

2010

Joint Transport Research Centre

*ROUND TABLE*

*4-5 February 2010, Paris*

2

DISCUSSION PAPERS



## **ROAD PRICING WITH COMPLICATIONS**

*Mogens FOSGERAU, DTU Transport, Denmark &  
Centre for Transport Studies, Sweden*

*Kurt VAN DENDER*

*OECD/ITF Joint Transport Research Centre*





ORGANISATION  
FOR ECONOMIC  
CO-OPERATION AND  
DEVELOPMENT



JOINT TRANSPORT RESEARCH CENTRE

---

## Discussion Paper No. 2010-2

Prepared for the ITF/OECD Round Table of 4-5 February 2010 on  
**Implementing Congestion Charging**

### Road pricing with complications

**Mogens FOSGERAU**

DTU Transport, Denmark &  
Centre for Transport Studies, Sweden

**Kurt VAN DENDER**

OECD/ITF Joint Transport Research Centre

January 2010

*The views expressed in this paper are the authors', and do not necessarily represent the opinions of DTU Transport, Denmark, the Centre for Transport Studies, Sweden, the International Transport Forum or the OECD.*

# Road pricing with complications

Mogens Fosgerau

DTU Transport, Denmark & Centre for Transport Studies, Sweden

Kurt Van Dender

Joint Transport Research Center, OECD

## Abstract

Standard textbook analyses of road pricing tend to assume that users are homogenous, that there is no travel time risk, and to have a view of congestion as static. The simple analysis also ignores that real pricing schemes are only rough approximations to ideal systems and that the general economic context may also have implications for optimal pricing. This paper reviews these issues and discusses how taking them into account may affect estimates of optimal tolls.

## INTRODUCTION

Road pricing can be used as a tool to reduce demand for travel when and where that is thought to be beneficial. It has some important advantages over other ways of reducing demand. By adding a toll to the cost of a trip, it removes just those trips that travellers themselves think are not worth the toll. So tolling ensures that the least beneficial trips are reduced first with drivers themselves assessing the benefits of their trips. Tolling allows the individual decisions about whether, when and where to travel to remain decentralised. A toll equal to the total cost of delay imposed on other drivers by one additional car, the marginal external congestion cost, ensures that the socially optimal level of congestion results.<sup>1</sup> Exactly those trips are then carried out that are worth the full cost. The toll payment is a loss for drivers, but the money does not disappear. The cost to drivers who choose to pay the toll is

---

<sup>1</sup> This is true in a first-best world, where all other prices equal marginal social cost.

offset by the gains of those who receive the payment.<sup>2</sup> Overall, the gain from pricing congestion can be considerable.

An obstacle to the introduction of road pricing is the fact that all drivers have to pay in order to deter the least beneficial trips. This transfer from drivers is only fully compensated by reduced travel times in some special circumstances. Drivers as a group will therefore tend to loose when the use of the revenues from road pricing is not taken into account.

The efficient toll is in effect a price on time. It reflects the value of the delays an additional driver imposes on other drivers. To design an efficient toll the first requirement is then to find the value of travel time (VTT).<sup>3</sup> This leads to the concept of generalised travel cost (GTC) that includes both monetary and time costs. For a certain trip, the GTC may be calculated as the monetary costs of the trip plus the duration of the trip times the VTT; in symbols  $GTC = c + \alpha t$ , where  $c$  is the monetary trip cost,  $t$  is the travel time and  $\alpha$  is the VTT.

Armed with an estimate of the VTT, the basic idea behind congestion pricing is simple: A supply curve relates traffic volume to delay. This can be used to compute the increase in travel time resulting from an additional car. Multiplying the increase in travel time by the traffic volume, the demand, yields the total delay caused by an additional car. Next, multiplying by the VTT yields the associated marginal external cost of congestion. A demand curve relates traffic volume to  $GTC$ . This can be used to find the efficient toll which is equal to the marginal external cost of congestion when the toll is implemented.

The point of this paper is to present an overview of what happens when one opens the door for some of the complications that are present in the real world.

The first complication is that travellers are not identical. On the contrary, they seem to be very heterogeneous. The evidence indicates that the VTT in a population can easily range over several orders of magnitude. There is variation in the VTT among individuals. There is even variation for the same individual depending on the exact context of a trip. The paper explores some consequences of this heterogeneity.

The second complication is that travel time is not a deterministic function of traffic flow. Travel times increase on average as flow approaches capacity. Inherent randomness in traffic conditions also gains importance and causes travel times to become more variable and less predictable. The associated cost is significant in comparison to the cost associated with the average travel time.

The third complication is that the relationship between traffic flow and travel time is not simple. There are important dynamics that must be taken into account. It is important to recognise that congestion is time specific. At any location there is sufficient capacity to allow

---

<sup>2</sup> This is also true in a first-best world.

<sup>3</sup> Becker [6] related the value of time to the wage rate by considering the allocation of time between work and leisure. This was later extended in various ways. Jara-Diaz [21] tells the history of the development of this theory.

traffic to flow, if only it was equally distributed over time.<sup>4</sup> Congestion occurs largely because people tend to want to travel at the same time. It is therefore important to recognise that the time of travel is a choice and that congestion is about time-specific preferences: scheduling preferences.

The fourth complication is that actual tolling systems are imperfect approximations to efficient tolls. For example it may only be possible to toll some links or it may not be possible (or desirable) to achieve the theoretically optimal spatial and temporal differentiation of tolls. Some implications of such limitations are explored.

Finally, tolls cannot be evaluated in isolation from the economy in general. In particular, commuting, a large part of private travel, is complementary to working. Hence the interaction of tolls with the labour market and labour income taxes must be considered.

Section 2 reviews the standard textbook analysis of tolls. Section 3 considers some implications of the fact that travellers are heterogeneous. Section 4 discusses issues that result since congestion also causes travel times to become variable and unpredictable for travellers. Section 5 discusses the need for a sophisticated understanding of the dynamics of the transport system. Section 6 discusses second-best issues related to constraints on the design of road pricing schemes and interactions with the wider economy. Section 7 concludes.

## THE STANDARD TEXTBOOK ANALYSIS

From an economist's perspective, the core problem of congestion is not that there is congestion but that there is too much of it. It is not congestion itself that is the problem, but the associated externality. Each driver takes only his own costs into account when deciding whether to travel. The individual will choose to drive when his benefit of doing so outweighs his cost of travelling such that he receives a personal net gain. However, he does not take into account the delay he imposes on others. When that delay is large then the trip entails a net loss. There is thus a potential gain that can be found if traffic can be reduced in such a way that only those travel whose personal gain from travelling outweighs the delay imposed on others.

### A static model

The standard analysis under simplifying assumptions is presented in the following figure. The horizontal axis is the traffic volume; the vertical axis is the *GTC*. The curve  $\alpha S$  is the supply curve. The supply *S* is the travel time as a function of traffic volume; this is multiplied by the VTT  $\alpha$  to convert travel time to monetary cost.<sup>5</sup> The supply curve is increasing to

---

<sup>4</sup> The alternative would require queues that keep growing indefinitely.

<sup>5</sup> Other monetary travel costs are ignored.

reflect a situation with congestion. It will generally be convex. All travellers experience the same travel time and hence the supply curve can also be understood as an average cost curve. The curve MC indicates the marginal cost. It is the change in total cost arising from an additional traveller. When the supply curve is increasing, the marginal cost curve will always be increasing, have a larger slope and lie above the supply curve.

