Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

University of Regensburg Working Papers in Business,
Economics and Management Information Systems

Traffic Congestion and Accidents

Andrea Schrage*

9. November 2006

Nr. 419

JEL Classification: *R41, R48, D62, H23*

Key Words: *Transport externalities, congestion pricing, traffic accidents*

* Andrea Schrage is a research assistant at the Department of Economics, Faculty of Business, Economics and Management Information Systems at the University of Regensburg, 93040 Regensburg, Germany
Phone: +49-941-943-2701, E-mail: andrea.schrage@wiwi.uni-regensburg.de

Traffic Congestion and Accidents

Andrea Schrage

Department of Economics, University of Regensburg

93040 Regensburg, Germany

e-mail: andrea.schrage@wiwi.uni-regensburg.de

Tel.: +49-941-943 2701, Fax: +49-941-943 2734

Abstract

Obstructions caused by accidents can trigger or exacerbate traffic congestion. This paper derives the efficient traffic pattern for a rush hour with congestion and accidents and the corresponding road toll. Compared to the model without accidents, where the toll equals external costs imposed on drivers using the road at the same time, a new insight arises: An optimal toll also internalizes the expected increase in future congestion costs. Since accidents affect more drivers if traffic volumes are rising than when they are declining, the efficient charge depends upon the demand for road use during the rest of the peak period.

Key words: Transport externalities, congestion pricing, traffic accidents

JEL-Classification: R41, R48, D62, H23

9 November 2006

1 Introduction

Traffic congestion is an extremely annoying feature of road transport. It consumes substantial amounts of valuable time, creates difficulties for scheduling and on-time deliveries, and thus reduces the potential advantages of road transport. Congestion typically occurs at times of high travel demand or as a consequence of accidents or other non-recurring incidents that temporarily reduce a road's capacity. It is associated with external costs in the sense that an additional driver on a road forces everyone else using this same road at the same time to adapt to the higher traffic volume by lowering driving speeds, so other drivers need more time to cover a given distance. The opportunity cost of the additional travel time is external to the marginal driver. Since there are no market prices reflecting this cost, drivers ignore it and the equilibrium traffic volume is excessively high compared to the welfare maximizing one.

The standard economic solution to congestion, dating back to Pigou (1920), consists in levying a toll equal to marginal external congestion cost. [1] A central result from the economic literature on traffic congestion (for a concise presentation of the general idea see Hau, 1992) is that external congestion costs take on different values depending on travel demand and the roads' characteristics. In order to correctly reflect the different values of this externality, the toll

[1] Solutions in the spirit of Coase (1960) fail to produce efficient results because of information asymmetries and high costs of negotiations among drivers. Commercial road management, as suggested by Roth (1996), suffers from limited competition within transportation corridors, so that profit maximizing tolls exceed efficient toll levels. Therefore, privatization of roads should be subjected to public price controls, and the optimal price equals the Pigou-toll.

needs to be higher on heavily used roads (typically in densely populated areas) than on uncongested ones, and it has to rise during rush hours. Tolls that are differentiated in time and space can increase welfare by inducing drivers to spread their trips more evenly and by reducing total traffic volumes to an efficient amount. Other externalities arising from road transport – external costs of accidents, emissions of pollutants and greenhouse gases, noise, road wear and tear – can also be addressed more efficiently with differentiated road charges than with the current practice of taxing gasoline and vehicle ownership (Calthrop and Proost, 1998). Since congestion and accidents are the most important categories of external costs related to vehicle kilometers traveled (and not so much to vehicle technology and other factors that are hard to influence through road charges), these will be the focus of attention in the following.

Based on the pioneering works of Pigou (1920), Walters (1961), and Vickrey (1969), the theory of congestion pricing has been explored in a variety of models (for a survey see Lindsey and Verhoef, 2000). Typically, road capacity is assumed constant in the short term. Congestion arises from high levels of travel demand or demand that is strongly concentrated during short peak periods, the latter resulting in pronounced rush hours. A congestion toll, via changing relative prices, induces some drivers to alter their travel behavior, e.g. to switch to public transport or to off-peak travel times. Thus, congestion is lowered by reducing traffic flow in relation to a given road capacity. In the long term, road building increases capacity, which also alleviates congestion.

