=1}^{k-1} \left(\frac{1}{S_{n1}} - \frac{1}{S_{n0}}\right) \times \left(\frac{\mathrm{FD}_{n0} + \mathrm{FD}_{n1}}{2}\right) \times \mathrm{VOT}_n \\ & + \sum_{n=k+1}^{N} \left(\frac{1}{S_{n1}} - \frac{1}{S_{n0}}\right) \times \left(\frac{\mathrm{FD}_{n0} + \mathrm{FD}_{n1}}{2}\right) \times \mathrm{VOT}_n.\end{aligned} \tag{5'}$$

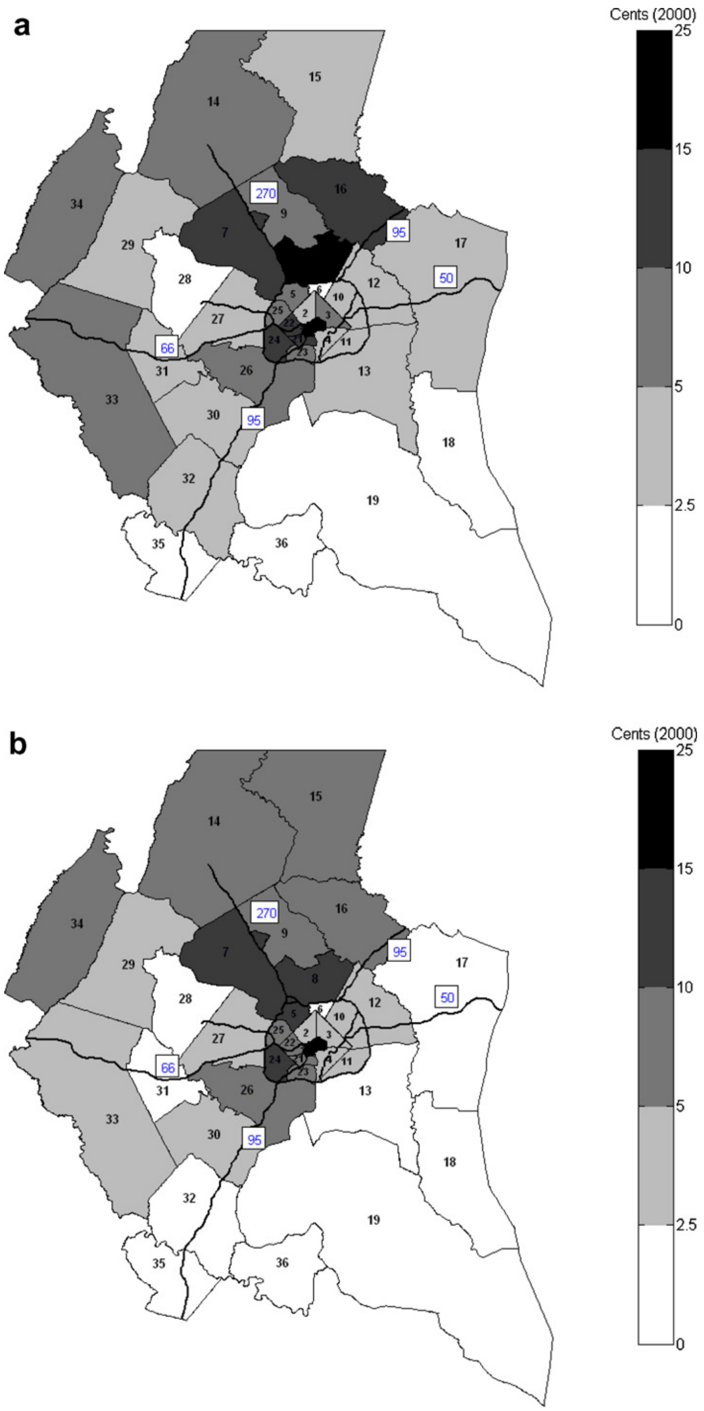

Fig. 2. Zonal average marginal congestion cost calculated, morning peak: (a) method 1, (b) method 2.

The first component represents the own-link effect, i.e. the marginal congestion costs on the link k resulting from an initial unitary increase in flow on the same link. The second component is the sum of the effects of the initial unitary increase in flow on link k on all other links in the network. These effects can be cumulatively considered the true network effects.