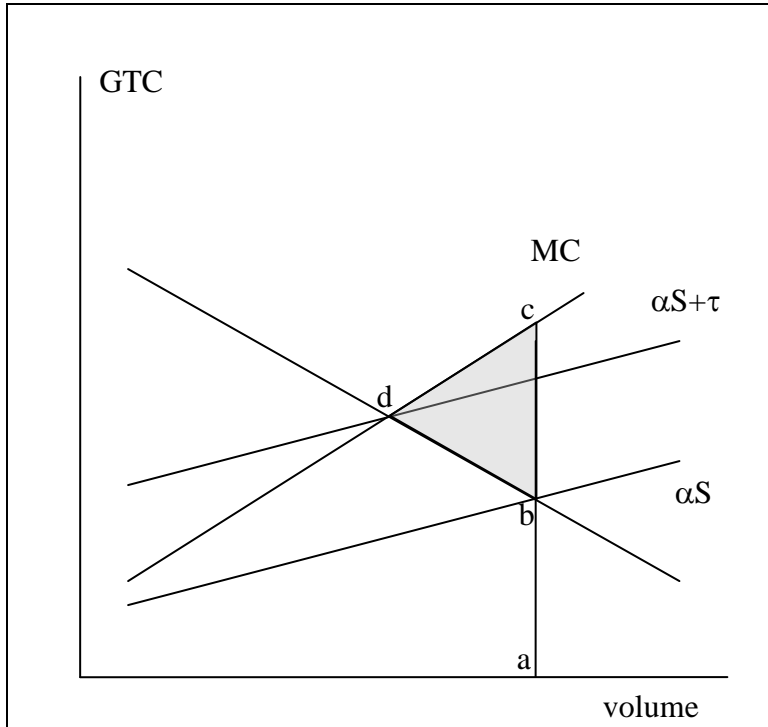


Figure 1. A static model

The demand curve is decreasing to reflect that demand decreases in price. The market equilibrium occurs at the intersection of the demand curve  $D$  with the supply curve  $\alpha S$ , at the point  $b$ . The marginal traveller at this point faces a cost corresponding to the line segment  $a-b$  and a corresponding benefit of the same size. The group of travellers as a whole, however, have a cost associated with the marginal traveller given by the  $MC$  curve. For the marginal traveller at point  $b$ , this cost corresponds to the line segment  $a-c$ . So the last traveller imposes a net loss corresponding to the line segment  $b-c$  on the group of all travellers. The loss is zero at the point  $d$  where the  $MC$  curve crosses the demand curve. The total loss is then represented by the shaded triangle  $b-c-d$  on the figure. The optimal toll implements the equilibrium at the point  $d$ .

Expressed symbolically, the marginal external cost of congestion (mecc) is  $mecc = S'(D) \cdot D \cdot \alpha$ . This arises since the total cost given realized demand  $D$  is  $TC(D) = S(D) \cdot D \cdot \alpha$ , the marginal total cost is  $TC'(D) = S'(D) \cdot D \cdot \alpha + S(D) \cdot \alpha$ , and the mecc is the difference between the marginal cost and the average cost.

We shall refer to this analysis as the 'simple static analysis'. It is useful as a basis for discussing more complicated models that take more aspects of reality into account.

## A dynamic model – the bottleneck model

The static analysis ignores the trip timing aspect of travel demand. Time is regarded merely as a resource of which a traveller can consume more or less, whereas in reality time at 8 AM is different from time at 9 AM. As has been noted, congestion arises because travellers prefer to travel around the same times. If they could be spread evenly over time, there would be no congestion. Moreover, congestion is inherently dynamic, since adding a vehicle to a queue at some instant will affect the evolution of the queue until it is gone. These aspects are captured in the bottleneck model [1, 36], which describes the time dimension more explicitly. This model generates a number of important insights regarding the pricing of congestion. A main insight is that large efficiency gains may be found through the effect of pricing on trip timing.

In the simplest incarnation of the bottleneck model, travellers are viewed as homogenous with VTT  $\alpha$ . They have a common preferred arrival time  $t^*$  and prefer not to be early or late at the destination relative to this time. The cost of earliness is usually denoted by  $\beta$  and the cost of lateness by  $\gamma$  and the generalised travel cost is then

$GTC(a, t) = \alpha \cdot (t - a) + \beta \cdot \min(0, t^* - t) + \gamma \cdot \max(0, t - t^*)$ . This specification of scheduling preferences was introduced by Vickrey [36] and later by Small [28] in a different context.<sup>6</sup>

A total of  $D$  travellers have to pass through a bottleneck in order to reach their destination. The bottleneck has a limited capacity of  $\psi$  users per minute. Assume for simplicity that travel time is zero before and after the bottleneck. Then the first traveller, departing at time  $t_0$  also arrives at the destination at this time, since there is no queue yet at this time. Denote the cumulative arrival rate at the bottleneck by  $R$ . Then  $R(t_0) = 0$ . Let the time of the last departure be  $t_1$ , such that  $R(t_1) = D$ .

The model assumes an equilibrium where no traveller can reduce his cost by changing departure time. This implies first that there is always queue during the interval  $[t_0, t_1]$  and second that the queue is exactly gone at time  $t_1$ , such that  $\psi \cdot (t_1 - t_0) = D$ . Third, equilibrium requires that the  $GTC$  is constant over the interval where travellers depart and higher outside. A queue builds up immediately as the first traveller departs, since travellers initially depart at a higher rate than capacity. The queue has maximum length at the departure time when a traveller would be at the destination exactly at the preferred time  $t^*$ . From that point, the departure rate is lower than capacity such that the queue gradually dissipates and the queue is exactly gone at the time of the last departure. This is illustrated in Figure 2.

---

<sup>6</sup> [38] develops the bottleneck model under more general scheduling preferences, while allowing for random queue sorting.

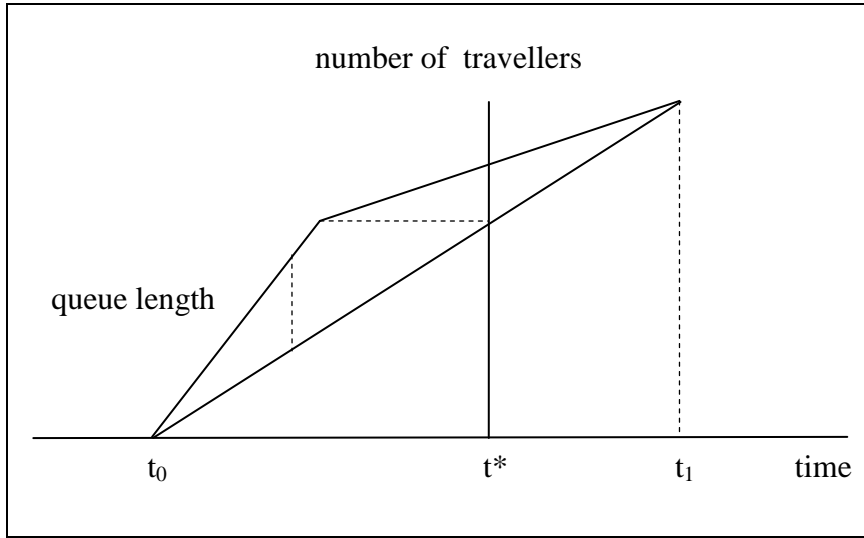


Figure 2

The first and last travellers experience no queue in this model. The first traveller is early at the destination while the last traveller is late. They incur the same  $GTC$  in equilibrium, which implies that  $-\beta \cdot (t^* - t_0) = \gamma \cdot (t_1 - t^*)$ . This fixes the interval  $[t_0, t_1]$  such that the equilibrium travel cost can be computed. The total travel cost in equilibrium for all travellers

is given by  $TC = \frac{\beta\gamma}{\beta + \gamma} \frac{N^2}{\psi}$ , such that the marginal external congestion cost is

$mecc = \frac{\beta\gamma}{\beta + \gamma} \frac{N}{\psi}$ . This is the marginal external congestion cost associated with the addition of

a marginal user to the equilibrium. It increases in the number of users and decreases in capacity.

We may regard the number of travellers  $D$  as being a function of the equilibrium generalised travel cost. Then connecting the bottleneck model with the simple static analysis, we would find that the optimal static toll is equal to the  $mecc$ . This toll would not remove congestion. The number of travellers would be reduced by the toll, but there would still be a queue during  $[t_0, t_1]$ .

This static toll does not vary over time. It is the same for all travellers, regardless of when they arrive at the bottleneck. It is hence only optimal if the toll must be constant over the day. It is possible to do better by letting the toll vary over time. The optimal time-varying toll is zero at time  $t_0$ . It increases until the preferred arrival time  $t^*$ , then it decreases again, until it is zero at time  $t_1$ . The average toll is equal to the optimal static toll. The optimal time-varying toll does remove congestion completely, since it ensures that travellers arrive at the bottleneck exactly at the rate  $\psi$ , which is the bottleneck capacity.