But even in the short term capacity changes if accidents and other incidents are taken into consideration. By temporarily reducing the amount of road space effectively available, accidents can be triggers of traffic congestion suffered by any given amount of subsequent traffic. Empirically, some 25–30 % of delays are estimated to be the consequence of traffic accidents (Skabardonis et al., 2002; Puls, 2004; FHWA, 2005; Kwon and Varaiya, 2005). Neglecting the impact of road crashes on later traffic flows therefore leads to errors in the calculation of congestion tolls. To get an impression of the size of the problem, consider the following back-of-the-envelope calculation: An accident typically blocks one road lane for 45–90 minutes, causing time losses of 1200–5000 vehicle-hours. Price this delay at the average value of travel time, say €15 per vehicle-hour, to calculate the congestion costs of an accident. If an additional vehicle increases the probability of an accident by $3 \cdot 10^{-7}$, the expected external costs of accident-induced congestion amount to €0.005–0.023 per vehicle kilometer (data from Grenzeback and Woodle, 1992; Schrage, 2005; Pöppel-Decker et al., 2003). This rough estimate of the externality neglects the cost increase suffered by drivers avoiding the congested road after the accident by switching to other routes or delaying their trip, so the optimal accident-related toll component would be somewhat higher.

Congestion caused by accidents and the characteristics of the Pigou-toll that accounts for this phenomenon have thus far received only limited attention in the transportation economics literature. Previous work has focused on how traffic volumes influence accident risk (Newbery, 1988; Vitaliano and Held, 1991; Jansson, 1994; Dickerson et al., 2000; Pöppel-Decker et al., 2003; Edlin and Karaca-Mandic, 2006) and accidents' severity (Shefer and Rietveld, 1997; Noland and Quddus, 2005), or how to design incentives for careful driving (Vickrey, 1968; Lindberg, 2001; Parry, 2004). Empirically there is no clear-cut answer to the question whether individual accident risk and accident severity change with traffic flow. Accident risk on interurban roads seems not to rise with traffic volume (Shefer and Rietveld, 1997; Dickerson et al., 2000; Pöppel-

Decker et al., 2003). For urban traffic, in contrast, the elasticity of risk with respect to flow is found to be strictly positive (Dickerson et al., 2000; Lindberg, 1999), so there is an external cost in this respect which should be added to a Pigou-toll for road use. More recently, defensive driving behavior has been recognized as an additional source of travel delays: Accident risk may not rise with traffic volume, but this comes at the cost of time losses (Peirson et al., 1998; Hensher, 2006).

The present paper turns the question of how traffic volume affects accident risk around and analyzes the impact which accidents have on congestion levels and how this affects optimal toll levels when drivers include the possibility of delays due to accidents in their expectations. I proceed by describing a model of traffic congestion that incorporates random accidents. Section 3 calculates the cost minimizing pattern of road usage, and section 4 derives the road toll that implements this pattern as an equilibrium.

2 A model of peak traffic congestion and accidents

The congestion model used in this paper is an extension of Henderson's (1974 and 1977) model of peak period congestion that includes random accidents. A fixed number of commuters N, one per car, drives to work on a given road. Commuting entails two kinds of private costs in addition to vehicle operating costs. One is the opportunity cost of time T spent driving from origin to destination, which is less productive or enjoyable than other possible uses of time. The second type of costs are schedule delay costs of arriving earlier or later than the official work start time. This specification was originally introduced by Vickrey (1969) in the context of the bottleneck model, which has been used to analyze a variety of congestion problems since (Arnott et al., 1998). The Henderson-model differs from this model class in the congestion technology used; instead of a queue waiting to enter a bottleneck, it captures congestion as travel delays which are monotonously increasing with traffic volume, which is consistent with empirical estimates of speed-flow-functions.

Commuters are homogeneous with respect to all relevant parameters such as trip origin and destination, time valuations, preferred arrival time and driving behavior. The generalized cost (i.e. operating costs, opportunity costs of time spent driving and schedule delay) per driver starting a trip at time t and arriving at work at time $t+T$ is

$$c = c(t, T) \tag{1}$$

Here I do not explicitly account for the expected private costs of an accident a driver might be involved in, which are assumed constant for ease of exposition, but $c(t, T)$ includes delays caused by previous accidents via travel time T.