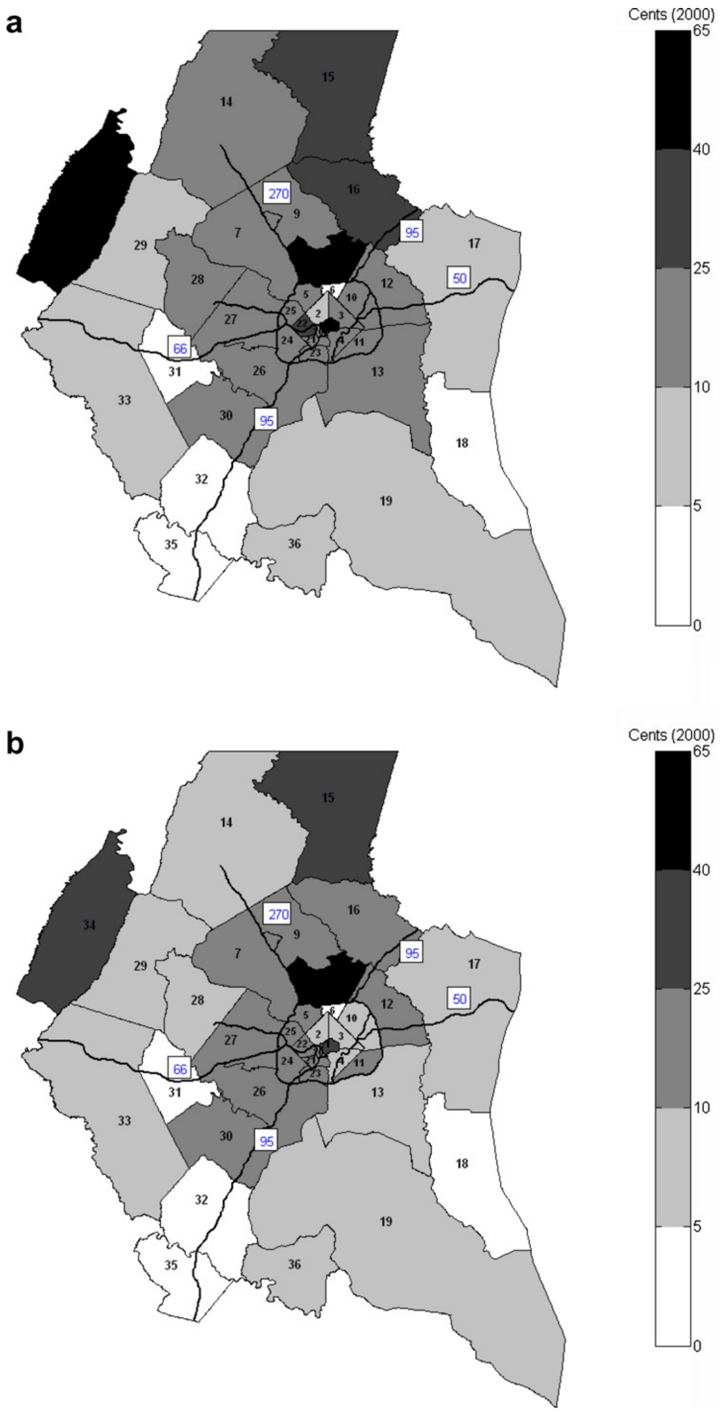

Fig. 3. Zonal average marginal congestion cost calculated, evening peak: (a) method 1, (b) method 2.

From Table 5 one can make several interesting observations regarding these two components. First, the MCC computed using method 2 tends to be higher than the own-link effect computed using method 3. Moreover, for most links, the total effect computed using method 3 is smaller than the own-link effect alone. Therefore, although we have results based on method 3 only for selected number of links, we hypothesize that the

Table 3
Average marginal costs of congestion over all the network by time period

	Morning	Evening
Method 1		
Average MCC: all links	6.5	18.3
Standard deviation	12.4	26.7
Skewness	12.6	7.7
Method 2		
Average MCC: all links	5.9	14.2
Standard deviation	11.4	20.3
Skewness	11.9	5.6

Note: Values in 2000 cents.