## HETEROGENEOUS TRAVELLERS

Where the simple analysis assumes that users are heterogeneous, it is clear that in reality users are heterogeneous. This section explores first the measurement of heterogeneous VTT, and second the consequences of heterogeneity of the VTT for the optimal congestion charge.

### Measurement of heterogeneous VTT

How can the VTT of travellers be inferred? If a traveller has a choice between two options for making a trip, where one is faster but more expensive than the other, then the traveller faces a trade-off between money and time. Assume that travellers have a generalised cost of a trip expressed as  $GTC = C + \alpha \cdot T$ , where  $C$  is the monetary cost of the trip,  $T$  is the travel time, and  $\alpha$  is an individual specific VTT. The difference in GTC between two trip options for a traveller with VTT equal to  $\alpha$  is  $\Delta GTC = \Delta C + \alpha \cdot \Delta T$ . Holding everything else constant, travellers with  $\alpha < -\Delta C/\Delta T$  will choose the slow option while travellers with  $\alpha > -\Delta C/\Delta T$  will choose the fast option. The trade-off thus entails an implicit price of travel time, namely  $v = -\Delta C/\Delta T$ . Through his choice, the traveller reveals whether his VTT is larger or smaller than the trade-off price, i.e. whether  $\alpha < v$  or  $\alpha > v$ .

Label by  $\Phi$  the cumulative distribution function describing the distribution of the VTT among users. Observing many travellers at a trade-off price  $v$  allows assessment of the share of travellers with  $\alpha < v$ . This share is the cumulative distribution evaluated at the point  $v$ , i.e.  $\Phi(v)$ . Observing travellers in different choice situations with different  $v$  allows assessment of  $\Phi$  over the range where  $v$  varies. To assess  $\Phi$  completely, it is necessary to observe choice shares for values of  $v$  ranging from a point where travel time is cheap and all travellers choose the fast and expensive option to a point where time is expensive and all travellers choose the cheap and slow option.<sup>7</sup>

In general, revealed preference data are preferable by virtue of relating to real choices where travellers actually feel the consequences of their choices. Suitable revealed preference data could come from situations where travellers face a choice between a slow and cheap route and a fast and expensive route. A notable example is routes combining free access regular freeway lanes with tolled express lanes [31]. It is however often difficult to achieve the necessary variation in the price of time needed to reveal the distribution of the VTT. This is a reason for relying on stated preference data, where travellers make choices between hypothetical options. It is possible to construct stated preference choice situations to meet many of the demands of econometric modelling. Stated preference data are, however, tainted by doubt whether they represent actual behaviour well.

---

<sup>7</sup> This may sound easier than it is. Fosgerau [12] discusses issues related to the identification of the distribution of VTT. De Borger & Fosgerau [10] discuss extension to take behavioral anomalies into account.

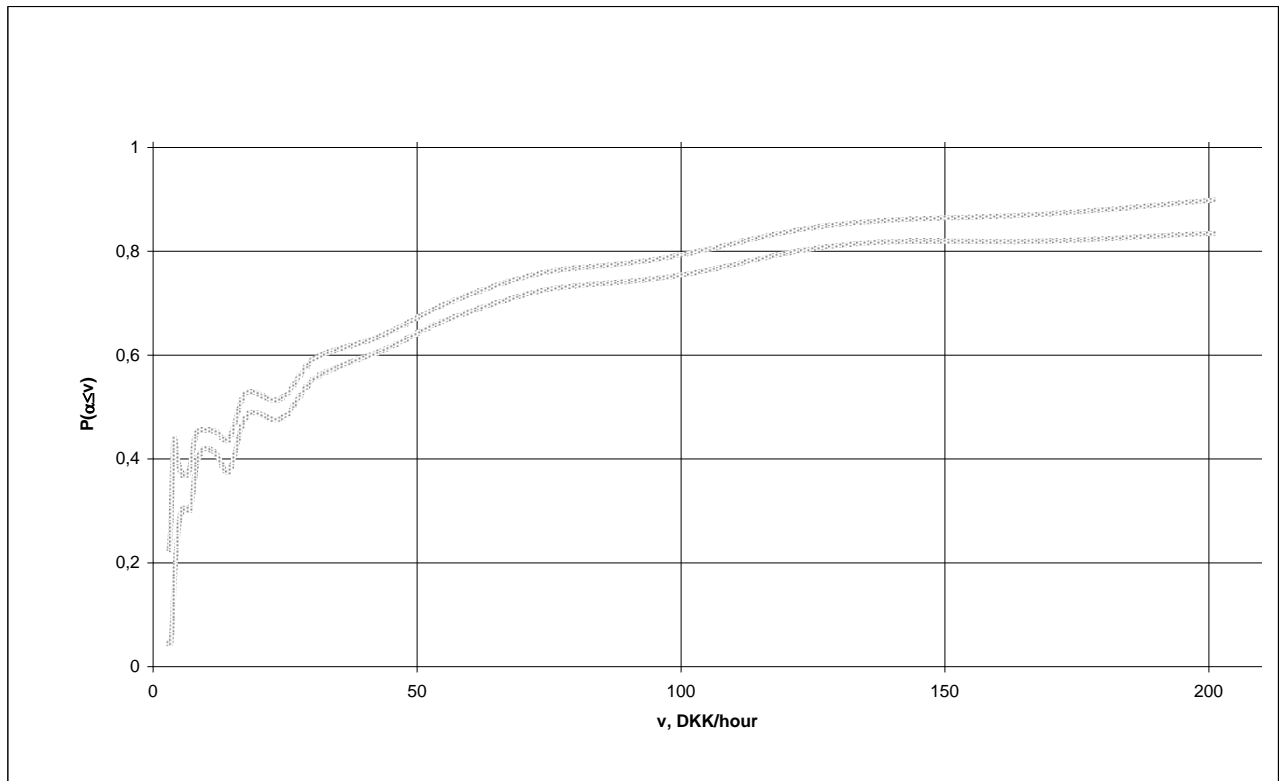


Figure 3. Confidence band for the cumulative distribution of VTT based on SP data. The unit is Danish Crowns (DKK) per hour: 1 EUR  $\approx$  7.5 DKK.

Figure 3 shows an estimate of a VTT distribution obtained from stated preference data [12].<sup>8</sup> The shape is broadly typical of many studies. It shows a right-skewed distribution with many travellers having low VTT and few travellers having large VTT. The median VTT is about 25 DKK/hour while the mean is considerably larger.

This estimate of the cumulative distribution of the VTT does not show the maximum of the VTT distribution. The largest trade-off price of time that was offered to respondents in the SP exercise was about 200 DKK/hour. A significant share of respondents, about 15%, indicated that they were willing to pay this amount per hour of travel time saved and hence that their VTT was higher than 200 DKK/hour. How much larger is impossible to say based on the figure. It is not possible to calculate the mean VTT without information about the right tail of the VTT distribution. Such information can be found using data that allows the right tail to be observed. Otherwise it is necessary to resort to more restrictive assumptions, which may be hard to justify. One popular approach is to assume a specific form for the VTT distribution. The resulting estimate of the mean VTT is extremely sensitive with respect to such an assumption [12].

<sup>8</sup> The estimate is computed using a nonparametric technique, which does not impose the restriction that the cumulative distribution should be increasing. It is therefore evidence of the internal validity of the SP data that an increasing function does result.

It is clear that there is enormous variation in the VTT with several orders of magnitude from low to high. The VTT depends on observable factors. It is generally found to increase with income but the size of the income elasticity is debated. The VTT is generally thought to vary substantially between individuals but also within individuals depending on the context. In general, a large part of the variation in VTT remains after controlling for observable factors. E.g., Fosgerau [12] controls for gender, income, trip duration, time difference between alternatives, share of delay time due to congestion in travel time, age and trip purpose, and finds that the remaining variation in VTT has more than a factor 50 between the 20% and 80% quantiles.