The total trip time T of a person departing from home at time t depends upon the traffic flow $f(t)$ entering the road at the same time as the vehicle in question,[2] where f is measured in vehicle-kilometers per unit of time. With low road usage and no congestion, travel time takes on its minimum value T_0. At higher volumes, congestion sets in and T rises with traffic flow. Groups of drivers starting out on the road at the same time travel at equal speeds and do

[2] Models with this characteristic have been termed "no propagation" models, as opposed to "instantaneous propagation" models, where driving speeds depend upon the average vehicle density on the road, not the density in a driver's immediate surrounding (Lindsey and Verhoef, 2000).

not draw apart during their journey. In equilibrium, $\mathrm{d}(t + T(t))/\,\mathrm{d}t > 0$ has to hold, which can be shown to be the case if the opportunity cost of travel time is positive (Schrage, 2005). This inequality guarantees two things: First, drivers cannot arrive earlier by unilaterally postponing their departure time, so the traffic pattern is actually a Nash-equilibrium, and second, small and therefore fast groups of drivers are not fast enough to catch up and interfere with possibly slower traffic further down the road.

Drivers departing at different times only influence each other indirectly in case of an accident. Since the vehicles involved block part of the road's width, they reduce the available capacity $\kappa(t)$ of the road for some time. For any traffic volume, a reduction in capacity makes travel more time consuming. Travel time for drivers starting their trip at time t is thus determined by the convex function

$$T = T(f(t), \kappa(t)) \tag{2}$$

satisfying

$$T_f \geq 0\ ,\ T_\kappa \leq 0\ ,\ T_{f\kappa} \leq 0, \tag{3}$$

where subscripts indicate partial derivatives with respect to that variable. The sign of the second partial derivative with respect to κ and f, $T_{f\kappa}$, is intuitively explained by the observation that at low traffic levels, when driving speeds are already high, an increase in capacity does little to improve these further. At high volume, on the other hand, adding capacity can relieve congestion considerably and make trip times much shorter.

Accidents are modeled as a Poisson process $q(t)$ with an arrival rate $\rho(f)$

equal to the expected number of accidents per unit of time. As $q(t)$ counts the cumulative number of accidents, $\mathrm{d}q = 1$ with each additional accident that happens and $\mathrm{d}q = 0$ with no accident. The probability that there is no accident during a time interval of duration Δ is $(1 - \rho(f)\Delta)$. The chance of more than one accident is close to zero if Δ is short (Ross, 1983), so the probability that one accident occurs is approximately $\rho(f)\Delta$. This probability is independent of the time that has elapsed since any previous accident, but increasing in traffic volume. This assumption is reasonable since the expected number of vehicles involved in accidents equals individual accident risk $r(f)$ times traffic volume f. Suppose the average number of cars per accident is a constant x, then the expected number of accidents is $\rho(f) = \frac{r(f)f}{x}$, which increases with rising f for empirically plausible non-negative values of the elasticity of risk with respect to traffic flow.

Immediately after an accident, the road's capacity is lowered by a fraction σ due to the presence of the stalled vehicles on the road. σ is assumed constant for simplicity. A more realistic specification would model this term as a random variable (possibly correlated to flow) reflecting accidents of different severity, causing different degrees of obstruction, but this complication promises little additional insight. After the accident, police, ambulance or rescue service start clearing the site, and capacity gradually recovers at a rate of $g(\kappa) \geq 0$ per unit of time, up to the point where it reaches its maximum of $\bar{\kappa}$ again, with $g(\bar{\kappa}) = 0$. The change in capacity is summarized by the following stochastic differential equation:

$$\mathrm{d}\kappa(t) = g(\kappa(t))\,\mathrm{d}t - \sigma\kappa\,\mathrm{d}q \ . \tag{4}$$

It is composed additively of jumps in capacity whenever an accident occurs and the deterministic, continuous function describing its recovery.

Using (2), the average social cost from equation (1) can be re-written as

$$c = c(t, f(t), \kappa(t)) \ , \tag{5}$$

which is convex and increasing in f [3] and decreasing in κ and satisfies $c_{f\kappa} \leq 0$. The total cost incurred by all commuters using the same departure time is $f(t) \cdot c(t, f(t), \kappa(t))$. Integrating over the duration of the rush hour from time $\underline{t}$, when the very first person departs, to $\bar{t}$, when the last commuter has left for work, gives the total cost TC of rush hour traffic on this road:

$$TC = \int_{\underline{t}}^{\bar{t}} f(t) \cdot c(t, f(t), \kappa(t)) \, \mathrm{d}t \ . \tag{6}$$

This cost includes the opportunity costs of travel time and wrong arrival times and those accident costs deriving from increased congestion levels, which are part of the travel time costs. In the next section, I derive the pattern of road usage for the rush hour that minimizes the expected social cost.