Table 4
Average of the differences in MCC between methods 1 and 2, comparisons across different link subsets

Link classification for selection of link subsets	Percentiles (%)	Average of the differences in MCC between methods 1 and 2			
		Morning		Evening	
		In absolute terms (2000 cents)	Relative to the MCC in the base case (%)	In absolute terms (2000 cents)	Relative to the MCC in the base case (%)
Links with greater differences in MCC, method 1 vs. method 2	95	8.4	39.9	18.7	26.3
	90	5.7	29.9	15.9	25.9
	75	3.5	24.5	10.6	24.4
	50	2.2	20.8	7.1	24.3
	25	1.7	20.6	5.4	23.8
	0	1.3	20.2	4.2	23.2
Most congested links	95	3.9	12.0	17.0	24.1
	90	3.1	12.1	14.9	23.6
	75	2.9	17.0	10.0	22.1
	50	2.0	17.8	6.9	22.7
	25	1.7	20.0	5.4	23.0
	0	1.3	20.2	4.2	23.3

results computed using method 3 would on aggregate show a lower average MCC than those computed by methods 1 and 2.

4.2. Congestion spillovers over the road network

The results obtained using method 3 provide an opportunity to learn to what extent a real transportation network serves as a conductor of congestion. In particular, it is interesting to see how strong the effects on other links are and how far from the point of impact they can be felt.

Looking at the Table 6, we conclude that the degree to which travel conditions on one link affect other links greatly depends on whether the affected link turns out to be a ‘‘bottleneck’’ on the network. In other words, if a link happens to be more heavily used by travelers along a large number of routes, a shock to that link would result in an impact on numerous other links. On the other hand, if a link primarily serves local travelers and is not very congested initially, only a limited number of other links turn out to be affected. We propose to measure the extent of the congestion spillovers by counting the number of links that are *significantly impacted* following a congestion impact on a given link. More precisely, we define a link as *significantly impacted* if its change in MCC is greater than 1 percent of the MCC of the link that was subject to the initial impact. Links on which changes in the MCC are below the 1% threshold are considered not significantly affected by conges-

Table 5
Marginal congestion costs from 3 methods on selected links

Selected link	Zone	Description	Time period	Method 1	Method 2	Method 3	
						Own-link	Total
1	1	DC Downtown inbound arterial link	Morning	34.4	43.6	27.7	27.1
2	1	DC Downtown outbound arterial link	Evening	341.9	534.1	168.6	110.1
35	5	American Legion Bridge	Evening	45.8	43.9	26.8	46.2
53	8	E Montgomery Co. outbound arterial link	Evening	91.6	69.7	42.2	28.0
57	8	E Montgomery Co. Beltway	Evening	57.1	43.5	23.3	14.2
65	9	NE Montgomery Co. circumferential arterial link	Evening	31.4	23.4	18.7	8.3
98	12	NE Prince George’s US 50 East of Beltway	Evening	38.9	34.3	20.7	6.5
137	20	E Arlington inbound arterial link	Morning	250.4	228.3	126.1	84.5
171	24	E Fairfax Co. I-395	Morning	47.1	38.2	33.5	13.0
189	25	NE Fairfax Co. Beltway	Evening	66.0	48.3	33.2	14.5
197	26	S Fairfax Co. I-95 North bound direction	Morning	26.9	24.7	15.0	4.7
198	26	S Fairfax Co. I-95 South bound direction	Evening	27.7	24.0	16.5	10.4

Note: Marginal cost values in 2000 cents.

tion spillovers.[8] According to this measure, we observe that it is not necessarily the most congested links (such as links 2 and 137) that have the greater extent of congestion spillovers. Those two links impact 13 and 95 links, respectively. In fact, a link that has a particular “bottleneck-prone” characteristic, such as being a bridge, proves to impact more links. Here, American Legion Bridge (link 35) impacts 20 links. Other links with low congestion and those not having the peculiarity of being a bridge, for example link 197, can also still have a greater extent of congestion spillovers (9 links impacted) than links that are significantly more congested.

At first it might seem counterintuitive that an increase in the costs of travel on one link leads to a *decrease* in the level of congestion on a number of other links. The major factor contributing to this is the so-called *bundling effect*. After the initial increase in MCC on the impacted link, some travelers will switch to other routes. Therefore, the number of travelers on routes containing the impacted link would have decreased and, consequently, other links along those routes would become less congested. From Table 6, we see that this *bundling effect* proves to be large on highly congested links (link 2 and link 137). For example, the DC Downtown circumferential arterial link has a negative MCC of 23 cents, in response to a marginal increase in congestion of the outbound arterial link in the same zone.