### Road pricing with heterogeneous travellers

Consider now a situation where the population of potential travellers differ in their VTT. Groups of travellers are indexed by their value of travel time  $\alpha$ . The group with VTT  $\alpha$  has a demand function  $D(p(\alpha)|\alpha)$ , where  $p(\alpha) = \tau + \alpha t$  is the generalised cost for travellers  $\alpha$ . The aggregated demand function is then  $D = E(D(\alpha))$ , the average over groups of travellers. The average VTT in the population is  $E(\alpha)$ . This is not the same as the average VTT of those who actually travel, which is the weighted average  $\bar{\alpha} = E[\alpha \cdot D(\alpha)] / E[D(\alpha)]$ . If travel time as a function of demand is  $t = S(D)$ , then the change in travel time resulting from the marginal traveller is  $S'(D)$ . Multiplying this by the number of travellers and by the average VTT among travellers indicates the marginal external cost of congestion. That is  $mecc = S'(D) \cdot D \cdot \bar{\alpha}$ .

This shows that the difference from the case of homogenous travellers is the VTT used to compute the marginal external cost of congestion. There is just a single VTT in the case of homogenous travellers. With heterogeneous travellers, this single VTT is replaced by the average VTT in the group of actual travellers. Hence the *mecc* depends in general on the toll, since introduction of a toll will change the composition of travellers and this affects the average VTT of road users.

This effect may be large. Consider a case where the VTT in the population follows a standard lognormal distribution. Then the mean VTT is  $E(\alpha) \approx 1.65$ . Imagine now a toll that causes a reduction in traffic of 10% and that it is those travellers with VTT above the 10% quantile that remain. Their average VTT is then about  $\bar{\alpha} \approx 1.81$ , which represents an increase of about 10%.

A toll will discourage some from travelling. If travellers with low VTT are discouraged more, as might be expected, then the average VTT  $\bar{\alpha}$  increases with the introduction of a toll. Then the optimal toll will be larger when travellers are heterogeneous than when they are homogenous.<sup>9</sup>

---

<sup>9</sup> Arnott, de Palma & Lindsey [2], discuss pricing with heterogeneous travellers in the context of the bottleneck model. Arnott & Kraus [4] discuss marginal cost pricing when travellers are heterogeneous but the differences are not observed.

We are not aware of empirical evidence concerning the likely size of this effect. It is reasonable to suppose that the effect is noticeable, since the distribution of VTT is generally thought to have a shape similar to that presented in Figure 3 with many travellers having a relatively low VTT.

Consider still the situation with heterogeneous travellers and imagine that the optimal toll  $\tau = mecc = S'(D) \cdot D \cdot \bar{\alpha}$  is in operation. Imagine then that road capacity is split in two halves and that travellers have to choose which part of capacity they want to use. Then they will divide equally on the two parts of capacity. If the toll is increased slightly on one part of capacity, then demand will decrease slightly there and shift to the other part or stop travelling on either part. Then the first part will be faster but more expensive than the other part. This will cause rational travellers to sort, such that those with the VTT above some threshold will use the first part and those with the VTT below the threshold will use the other. As a consequence, the average VTT is high on the first part and low on the second. Then the toll can be raised on the first part and reduced on the second part to produce a net welfare gain.

Verhoef & Small [35] consider differentiated tolls in a static network with serial and parallel links and with heterogeneous users. They are particularly concerned with second-best policies whereby only a part of the network is tolled. Such policies, they find, are in danger of losing much of their potential effectiveness if heterogeneity is ignored when setting toll levels. Ignoring heterogeneity in VTT may cause the welfare benefits of second-best policies to be drastically underestimated.

A real-world counterpart to such examples is value pricing as implemented various places in the US. In these schemes, drivers can choose between free but congested lanes and tolled but less congested lanes. Clearly, as shown by Small & Yan [29], heterogeneity is important for assessing the welfare impacts of such policies as drivers with high VTT may be expected to choose the tolled lanes.

Pricing is not the only means to increase the efficiency with which existing road capacity is used. A recent paper [9] shows that there are unexpected benefits from car pool lanes that do not have to do with pricing. The benefit arises because the car pool lanes reduce disruptive vehicle lane changing. Even a severely underused carpool lane can in some instances increase a freeway bottleneck's total discharge flow.<sup>10</sup>

In a dynamic setting there can be benefits from differentiation, even with homogenous travellers and without pricing. Ongoing research (Fosgerau & de Palma) uses the bottleneck model to analyse a situation in which travellers are arbitrarily divided into groups. The immediate cause of congestion in the bottleneck model is that travellers initially depart at a rate that is higher than capacity. Congestion is reduced by a toll that makes travellers decrease the initial departure rate. This effect may be induced in other ways. One way is to divide

---

<sup>10</sup> A theoretical investigation of these issues is undertaken in [23].

travellers into, say, two groups. A more than proportional share of capacity is allocated to the first group. The second group can use the remaining capacity. The second group can also use the share of capacity allocated to the first group when the first group is not using it. The first group would find a new equilibrium where departures occur during a shorter interval of time and would hence experience a cost reduction relative to the situation without grouping. The cost of the second group would not increase since it is determined by the length of the interval during which departures take place and this is unchanged relative to the situation without grouping.

### TRAVEL TIME RISK

Increasing travel demand leads to congestion and increasing travel times. As demand approaches capacity, travel times also become increasingly variable and unpredictable for users. This travel time variability (TTV) may add significantly to the generalised travel cost.

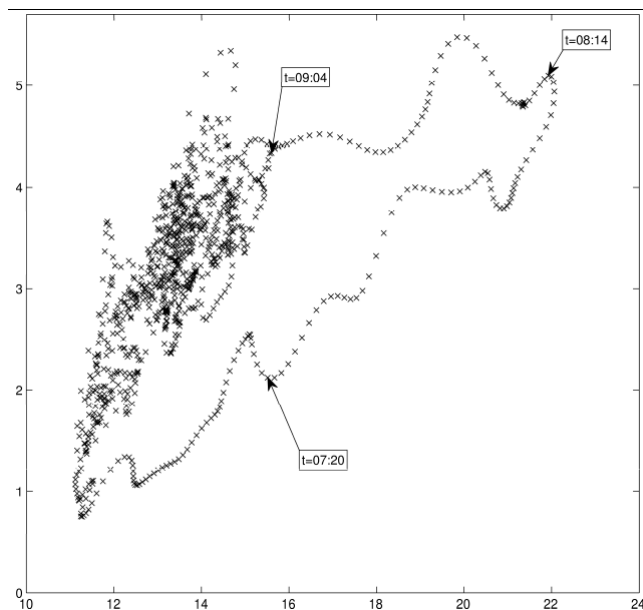


Figure 4. Scatter plot of the standard deviation of travel time (minutes) against the mean travel time (minutes) for a congested urban road.

Figure 4 shows a scatter plot of the standard deviation of travel time against the mean travel time for a congested urban road with a distinct morning peak. Each point on the plot corresponds to a time of day. The numbers have been computed using data covering a period of three months. Both the mean and the standard deviation are small in the morning. The mean and standard deviation increase and then decrease over the peak. The standard deviation peaks later than the mean, indicating that there is not a constant relation between the mean

and the standard deviation. This creates the loop that is evident on the figure. It is a characteristic pattern that has been observed many times.<sup>11</sup> An implication is that the mean and the standard deviation of travel time must be accounted for separately.

The definition of GTC must be extended in order to account for the cost associated with travel time variability. Some authors have simply added a term to the definition of GTC to reflect the increase in cost due to TTV. For example using  $GTC = c + \tau + \alpha T + \eta X$ , where  $X$  is some measure of TTV. Various measures of TTV have been employed such as the standard deviation or the variance of travel time or a range between two quantiles [31]. Studies have then proceeded to estimate a value  $\eta$  of  $X$  based on revealed or stated preference data.

This is however not a completely satisfactory solution without some arguments to indicate why one measure of TTV should be preferred to another. The problem is complicated since a travel time distribution is a shape rather than a number. **Figure 5** shows an example of a travel time distribution. There are (infinitely) many possible such shapes and they cannot be described completely by a few numbers.