3 Socially optimal pattern of road use

The efficient traffic volume for each departure time minimizes the expected social cost of having a total of N commuters use the road during the peak period:

[3] The sign of the partial derivative $c_f \geq 0$ implicitly uses the assumption that the opportunity cost of time early at the drivers' destination is less than the opportunity cost of travel time. This is a common assumption in the analysis of rush hour traffic and is empirically verified by Small (1982).

$$\min_{f(t),\underline{t}} \; E\left\{\int_{\underline{t}}^{\bar{t}} f(t)\cdot c(t, f(t), \kappa(t))\,\mathrm{d}t\right\} \tag{7}$$

$$\text{s.t.} \;\; \kappa(\underline{t}) = \bar{\kappa} \;\; , \tag{8}$$

$$\mathrm{d}\kappa(t) = g(\kappa(t))\,\mathrm{d}t - \sigma\kappa\,\mathrm{d}q \;, \tag{9}$$

$$\int_{\underline{t}}^{\bar{t}} f(t)\,\mathrm{d}t = N \;. \tag{10}$$

Depending on changes in capacity, $f(t)$ is constantly readjusted to fulfill the objective equation (7). At the beginning of the rush hour there is no obstacle on the road (boundary condition (8)), but traffic can cause accidents and reduce future capacity (state equation (9)). The road's capacity at time $\bar{t}$ does not constrain the problem, since no later driver is hindered by it anymore.

The period of road use is also determined optimally: $\underline{t}$ is the first time the cost minimizing $f(t)$ becomes strictly greater than zero; $\bar{t}$ is the time by which all drivers have left home (constraint (10)), so $f(t)$ is back to zero for $t \geq \bar{t}$ and no further costs arise. Since the optimal pattern of traffic flow adapts to possible accidents, $\bar{t}$ cannot be determined ex ante.

To handle restriction (10) as part of the cost minimization problem, it is replaced by state equation (11) counting the cumulative number of drivers $A(t)$ that has already set out on the trip up to time t:

$$\mathrm{d}A(t) = f(t)\,\mathrm{d}t \;\; , \tag{11}$$

$$A(\underline{t}) = 0 \;, \quad A(\bar{t}) = N \;. \tag{12}$$

Boundary condition (12) requires that the rush hour ends when the cumulative number of people having left for work equals total traffic demand N.

The efficient traffic pattern has to balance three effects: A marginal increase in traffic flow at time t increases the social cost of congestion at that instant. It also raises the probability of an accident which might lower future road capacity, thus increasing expected congestion costs for subsequent drivers. But at the same time, having an additional driver use the road at time t raises $A(t)$ and reduces the number of drivers left to use the road during the rest of the rush hour, which reduces congestion costs at later times. To solve for the pattern of traffic flow that weighs these effects in the best possible way, define the optimal value function

$$V(t, A(t), \kappa(t)) = \min_f E\left\{\int_t^{\bar{t}} f(x) \cdot c(x, f(x), \kappa(x))\,\mathrm{d}x\right\} \; . \tag{13}$$

For given initial conditions $A(t)$ and $\kappa(t)$ at time t, this is the lowest expected total cost at which the remaining $N - A(t)$ people can reach their destination if traffic flows are chosen optimally during the remainder of the rush hour.

By manipulating the optimal value function (see appendix A), the following optimality condition for this problem is obtained:

$$c(t) + f(t) \cdot c_f + \rho_f \cdot [V(t, A, \kappa(1-\sigma)) - V(t, A, \kappa)] = -V_A \; . \tag{14}$$

The optimal solution also satisfies state equations (9) and (11) and boundary conditions (8) and (12). Since the peak period's beginning is chosen optimally, the transversality conditions is

$$V_t(\underline{t}) = 0 \; , \tag{15}$$

and the transversality condition with respect to the value of $\kappa(\bar{t})$ is

$$V_\kappa(\bar{t}) = 0 \; . \tag{16}$$

Equation (14) implicitly defines the optimal traffic flow pattern t, $f^* = f^*(t, A(t), \kappa(t))$. The left hand side of the equation is the expected marginal social cost of an additional trip at time t. It consists of private travel cost c, the congestion externality imposed on other people departing along with the marginal driver, $f \cdot c_f$, and external costs to drivers departing later on (the change in probability of an accident ρ_f that might narrow road space from κ to $\kappa(1-\sigma)$, increasing the value of V). With f^*, the cost of marginally increasing the current traffic flow exactly offsets the future savings from having one (marginal) driver less use the road at later times (as given by the right hand side of the equation).