4.3. Implications for congestion pricing

The results presented above provide a few lessons on design of congestion pricing schemes. First of all, using non-network marginal congestion costs as a basis for estimating average levels of congestion tolls, for a given zone, for example, may be very accurate since they are highly correlated with network-based marginal congestion costs. However, marginal congestion costs vary considerably link-by-link and therefore setting uniform tolls for all roads during rush hour would result in an outcome significantly different from the first-best. Furthermore, we have found that the discrepancies between own-link MCC and network MCC are larger on more congested links. This result has important implications for practitioners. Failing to account for the network effects while implementing congestion pricing will result in less accurate toll levels on the most congested links, where the price of mistakes is higher, both because of high congestion and high traffic volumes. Finally, our research suggests that pricing individual links on the network, while keeping the rest of the network un-priced, would result in significant congestion spillovers in a large part of the metro area. World practice shows that congestion pricing is being introduced by small steps and in localized areas (Ramjerdi et al., 2004; Santos, 2004b; DeCorla-Souza, 2004). Under such circumstances, it is very important to take into account congestion spillovers from priced to un-priced facilities (Liu and McDonald, 1998).

---

[8] Although the 1% threshold is set up in an ad hoc manner, it allows sorting out links with numerically significant spillover effects.

Table 6
Distribution of marginal congestion costs using method 3

Link description	Selected link	Zone	Outbound arterial link DC Downtown (zone 1, link 2)	Inbound arterial link E Arlington (zone 20, link137)	American Legion Bridge (zone 5, link35)	I-95 North Bound S Fairfax Co. (zone 26, link 197)
DC Downtown inbound arterial link	1	1		1.3		
DC Downtown outbound arterial link	2	1	**168.6**	−0.2		
DC Downtown circumferential arterial link	3	1	−23.0	−8.6	0.1	
NW DC inbound arterial link	11	2	−2.1	−0.1		
NW DC circumferential arterial link	12	2	−1.6	0.1		
SE DC outbound arterial link	19	4	−3.8	−0.1		
SW Montgomery Co. Inner Beltway	32	5	−0.6	0.4	1.8	
SW Montgomery Co. Outer Beltway	33	5	0.6	−0.1	0.8	
American Legion Bridge	35	5			**26.8**	
W Montgomery Co. outbound arterial link	44	7	−1.7		−0.1	
W Montgomery Co. I-270 North of Beltway	46	7	−1.4		0.3	
E Montgomery Co. outbound arterial link	53	8	0.6		1.5	
E Montgomery Co. US 50	57	8	−0.6		0.8	
NW Prince George Co. Beltway	76	10	−2.6			
SW Prince George Co. outbound arterial link	87	11	−1.4			
SW Prince George Co. Beltway	93	11	0.6		0.4	
NE Prince George Co. US 50	98	12	−1.8		0.1	
SE Prince George Co. outbound arterial link	109	13	−1.3		0.1	
E Arlington inbound arterial link	137	20		**126.1**		
E Arlington circumferential arterial link	138	20	−3.1	−0.9	−0.2	
S Arlington outbound arterial link	146	21	−2.0	−0.1		
S Arlington I-395	148	21	−0.5	−1.7		−0.8
W Arlington inbound arterial link	152	22	−0.5	−22.4		
W Arlington outbound arterial link	153	22	7.1		0.2	
W Arlington I-66 HOV	155	22	−0.4	0.9		
Alexandria I-395	160	23	−0.2	−1.1		−1.4
Alexandria I-395	161	23	1.8		0.4	
E Fairfax Co. inbound arterial link	168	24	−0.1	−0.9		

Table 6 (*continued*)

Link description	Selected link	Zone	Outbound arterial link DC Downtown (zone 1, link 2)	Inbound arterial link E Arlington (zone 20, link137)	American Legion Bridge (zone 5, link35)	I-95 North Bound S Fairfax Co. (zone 26, link 197)
E Fairfax Co. outbound arterial link	169	24	−1.8		0.1	
E Fairfax Co. I-395	171	24				−1.8
E Fairfax Co. Beltway I-66-I-95	175	24				−0.6
E Fairfax Co. Beltway	176	24	0.6		1.4	
N Fairfax Co. inbound arterial link	181	25	−0.2	−1.0	−0.2	
N Fairfax Co. Beltway	189	25			−1.2	
S Fairfax Co. I-95 South of Beltway	197	26				**15.0**
S Fairfax Co. I-95 South of Beltway	198	26	−0.3		2.1	
NW Fairfax Co. outbound arterial link	208	27	−4.5	0.0	−0.1	
NW Fairfax Co. VA 267	211	27	1.7		−0.1	
NW Fairfax Co. I-66 West of Beltway E-bound	212	27	−0.6	−4.2	0.8	
NW Fairfax Co. I-66 West of Beltway W-bound	213	27	0.3		1.0	
E Loudon Co. outbound arterial link	223	28	−1.8			
S Prince William Co. I-95 South of Beltway	235	30				−3.3
S Prince William Co. I-95 South of Beltway	236	30	−0.1		2.1	
Other Links	–	–	−13.7	−2.8	7.4	−2.4
Total			110.1	84.5	46.2	4.7
Number of links affected > 1% × MCC own link			13	5	20	9