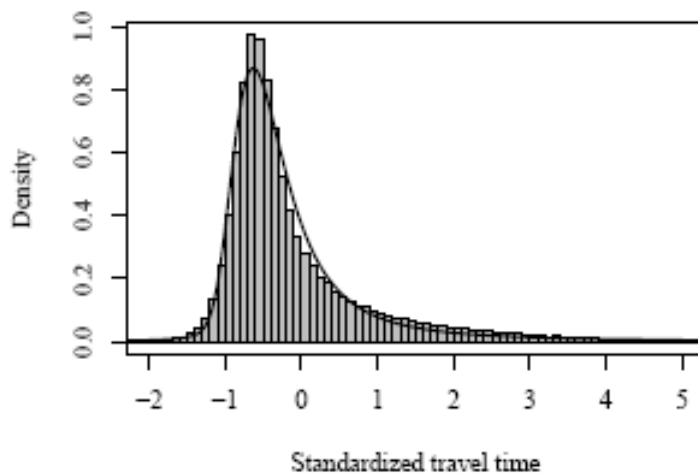


Figure 5. An empirical travel time distribution

At this point it can be noted that all measures of scale are essentially equivalent when the shape of the travel time distribution is constant. In this case, the standard deviation is proportional to the range between any two given quantiles. A change from one scale measure to another is then reflected in an inverse change in the value of TTV given by  $\eta$ .

Intuitively, the cost associated with TTV is related to scheduling considerations. Compare two situations, one in which travel time is variable and one in which it is constant. The mean travel time is the same in both situations. Travellers have to decide when to embark on a certain trip. When travel time is constant, travellers choose an optimal time of departure which is directly associated with an optimal time of arrival at the destination. Faced with

<sup>11</sup> This pattern is generated by the random capacity bottleneck model for any distribution of capacity [13].

TTV, travellers may embark on the trip earlier than they otherwise would have. On average they will therefore arrive earlier than otherwise. Sometimes they arrive later than they would have chosen with constant travel time.

To make this more formal, economic theory generally assumes that travellers have preferences that encompass scheduling considerations regarding when they depart from the origin of the trip and when they arrive at the destination. Travellers are pictured as knowing the travel time distribution and choosing the departure time optimally. A specification of scheduling preferences then leads to a relationship between the travel time distribution and the GTC. This relationship is not in general tractable and there is generally no obvious candidate for defining a measure of TTV. There are however a few special cases, where simplifying assumptions enable a simple measure of TTV to be defined.

Fosgerau & Karlstrom [15] consider the departure time choice of a traveller facing TTV. The traveller cares about travel time and about being early or late at the destination according to the Vickrey/Small scheduling preferences described above. The distribution of random travel is independent of the departure time, such that a change in departure time does not affect the shape of the travel time distribution but only shifts it earlier or later. Similarly, the monetary trip cost does not depend on the departure time.

When the traveller knows the travel time distribution and chooses the optimal departure time, it turns out that the expected GTC becomes linear in the mean and the scale of the travel time distribution, regardless of what the travel time distribution is.<sup>12</sup> More specifically, when the traveller chooses the optimal departure time, then  $GTC = c + \alpha \cdot E(T) + \eta \cdot \sigma \cdot H$ . In this expression,  $c$  is the monetary cost of the trip,  $\alpha \cdot E(T)$  is the VTT multiplied by the mean travel time. The last term captures the effect of TTV:  $\eta$  is the value associated with TTV and is given in terms of scheduling parameters  $\beta$  and  $\gamma$ ;  $\sigma$  is a measure of the scale of the travel time distribution and  $H$  depends on scheduling parameters and on the shape of the travel time distribution but not on the scale of the travel time distribution.

Fosgerau & Karlstrom provide an example of a congested urban road in Copenhagen where the cost of TTV comprises about 15% of the time cost to travellers. Including TTV in the GTC is likely also to lead to an increased estimate of the mecc. This is because TTV tends to increase with demand just as does the mean travel time.

There is thus a basis for including a measure of the scale of the distribution of travel time as a measure of TTV. Given estimates of the scheduling parameters and the shape of the travel time distribution it is possible to calculate the contribution of TTV to the GTC. It is not necessary to know the preferred arrival time of travellers since this does not appear in the GTC when the departure time is optimally chosen.

While the Fosgerau-Karlstrom result has some advantages for application, there are also some drawbacks. First, the value of TTV depends on the shape of the travel time distribution. It may therefore vary across different contexts. It still remains to gather sufficient empirical

---

<sup>12</sup> Previously, this had only been shown for some special cases [5, 24].

evidence to be able to judge whether this is a serious drawback in practice, given that there are many other uncertainties and approximations in play.

Second, the result does not apply to scheduled services. The issue is that users of scheduled services are, by definition, not able to choose their departure time freely. Therefore they are not able to choose the optimal departure time as required by the Fosgerau-Karlstrom model.

Third, the scale of the travel time distribution can be hard to compute in networks comprising many links. This is not an issue with the mean travel time, since the mean travel time may be computed at the link level and then summed over links to obtain a trip level mean travel time. The standard deviation is not additive in this way and so the GTC cannot just be computed at the link level and summed.

In a broader perspective, it is important that the specification of scheduling preferences is consistent with empirical evidence. The Vickrey/Small specification of scheduling preferences entails the prediction for an individual traveller that an isolated increase in travel time will cause a proportional change in departure time that leaves the arrival time unchanged. An isolated increase in the standard deviation of travel time would lead to earlier departure and earlier arrival on average. This may or may not be an adequate description of actual behaviour.

There is an alternative formulation of scheduling preferences that also leads to a tractable expression for the value of TTV. It is based on a less known paper by Vickrey [37] in which he defines scheduling preferences in terms of time-varying utility rates at the origin and at the destination of the trip. The traveller receives utility at some rate specific to the trip origin until he departs. When he arrives he begins receiving utility at a rate specific to the destination. The cost of the trip is an opportunity cost associated with the foregone utility at the origin or at the destination. When the utility rate at the origin is decreasing and the utility rate at the destination is increasing, then there is a time at which the individual would optimally travel from the origin to the destination. This view of scheduling preferences is attractive, since it treats the origin and the destination of the trip symmetrically. In general, it is hard to argue why timing at one trip-end should be more important than at the other as implied by Vickrey/Small scheduling preferences.

Using a simplified version of these scheduling preferences, Fosgerau [14] finds an expression for the value of travel time variability that does not depend on the shape of the travel time distribution. The related measure of travel time variability is the variance of travel time. This applies equally to travellers who can freely choose departure time and to travellers who use a scheduled service with fixed headway. Depending on parameters, travellers may be risk averse or risk seeking and the value of travel time may increase or decrease in the mean travel time.

This model has some advantages over the Fosgerau & Karlstrom model. First, the value of TTV does not depend on the shape of the travel time distribution. Second, the result applies



equally to scheduled services. Third, the variance of travel time is additive over links in a network, provided travel times on links are independent.<sup>13</sup>

Ultimately, the choice between formulations of scheduling preferences and the associated measures and value of travel time variability should not be based on convenience but on conformity with observable behaviour.

Randomness is a lack of information. And so information provision is a natural policy measure in the context of TTV. Consider a situation in which travel time is variable from day to day but perfect information about tomorrow's conditions is provided to travellers. Then every day they can choose the optimal departure time and the GTC in the Fosgerau-Karlstrom model reduces to  $GTC = c + \alpha E(T)$ , which omits the term relating to TTV.

The information does not have to be perfect in order to reduce the GTC. In general, it can be just some signal that contains some information about tomorrow's travel time, i.e. it must have some dependence with tomorrow's travel time.<sup>14</sup> This reduces the risk that travellers face. The value of this information may be assessed with the same models that are used to assess the cost associated with random travel time variability.

In using these results, it is important to keep in mind that no consideration has been given to equilibrium. The departure time choice of a single traveller is considered, taking the choices of all other travellers as given. The random distribution of travel time affects each traveller's choice of departure time. But there is also a causal relationship in the other direction whereby the combined departure time choices of travellers affect the distribution of travel time [3].

## MEASURING AND MODELLING SUPPLY

The basis for efficient congestion pricing is the marginal external cost of congestion. It involves, essentially, the VTT and the supply relationship. So it is clearly crucial for the design of congestion pricing schemes to have an adequate understanding of the supply-side. The description of supply relationships for road travel has traditionally been considered the domain of engineering or physics and economics has tended to ignore the complexities involved. This might have been reasonable in times when the main issues were to do with the design and capacity of road networks. When it comes to the design of road pricing schemes it becomes necessary to have a deeper understanding of the supply side. In particular,

---

<sup>13</sup> Travel times are not likely to be independent since delays on different links may have common causes. Still, additivity must be considered an improvement over no additivity.