In a deterministic model of peak congestion with no accidents, the optimal traffic pattern equalizes marginal social congestion costs $c + fc_f$ during the rush hour. This is intuitively plausible because it leaves no margins for arbitrage by shifting drivers across different travel times. Here, this result is modified. The expected value of V_A can be shown to be constant in optimum (see appendix B):

$$E\left\{\frac{\mathrm{d}(-V_A)}{\mathrm{d}t}\right\} = \Big[(c + fc_f) + \rho_f(V(\kappa(1-\sigma)) - V(\kappa)) + V_A(\kappa)\Big]\frac{\partial f^*}{\partial A} \\ = 0 \,, \tag{17}$$

where the last equality follows from the first order condition for road usage. In connection with equation (14) this means that with the optimal traffic pattern, the *expected* marginal social congestion cost, including expected costs inflicted upon later drivers, is equalized over the rush hour's duration.

If accidents trigger congestion, static congestion costs $c + fc_f$ ought to change during the peak in order to compensate for expected changes in later congestion costs, $\rho_f(V(\kappa(1-\sigma)) - V(\kappa))$. This latter intertemporal component of marginal social cost is not constant, as neither of its factors is: ρ_f varies with f^*, and the effect of a change in capacity from κ to $\kappa(1-\sigma)$ on the value function depends upon how many drivers are affected by that change. A marginal change in capacity influences total cost as follows (differentiating (A.1) with respect to κ):

$$\begin{aligned} V_\kappa(t) = & f(t)\frac{\partial c}{\partial \kappa}\Delta + E\{V_{\kappa(t+\Delta)}(t+\Delta)\}\frac{\partial \kappa(t+\Delta)}{\partial \kappa(t)} + \\ & \left[c(t) + f(t)\,c_f + \rho_f[V(\kappa(1-\sigma)) - V(\kappa)] + V_A(\kappa)\right]\frac{\partial f^*}{\partial \kappa}\Delta \\ = & f(t)c_\kappa\Delta + E\{V_{\kappa(t+\Delta)}(t+\Delta)\}(1+g_\kappa\Delta) \end{aligned} \tag{18}$$

A decrease in capacity at time t, all other things equal, causes static congestion costs to increase by $(-fc_\kappa)$ during a short time interval Δ. This effect is larger for higher traffic volumes, because with $c_{f\kappa} \leq 0$ one can see that $\partial(fc_\kappa)/\partial f \leq 0$. The capacity reduction at time t also leads to a change in future capacity, both directly and via its effect on the rate of recovery $g(\kappa)$. This in turn influences static congestion costs at later times as long as capacity is affected - more strongly so at higher traffic flows. Further, the decrease in capacity tightens one of the restrictions of the optimization problem and requires an adjustment of the traffic pattern $(\partial f^*/\partial\kappa)$, but the marginal effect of this on total cost is zero (from the envelope-theorem).

Note that unlike what is derived in the deterministic case without accidents (Henderson, 1974, 1977; Chu, 1995), the marginal external costs are not determined by current road usage $f(t)$ only. Furthermore, even with the same traffic flow and equal capacity, the external costs are not equal at different times during the peak. Identical traffic volumes cause higher expected congestion costs when traffic flows are rising then when road usage is declining. This result relates to the necessity to assign the commuters to a sequence of departures: There has to be a first, second, third, ..., last driver. In case of an accident of the very first driver, for example, all other drivers have to adjust their travel speeds or departure times compared to the case that he has no accident. An accident of the last driver, in contrast, does not trigger such responses, since there are no subsequent drivers. But postponing the first driver's departure to a later time cannot eliminate this disparity, since it is not departure time but position in the sequence of drivers that is responsible for his high expected accident cost. There is no possibility for arbitrage in this respect because it is logically impossible to just skip the first driver and add him at some later position. The inequality between different drivers' expected external accident costs is a consequence of this, and at the optimum it is offset by an adjustment of the deterministic congestion costs $c + fc_f$ in the opposite direction.