5. Conclusions

This paper investigates the question of whether marginal congestion costs computed on a network using a realistic urban-scale model significantly differ from the marginal congestion costs computed based on individual congestion functions on each link. The strategic transportation planning model START was calibrated for Washington DC metropolitan area and used to simulate the impact of changes in flow on network links.

We have concluded that a straightforward link-by-link method can be utilized to compute the region-wide average levels of marginal congestion costs. Such average values could be applied to estimate spatially aggregate policies (such as fuel tax or even cordon toll (Santos, 2004a)). However, this method may not be appropriate for designing finer policies such as geographically differentiated congestion tolls. Also, we observe that DC area urban network is a good conductor of traffic congestion. Therefore, policymakers should be careful with implementation of transportation pricing policies in one part of the urban area without taking into account probable changes in congestion in other parts.

An important caveat of this research is that the START model employed in this study treats routes chosen by travelers exogenously and the route choice as probabilistic. Although we have no reason to believe that removing those features can significantly change our results, it would be useful to investigate the issue of network proliferation of congestion using standard traffic assignment networks, bottleneck-style models (e.g., De Palma et al., 2005), models explicitly incorporating queuing at intersections and traffic lights (e.g., Dewees, 1979), or other methods of modeling road supply (e.g., May et al., 2000). It would be interesting to see whether using other models or applying START to other urban areas can corroborate the results of this paper.

Finally, from a policy perspective, this paper indicates that theoretically appealing marginal congestion costs containing networks effects could potentially be applied to design first-best marginal pricing schemes on urban transportation networks.

Acknowledgements

The authors would like to thank Phil Goodwin, Winston Harrington, Lars Lundqvist, Peter Nelson, Michael Roberts, and Kenneth Small for helpful suggestions. All remaining errors are exclusively ours.

References

Anderson, D., Mohring, H., 1997. Congestion costs and congestion pricing. In: Greene, D.L., Jones, D.W., Delucchi, M.A. (Eds.), The Full Costs and Benefits of Transportation: Contributions, Theory and Measurement. Springer, New York, pp. 315–336.

Armelius, H., 2005. An integrated approach to urban road pricing. Journal of Transport Economics and Policy 39 (1), 75–92.

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

Council of Governments, 1999. Traffic Quality on the Metropolitan Washington Area Freeway System. Prepared by Skycomp, Inc, Columbia, MD, for the Washington Metropolitan Council of Governments.

Croombe, D.P., Scholefield, G., Skinner, A., 1997. Study of parking and traffic demand. Part 3: the effects of parking control strategies in Bristol. Traffic Engineering and Control 38 (4).

DeCorla-Souza, P., 2004. Recent US experience: pilot projects. In: Santos, G. (Ed.), Road Pricing: Theory and Evidence. Elsevier Science, Amsterdam, pp. 179–206.

De Palma, A., Kilani, M., Lindsey, R., 2005. Congestion pricing on a road network: a study using the dynamic equilibrium simulator METROPOLIS. Transportation Research A 39, 588–611.

Dewees, Donald N., 1979. Estimating the time costs of highway congestion. Econometrica 47, 1499–1512.

Graham, D.J., Glaister, S., 2006. Spatial implications of transport pricing. Journal of Transport Economics and Policy 40 (2), 173–201.

Hearn, D.W., Yildirim, M.B. 2002. A Toll pricing framework for traffic assignment problems with elastic demand. In: Michel Gendreau, Patrice Marcotte, (Eds.), Current Trends in Transportation and Network Analysis – Papers in Honor of Michael Florian. In: Kluwer Applied Optimization Series, vol. 63, pp. 135–145.