<sup>14</sup> There are cases where equilibrium effects imply that imperfect information is not necessarily welfare improving [3].

it is essential to be able to estimate the effect on travel times of changing demand. Perhaps economics should get involved in this. Small & Verhoef [30] discuss congestion from this point of view.

In economic models, supply is generally taken as given in the form of a supply curve for a road or a simple network. In the case of the bottleneck model, it is simply the bottleneck capacity, which is a single number. In reality, congestion is a hugely complicated phenomenon. Research into the relationship between travel demand and travel time has been active for at least 75 years and we shall make no pretence of being able to summarise the state of the field. It is however important to point out some of the issues involved. There are still very many open questions that appear when we ask what is the consequence for mean travel time and variability of adding an extra traveller.

There are many causes and corresponding types of congestion. There is flow congestion which arises as traffic slows down due to increased density, independently of upstream or downstream links. Flow congestion is related to the microdynamics of traffic; it may arise due to small random fluctuations in flow and may be involved in the phenomenon of hypercongestion. The importance of flow congestion is however debated with some arguing that congestion is more likely to be related to bottlenecks such as intersections and merge lanes and to accidents that create temporary bottlenecks. Congestion involves spill-backs such that delays on a link may be due to upstream delays. A particularly clear instance of spill-backs is when a queue behind a bottleneck blocks upstream intersections for crossing traffic. Delay for the crossing traffic is then unrelated to that demand.

Consider now that the objective often is to price urban networks comprising thousands of links and nodes and even more combinations of origins and destinations. It is clearly a daunting task to try to describe such systems in detail. Here rescue may come from the existence of urban-scale volume-delay relationships [17], that allow the complexities of the network to be compressed into a single expression.

It is not in general sufficient to consider only the mean travel time, when travel time variability accounts for a significant share of GTC. It is necessary to be able also to predict the impact of TTV of changes in demand. In general there is no simple relationship whereby TTV can be expressed as a function of mean travel time. This is evident in Figure 4 where the standard deviation of travel time is rising at times when the mean travel time is decreasing.

The example of the loop in Figure 4 also indicates the dynamic nature of congestion whereby even a small event at one time and place may have significant effects later and elsewhere in the system. Such dynamic phenomena also create problems for the empirical measurement of speed-flow relationships.

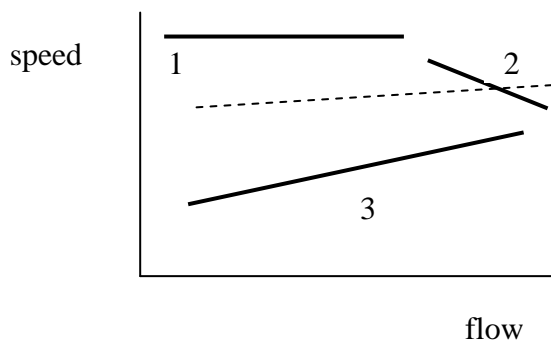


Figure 6. A stylised speed-flow relationship

The fundamental diagram of traffic flow incorporates a relationship between speed and traffic flow on a road link, depicted as a backward bending curve as shown in **Figure 6**. It depicts a situation with three regimes. First, a free flow regime in which speed is about constant with flows ranging from zero to some point. Second, a congested regime in which higher flows are associated with lower speeds. Third, a hypercongested regime in which speeds are lower than in the other regimes at the same flow levels and where higher flows are associated with higher speeds.

Consider a scatter of observations from a speed-flow curve. Consider a regression of speed against flow. If there are many observations from the hypercongested regime, then it can happen that an increasing mean relationship seems to be present. This would imply the perverse prediction that increasing flow would lead to an increase in speed. The problem is that there is causality in two directions. The causal effect of interest is the effect of flow on speed. There is however also a causal effect in the opposite direction whereby low speed creates a blockage which causes flow to be low. The problem of reverse causality is a classical econometric problem and a range of econometric techniques exist to tackle it. In the present context it is important to realise that the measurement and modelling of supply should be taken seriously and requires sophisticated methodology.

## SECOND-BEST ISSUES

The key idea of congestion pricing is that it reduces waste by alleviating the misalignment of private and social costs of travel that is caused by the congestion externality. The previous sections have pointed out how hard it is to establish the magnitude of this misalignment, making abstraction of the broader context in which pricing might be introduced and assuming that sophisticated instruments for charging are available. Second-best analysis asks the broad question of how the basic analysis is modified when these simplifications are

abandoned. The literature on the subject is vast and we make no attempt at providing an overview, referring the interested reader to a concise discussion in Small and Verhoef [30, pp. 137-147]. Instead, we limit ourselves to discussing some examples of second-best reasoning, and draw conclusions on how second-best analysis can help improve the practical implementation of congestion charging systems. Section 6.1 investigates the consequences of the fact that practical systems are approximations to the ideal charging system, and section 6.2 asks what are the consequences of the fact that congestion charging – even if potentially ideal – is implemented in an imperfect economy. Section 6.3 presents some guidelines for practical analysis.

### **Imperfect implementation**

If the second-best issues discussed below in section 6.2 are ignored, then an ideal congestion charging system is one that charges the marginal external congestion cost at each time and place in the road network. A glance at existing and planned systems shows this ideal is not reached: charges are levied on only a very limited number of network segments and those charges differ from the marginal external cost. If the first-best welfare gain from congestion charging is taken as a benchmark (which on the basis of the arguments of section 6.2 may be a disputable choice), this means lower welfare gains but not necessarily negative gains: congestion charging is still likely to be worthwhile, although the risk of counterproductive systems does exist and the loss of efficiency compared to ideal systems can be large.

The problem of which links to charge and what tolls to set when only some links in a network can be charged has a conceptual solution that is difficult to translate into a practicable one (e.g. Verhoef [34]). Simulations using detailed network models suggest that reasonably performant pricing schemes can be designed even with a small number of tolled links or cordons, e.g. by choosing high volume and high-speed links with poor substitutes [27]. While this is in line with common sense, no links may fit the bill perfectly, so that choices can be hard in practice. Furthermore, the question of how much to charge remains unresolved. Also, when the choice is where to place a cordon instead of what link to charge, deciding where to place one or more cordons and what charge to levy appears to be particularly challenging, with the results from simulation work sometimes differing from common sense judgment [32].

In sum it seems reasonable to conclude that imperfect implementation is unavoidable but that nevertheless good results can be obtained. However, systematic assessments of where and how much to charge can improve considerably on common sense judgment or at least help avoid big mistakes. Detailed analysis using traffic models is likely to have considerable payoffs.

### **Distortions in the wider economy**

The first-best analysis focuses on the congestion externality in transport. It implicitly assumes that there are no distortions (deviations from efficiency) in the rest of the economy, or at least none that should be taken into account when thinking about charging the external cost of congestion. In reality, the economy is rife with distortions that potentially do matter.

One example, which has attracted considerable research attention because of its quantitative importance, is that labour income is taxed. A tax on labour income is a distortion because it reduces the supply of labour, resulting in a less than efficient level of employment. How might this distortion affect prescriptions for a congestion charge?

Consider the simplest possible case where all road users are commuters and all commuters are road users: In order to work, it is necessary to commute. Say that the marginal external congestion cost per roundtrip (or, equivalently, per workday) is \$20, whereas the tax on labour income per workday is \$50. In this situation the labour tax and the congestion charge have the same effect since they both affect the same choice margin, whether to work or not. Since the labour tax more than covers external costs, no congestion charge is justified based on the argument of untaxed externalities. This polar case highlights the importance of interactions with other distortions.

Next, consider the same proportionality between labour supply and travel volumes, but introduce a second transport mode, “transit”. Assume that this mode is free of congestion and other external effects, and it is priced at marginal cost. A congestion charge for car commuting now is potentially useful, as it helps attain the socially desirable modal split. In fact, setting the congestion charge equal to the marginal external congestion cost guarantees that commuters face the marginal social cost of each mode, so the optimal modal split will result. At least, this is the case when the revenues from the charge are used to reduce labour tax revenues by the same amount, through a modification of the labour tax rate. For if revenues were used differently, the effective labour tax – and with it the labour market distortion – would increase. This example shows that revenue use is of crucial importance in evaluating the effects of congestion charging.