4 The optimal road toll

The unregulated equilibrium traffic pattern results from the households' cost minimizing departure time decisions. Because of the presence of externalities, it does not coincide with the optimal pattern of road usage. In order to reproduce the latter, individuals' behavior can be influenced by levying a toll for road use that is variable over time to reflect changing levels of external costs. In the following, a rule for calculating the toll level $\tau(t)$ for departure times during the rush hour is derived.

Drivers choose their departure time such as to minimize the expected private costs. These consist of the generalized cost of commuting $c(t)$ and of payments for road use if a toll is levied. As in the previous section, accident costs other than congestion will be neglected for ease of exposition. When making their decision, drivers are assumed to be perfectly informed about current road capacity and toll levels, so they know at which cost they can travel instantly with any given traffic flow. But due to the random nature of accidents they can only form rational expectations about how these variables will change over the rest of the rush hour.

Risk-neutral drivers compare the current private cost of a trip, $c(t)+\tau(t)$, to the expected future cost and choose to drive at the cheaper time. For example, if a decrease in private cost is expected, some drivers who had considered to leave immediately will postpone their trip to benefit from lower costs. This lowers current travel costs and increases their expected value later on. A Nash-equilibrium evolves where a unilateral shift in departure time does not promise any cost savings, i. e.

$$c(t)+\tau(t)=E\Big[c(t+\Delta)+\tau(t+\Delta)\Big] \tag{19}$$

is satisfied at every instant that the road is actually being used, and the cost is higher outside the period of road use when schedule delay costs grow too high. With $\Delta \to 0$, equation (19) is equivalent to

$$E\left[\frac{\mathrm{d}c}{\mathrm{d}t}+\frac{\mathrm{d}\tau}{\mathrm{d}t}\right]=0\ . \tag{20}$$

If no toll is charged, traffic flow will constantly adjust such as to equalize the *expected* private cost of driving at different times, including the expected travel delays from possible accidents, but ignoring externalities. Actual costs in contrast can – and do – change. If, for example, there is no accident during a short time interval, then the actual capacity after that time is higher than what was expected. Ceteris paribus, this lowers travel costs compared to their expected value. Traffic flow adjusts, which in turn influences costs, until present and expected future costs are again equalized, given the actual state of information – and as this changes constantly, so does average cost.

An optimal toll induces drivers to choose the efficient traffic flow at every departure time. Comparing the derivative of equation (14) with respect to departure time t,

$$E\left[\frac{\mathrm{d}c}{\mathrm{d}t} + \frac{\mathrm{d}\{f\,c_f + \rho_f \cdot [V(t, A, \kappa(1-\sigma)) - V(t, A, \kappa)]\}}{\mathrm{d}t}\right] = 0\ , \tag{21}$$

to equation (20), it can be seen that the efficient charge for road usage is determined as follows:

$$\tau(t) = f(t)c_f + \rho_f\,[V(t, A, \kappa(1-\sigma)) - V(t, A, \kappa)] + Z \tag{22}$$

If drivers anticipate that the toll is set according to this rule, their individually cost-minimizing departure time decisions will decentralize the optimal pattern of road usage. The first term of the toll is a standard congestion toll charged to internalize the external costs at the time of use resulting from a marginally higher traffic flow. The second term internalizes the marginal increase in the probability of an accident's negative consequences for following traffic. Finally, since total road usage was assumed to be perfectly inelastic, the toll can be increased by an arbitrary constant term Z (resulting from integration of $\mathrm{d}\tau/\,\mathrm{d}t$). The expected variation of the toll is responsible for decentralizing the optimal traffic pattern, and the constant term Z brings about a lump sum transfer between the road's users and its operator without distorting the drivers departure time decision. An intuitively appealing value for Z is zero; in this case, drivers pay a toll exactly equal to their expected external congestion cost. If Z where set to some positive value instead, they would pay a lump-sum tax on top of the congestion toll, or receive a subsidy if $Z < 0$.