Jenelius, E., Petersen, T., Mattsson, L.-G., 2006. Importance and exposure in road network vulnerability analysis. Transportation Research A 40, 537–560.

Lee, D.B., 1997. Uses and meanings of full social cost estimates. In: Greene, D.L., Jones, D.W., Delucchi, M.A. (Eds.), The Full Costs and Benefits of Transportation: Contributions, Theory and Measurement. Springer, New York, pp. 113–148.

Liu, L., McDonald, J.F., 1998. Efficient congestion tolls in the presence of unpriced congestion: a peak and off-peak simulation model. Journal of Urban Economics 44, 352–366.

May, A.D., Roberts, M., Mason, P., 1992. The Development of transport strategies for Edinburgh. In: Proceedings of the Institution of Civil Engineers: Transportation, vol. 95, pp. 51–59.

May, A.D., Shepherd, S.P., Bates, J.J., 2000. Supply curves for urban road networks. Journal of Transport Economics and Policy 34, 261–290.

Mayeres, I., Proost, S., 2001. Marginal tax reform, externalities, and income distribution. Journal of Public Economics 79 (2), 343–363.

Nelson, P., Gillingham, K., Safirova, E., 2003. Revving Up the Tax Engine: Gas Taxes and the DC Metro Area's Transportation Dilemma. Issue Brief 03–05. Resources for the Future, Washington, DC.

Nelson, P., Baglino, A., Harrington, W., Safirova, E., Lipman, D.A., 2006. Transit in Washington, DC: Current Benefits and Optimal Level of Provision. Discussion Paper 06-21. Resources for the Future, Washington, DC.

Newberry, David, 1990. Pricing and congestion: economic principles relevant to pricing roads. Oxford Review of Economic Policy 6 (2), 22–38.

Ozbay, K., Bartin, B., Berechman, J., 2001. Estimation and evaluation of full marginal costs of highway transportation in New Jersey. Journal of Transportation Statistics 4 (1), 81–104.

Parry, I.W.H., Bento, A.M., 2001. Revenue recycling and the welfare effects of congestion pricing. Scandinavian Journal of Economics 103, 645–671.

Ramjerdi, F., Minken, H., Øsmoe, K., 2004. Norwegian urban tolls. In: Santos, G. (Ed.), Road Pricing: Theory and Evidence. Elsevier Science, Amsterdam, pp. 179–206.

Safirova, E., Gillingham, K., Harrington, W., Nelson, P. 2003. Are HOT Lanes a Hot Deal? The Potential Consequences of Converting HOV to HOT Lanes in Northern Virginia. RFF Issue Brief 03-03.

Safirova, E., Parry, I.W., Gillingham, K., Harrington, W., Nelson, P., Mason, D., 2004. Welfare and distributional effects of road pricing schemes for metropolitan Washington, DC. In: Santos, G. (Ed.), Road Pricing: Theory and Evidence. Elsevier Science, Amsterdam, pp. 179–206.

Safirova, E., Gillingham, K., Harrington, W., Nelson, P., Lipman, D., 2005. Choosing congestion pricing policy: cordon tolls vs. link-based tolls. Transportation Research Record 1932, 169–177.

Safirova, E., Houde, S., Lipman, D., Harrington, W., Baglino, A. 2006a. Congestion pricing: long-term economic and land use effects. Resources for the Future, Discussion Paper 06-37.

Safirova, E., Houde, S., Coleman, C., Harrington, W., Lipman, D. 2006b. A small cordon in the hand is worth two in the bush: long-term consequences of road pricing. Resources for the Future, Discussion Paper 06-42.

Santos, G., 2004a. Urban congestion charging: a second-best alternative. Journal of Transport Economics and Policy 38 (3), 345–369.

Santos, G., 2004b. Urban road pricing in the UK. In: Santos, G. (Ed.), Road Pricing: Theory and Evidence. Elsevier Science, Amsterdam, pp. 179–206.

Schrank, D., Lomax, T., 2005. The 2005 Urban Mobility Report. Texas Transportation Institute, Texas A&M University College Station, TX.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice Hall, Englewood Cliffs, NJ.

Verhoef, E.T., 2002. Second-best congestion charging in general networks. Heuristic algorithms of finding second-best optimal toll levels and toll point. Transportation Research B 36, 707–729.

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

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

Zhang, L., Levinson, D., 2006. Economics of road network ownership. Paper presented at the First International Conference on Funding Transportation Infrastructure, Banff, Canada, August 2006.