These examples are extreme: in reality there is no strict proportionality between labour supply and commuting, the labour market is not perfectly competitive, not all transport users are commuters, and commuters are a heterogeneous group. In general this means that the strong results of the examples will play out weaker in reality. Some more specific insights are as follows:

De Borger [11] investigates the consequences of replacing the assumption of a competitive labour market by one where unions and employers bargain about wage levels. This arguably is a closer approximation of prevailing labour market conditions in some countries or some segments of the labour market, and it is seen to matter for the impact of the interaction of labour market distortions and the congestion externality. As long as congestion tolls are not differentiated between those who are and are not working, it makes sense to raise the toll above marginal external congestion costs as this shifts part of the tax burden to non-workers. Given the large distortion from labour taxes at the margin, such a shift is efficiency-improving. However, when differentiation of tolls between workers and non-workers is feasible, the toll for workers is best set below marginal external congestion costs and that for non-workers exceeds it.

The importance of distinguishing between work and non-work trips is also highlighted by Van Dender [33], who extends the Parry and Bento [20] analysis by considering two trip types that are strict complements to labour supply and to leisure respectively, and by

assuming that transit and cars share a congested road. One result is that introducing a congestion toll can make sense even if the labour tax does not change, on the condition that the toll can be differentiated between both trip types. But if no differentiation is possible and the labour tax cannot be changed, then the scope for efficiency-improving tolls is quite limited (in a model where no possibilities for welfare-improving revenue use exists apart from reducing labour taxes). These analyses suggest that differentiation of congestion charges can be very useful not only because road users differ (as highlighted in previous sections), but because of efficiency costs of distortionary taxation. In particular, it makes sense to charge less for commuting trips.

Introducing heterogeneity into the analysis forces us to abandon a strict focus on efficiency. If individuals differ, we need to be explicit on how we care about the distribution of welfare among them. In second-best, the strict separation between efficiency and equity cannot be maintained because there are no lump-sum taxes suitable for supporting that separation. This tells us that the equity impacts of congestion charges and the use of revenues need to be considered.

However, one should not forget that the labour income tax is there for more than one reason; a notable motivation is a desire for redistribution. Kaplow [22] argues this point of view and suggests that cost-benefit tests should not be corrected with the marginal cost of funds. In the context of road pricing, Kaplow might similarly argue that income taxation should not be counted as distortionary.

There are several important market imperfections other than labour income taxes that interact with congestion charges. First, search unemployment occurs because it takes time for separated workers to match with new employment. The duration of search depends on the number of job openings within reach of the unemployed which in turn is affected by transport costs. Thus increasing transport costs can lead to increased duration of unemployment spells and a lower employment rate. This effect is likely to be significant [26]. Agglomeration effects constitute another relevant externality to take into account when thinking about congestion charges. There is substantial evidence that production in cities is subject to agglomeration economies, i.e. economies of scale that are external to firms. This is relevant to the analysis of congestion, as workers contributing to and suffering from congestion also contribute to the positive externality of agglomeration. If all workers contribute equally to congestion and agglomeration, this suggests congestion tolls should be reduced by the value of the agglomeration externality unless a separate instrument is available to stimulate agglomeration. But if workers differ in their values of time and their effects on agglomeration (as is the case according to [Graham , 200x](#)), tolls should differ among workers [19]. Of course, agglomeration effects are far more complex than suggested in these simple models and our understanding of them is limited [18]. In addition, policies to improve the benefits from agglomeration can be separate from congestion management policies. Nevertheless, the warning that congestion should not be considered in isolation from the productivity of a city's economy is valid.

## **Implication for implementation**

One response to the reality that several potentially important market imperfections interact with the congestion externality is to construct a model that encompasses the main imperfections (as judged by who builds the model) and derive a rule for the assessment of charges in this context. For example, Calthrop et al. [8] propose a rule for transport infrastructure investments, which could be modified for transport pricing, which includes many of the interactions discussed above. Their framework emphasizes the role of distortions, which comes at the cost of a strongly simplified representation of the transport markets. Fosgerau and Pilegaard [16] take the opposite route, showing how some general equilibrium interactions related to tax distortions can be integrated into cost-benefit analyses based on traffic models. This has the advantage of allowing a detailed model of the transport market (relying on traffic models that are often used in the practice of transport project appraisal), but the range of general equilibrium interactions is more narrow.

The use of such sophisticated rules is sometimes thought to be superior to the simple first-best rule on the grounds that the latter is easily shown to imply large mistakes in some cases, i.e. it is not a very good approximation. However, given our imperfect understanding of the broader context in a conceptual sense and even more in an empirical sense, it is not obvious that the approximation of the sophisticated rule is necessarily better. The fact that analyses of theoretical rules are typically illustrated by highly stylized numerical illustrations suggests that not too much faith should be put in the practical relevance of the numbers, but rather that they should be considered as rough indications of orders of magnitude associated with the mechanisms contained in the theory. This view is similar to the one expressed by Böhringer et al. [7], who discuss second-best analysis of climate change policies:

“The above results should be treated with caution. The numbers are neither accurate nor precise. They are ballpark estimates. What really matters are the insights. Climate policy need not cost a lot, but imperfect implementation implies excess costs. The excess costs are substantial relative to the costs of the first-best policy.”

In addition, even the most comprehensive models on general equilibrium interactions are (by definition) highly stylized representations of reality. They highlight distortions thought to be of particular importance (with judgment ideally based on evidence) while ignoring others, and they miss features that matter in the applied analysis of proposals for congestion charging. For example, few models contain a sufficiently detailed representation of the capacity and usage of multimodal transport systems that would allow a comparison of revenue use to reduce labour taxes or to improve the supply of public transport. Clearly, such comparison would be relevant to the design of charging proposals. Parry and Small [25] present a detailed analysis of rationales for subsidizing public transport. They do not focus on the interaction with broader distortions, but do suggest that their impact on the optimal subsidy is limited.

Our view, then, is that the models generate insights that ought to be part of debates on the implementation of congestion charging. Applications of simple models can help clarify the importance of second-best concerns, but we should not expect a full-fledged general equilibrium analysis to be carried out.

## CONCLUSION

Unregulated congestion entails an efficiency loss and a corresponding possibility for obtaining a welfare gain. This gain can be achieved through road pricing, decentralising the decision about who should travel when and where. In the first step of analysis, the toll should equal the value of the delay that a marginal car imposes on other travellers. The size of the total delay associated with the marginal car is determined from traffic models, ranging from simple supply curves to complex traffic models. The transformation to monetary value accomplished through the value of travel time (VTT), which can be measured in various ways.

Congestion arises because people tend to travel at the same times. With an even distribution of traffic over time, there would be no congestion. As the example of the bottleneck model shows, there are potentially large benefits that can be achieved if the trip timing aspect of demand is taken into account by varying tolls over peaks, inducing travellers to distribute departure times more evenly.

The simplifications involved in the textbook analysis of congestion pricing allow the central insights to be easily communicated. There are however a number of complications that must be taken into account when this theory is taken to practice. First, people are different. The stylised facts state that many people have low VTT but some have very large VTT. The mean VTT is larger than the median. There is much variation between people but also between seemingly identical individuals and even within the same individual in different contexts. Recognition of this heterogeneity will tend to lead to higher suggested tolls and will tend to reveal larger benefits from price differentiation between roads.

Travel times are inherently random. As congestion increases, travel times become not only longer but also increasingly variable and unpredictable. This travel time variability contributes significantly to travel costs. Taking travel time variability into account will generally lead to higher suggested tolls. Theory exists whereby the value of travel time variability can be expressed in terms of the standard deviation or the variance of travel time in a simple and readily applicable way.

Traffic systems are hugely complicated and the complexities of measuring and modelling supply should not be underestimated. It is difficult to establish the size of the delay associated with a marginal vehicle and even more difficult to establish the consequences for travel time variability. The simple descriptions of the supply side often employed in economic papers are not adequate for real world assessment of road pricing systems.

Constraints on charging instruments lead practical systems to fall short of ideal (in a first-best or second-best sense) congestion charging, and this implies lower benefits than would be obtained in that ideal. Charging can of course still produce net benefits (after subtraction of investment and operational costs). Careful assessment of charging systems becomes crucial however, because of the multitude of design options available and the large differences among them in terms of benefits produced. Such assessment should tackle the second-best



aspects explicitly, and not evaluate a system as if it were first-best. Changes in travel speeds are not a sufficient indicator of the benefits of charging.

Congestion charging should not be considered in isolation, as there are economic interactions that have large potential effects on the benefits of charging. As indicated, we are sceptical about the possibility of capturing these interactions in one elaborate model, but do think it valuable to discuss explicitly how any charging system might interact with other taxes and other externalities, with the aid of models and data, however sketchy and imperfect they are.

Tolling may be suggested as part of a policy package that also specifies how revenues are to be used. However, such earmarking of revenues runs the risk of being wasteful. The justifiability of a package that combines tolling with earmarking depends on the way revenues are to be used. An otherwise justifiable tolling scheme may become unjustifiable in combination with a use of revenues that is unjustifiable. It is therefore helpful to evaluate the parts of such policy packages in isolation in order to make clear the merits of each component.

One specific recommendation that emerges is that charges for commuting should be lower than for non-commuting. Since a trip purpose cannot be observed at the time charges are levied, a correction through the income tax system could be envisaged.

The second-best analysis suggests that in many cases the best tolls are lower than the marginal external cost of congestion because tolls match external costs imperfectly and they exacerbate pre-existing distortions. These, however, are partial results, depending to some extent on the interactions considered. If, for example, congestion tolls could be designed to fall mainly on non- or lower-taxed activities (“leisure”) and the revenues would be used to reduce highly distortionary taxes, then this would be an argument for raising them above the marginal external cost of congestion. What matters hence is to figure out what activities would be affected by the introduction of a charge.

In summary, it seems reasonable to use the basic analysis of congestion charging as a first approximation when considering its implementation, but to check whether important interactions with other market imperfections can be expected and revise the analysis when there are concerns about large indirect effects. Roy (2008) goes a step further, arguing for reliance on the simple rule and focus on removing second-best obstacles rather than taking them as given. We are less optimistic about the possibility of removing at least some of the obstacles. Nevertheless, our view and Roy’s are not mutually exclusive if charging mechanisms are seen as malleable through time.

Reference List

- [1] R.A.Arnott, A.de Palma, R.Lindsey, A structural model of peak-period congestion: A traffic bottleneck with elastic demand, *American Economic Review* 83 (1993) 161-179.
- [2] R.A.Arnott, A.de Palma, R.Lindsey, The Welfare Effects of Congestion Tolls with Heterogeneous Commuters, *Journal of Transport Economics and Policy* 28 (1994) 139-161.
- [3] R.A.Arnott, A.de Palma, R.Lindsey, Information and time-of-usage decisions in the bottleneck model with stochastic capacity and demand, *European Economic Review* 43 (1999) 525-548.
- [4] R.Arnott, M.Kraus, When are anonymous congestion charges consistent with marginal cost pricing?, *Journal of Public Economics* 67 (1998) 45-64.
- [5] J.Bates, J.Polak, P.Jones, A.Cook, The valuation of reliability for personal travel, *Transportation Research Part E: Logistics and Transportation Review* 37 (2001) 191-229.
- [6] G.S.Becker, A theory of the allocation of time, *Economic Journal* 75 (1965) 493-517.
- [7] C.Böhringer, T.F.Rutherford, R.S.J.Tol, The EU 20/20/2020 targets, ? (2009).
- [8] E.Calthrop, B.De Borger, S.Proost, Cost-Benefit analysis of transport investments in distorted economies, *Transportation Research Part B Forthcoming* (2009).
- [9] M.J.Cassidy, K.Jang, C.F.Daganzo, The smoothing effect of carpool lanes on freeway bottlenecks, *Transportation Research Part A Forthcoming* (2009).
- [10] B.De Borger, M.Fosgerau, The trade-off between money and time: a test of the theory of reference-dependent preferences, *Journal of Urban Economics* 64 (2008) 101-115.
- [11] B.De Borger, Commuting, congestion tolls and the structure of the labour market: optimal congestion pricing in a wage bargaining model, *Regional Science and Urban Economics* 39 (2009) 434-448.
- [12] M.Fosgerau, Investigating the distribution of the value of travel time savings, *Transportation Research Part B: Methodological* 40 (2006) 688-707.
- [13] M.Fosgerau, On the relation between the mean and variance of delay in dynamic queues with random capacity and demand, *Journal of Economic Dynamics and Control Forthcoming* (2009).
- [14] M.Fosgerau, The value of travel time variance, *Working Paper* (2009).
- [15] M.Fosgerau, A.Karlstrom, The value of reliability, *Transportation Research Part B: Methodological In press* (2009).

- [16] M.Fosgerau, N.Pilegaard, Cost-benefit rules for transport projects when labor supply is endogenous and taxes are distortionary, Munich Personal RePEc Archive 3902 (2007).
- [17] N.Geroliminis, C.F.Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological* 42 (2008) 759-770.
- [18] D.Graham, K.Van Dender, Estimating the agglomeration benefits of transport investments: some tests for stability, *JTRC Discussion Paper* 32 (2009).
- [19] D.Graham, K.Van Dender, Pricing congestion with heterogeneous agglomeration externalities and workers, Working Paper (2009).
- [20] Ian W.H Parry, Antonio Bento, Revenue Recycling and the Welfare Effects of Road Pricing, *Scandinavian Journal of Economics* 103 (2001) 645-671.
- [21] S.R.Jara-Diaz, *Transport Economic Theory*, Elsevier, Amsterdam 2007.
- [22] L.Kaplow, The Optimal Supply of Public Goods and the Distortionary Cost of Taxation, *National Tax Journal* 49 (1996) 513-533.
- [23] M.Menendez, C.F.Daganzo, Effects of HOV lanes on freeway bottlenecks, *Transportation Research Part B* 41 (2007) 809-822.
- [24] R.B.Noland, K.A.Small, Travel-Time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes, *Transportation Research Record* 1493 (1995) 150-158.
- [25] I.Parry, K.A.Small, Should urban transit subsidies be reduced?, *American Economic Review* 99 (2009) 700-724.
- [26] N.Pilegaard, M.Fosgerau, Cost-benefit analysis of a transport improvement in the case of search unemployment, *Journal of Transport Economics and Policy* 42 (2008) 23-42.
- [27] E.Safirova, K.Gillingham, I.Parry, P.Nelson, W.Harrington, D.Mason, Welfare and distributional effects of road pricing schemes for metropolitan Washington DC, in Georgina Santos (Ed.), *Road pricing: theory and evidence*, Elsevier JAI, Amsterdam, 2004, pp. 179-208.
- [28] K.Small, The scheduling of Consumer Activities: Work Trips, *American Economic Review* 72 (1982) 467-479.
- [29] K.A.Small, J.Yan, The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation, *Journal of Urban Economics* 49 (2001) 310-336.
- [30] K.A.Small, E.T.Verhoef, *Urban transportation economics*, Routledge, London and New York 2007.

- [31] K.A.Small, C.Winston, J.Yan, Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability, *Econometrica* 73 (2005) 1367-1382.
- [32] A.Sumalee, T.May, S.Shepherd, Comparison of judgmental and optimal road pricing cordons, *Transport Policy* 12 (2005) 384-390.
- [33] K.Van Dender, Transport taxes with multiple trip purposes, *Scandinavian Journal of Economics* 105 (2003) 295-310.
- [34] E.T.Verhoef, Second-best congestion pricing in general networks: heuristic algorithms for finding second-best optimal toll levels and toll points, *Transportation Research Part B* 36 (2002) 707-729.
- [35] E.T.Verhoef, K.A.Small, Product differentiation on roads: constrained congestion pricing with heterogeneous users, *Journal of Transport Economics and Policy* 38 (2004) 127-156.
- [36] W.S.Vickrey, Congestion theory and transport investment, *American Economic Review* 59 (1969) 251-261.
- [37] W.S.Vickrey, Pricing, metering, and efficiently using urban transportation facilities, *Highway Research Record* 476 (1973) 36-48.
- [38] A.de Palma, M.Fosgerau, Random queues and risk averse users, Working Paper (2009).