An important point to note about the toll rate is on which variables it depends. In contrast to deterministic congestion models, it is not sufficient to know the road's design capacity (which determines the shape of the average cost function $c(f)$) and measure traffic flows to calculate the efficient toll. Here, the effectively available capacity at a driver's departure time, $\kappa(t)$, and the expected change of traffic flow also need to be taken into consideration in order to determine his external effect upon other drivers, because the former changes c_f and the latter influences $V(\kappa(1-\sigma)) - V(\kappa)$.

5 Summary and policy implications

Traffic accidents are an empirically important trigger of congestion, and they exacerbate pre-existing congestion levels. Theoretical analysis of accidents and congestion in the present paper reveals some important aspects of how to deal efficiently with this problem. The main result is that the marginal social cost of congestion in a static sense, $c + fc_f$, should not be equalized across different times of road usage. Instead, the optimal static marginal cost of congestion tends to be lower early during the rush hour than at later times. This is because a higher traffic flow causes not only congestion at the time of usage but also higher expected congestion later on. Early during the rush hour, when more subsequent drivers are affected, this dynamic cost component is more important than later on, when road usage is declining.

Variable road pricing can enforce the optimal pattern of road usage as an equilibrium allocation. The road toll changes to reflect the changing value of the marginal external effects, which consists of three components. The first is a component familiar from deterministic congestion pricing models, which depends only upon the traffic volume at the time it is being levied. It is responsible for leveling the peak and elongating the period of road use in order to achieve a more even utilization of the road. The second toll component captures the expected consequences of an accident, and its value depends upon the number of people that are left to use the road over the remainder of the rush hour. Finally, the toll can be increased or lowered by a time-invariant term if total demand is perfectly inelastic, changing the distribution of welfare gains among the road's operator and its users in a lump-sum fashion. If total demand is not fixed, this constant toll component has to be chosen to ensure optimal total road usage.

In order to set the correct Pigou-toll for road use, it is important to distinguish between congestion as a pure traffic flow problem and additional congestion that results from traffic accidents. Since these add up to the actual level of congestion, Lindberg (1999) suggests that the "distinction is not so important when designing the charge, the label of the components will differ but the sum of charges will be the same." The present model shows that this is true for the toll's static congestion component only, which equals the external congestion cost at the time of driving, no matter whether it results from high traffic flow or from low capacity after an accident. But an efficient road toll also charges the additional congestion costs caused by an accident to the drivers responsible for that accident (or rather the expected value of that cost to every driver), before the extra congestion sets in. This cannot be accomplished by a static congestion toll.

Taking accident-related externalities other than subsequent congestion into account does not change the picture. The efficient toll for road usage would need to be amended to reflect the system and traffic volume externality (Lindberg, 1999). But since they are both static in nature (i.e., their value depends on current traffic flows only), there is no modification to the intertemporal toll component, only a strengthening of the general case for road pricing to incorporate external accident costs.

Current methods of charging for road use are obviously inadequate to reflect the marginal external costs of accidents and congestion. Automobile insurance, motor vehicle or gasoline taxes are not nearly variable enough by distance driven, road characteristics or traffic volumes. But differentiated road pricing has in recent years evolved from a theoretical idea into a very real possibility. Technologies for toll collection have improved to the point where toll rates can be changed every few minutes based on real time traffic counts. Experience with time-variable tolling in Singapore and the US shows that drivers respond to the incentive of high peak period tolls by shifting their departure to off-peak times, much as theory assumes. Using satellite navigation and mobile communication, it is also possible to charge for driving on complex road networks with no need for toll bridges or other roadside charging infrastructure on countless road sections. Public acceptance for road charging is generally low when first introduced, but improves when drivers get accustomed to it (Schade and Schlag, 2003) and start to appreciate the reduction in congestion and other advantages accomplished by using the toll revenues, e.g. for road improvements.

One obstacle for putting the toll derived in section 4 into practice is that its calculation relies on knowledge of the optimal value function or, as a proxy for $V(\kappa(1-\sigma)) - V(\kappa)$, its derivative with respect to road capacity. Both functions are not readily measurable. Therefore, real-world implementations would have to rely on numerical approximations to equation (22). Another practical drawback is that it is necessary to announce toll rates before they come into effect: Only if they know the cost of a trip in advance can drivers adapt their travel plans. For this reason, the charge for road usage would have to lag behind current traffic conditions, losing part of the theoretically possible efficiency gains.

The important qualitative conclusion with respect to the "accident-congestion-toll" is that the marginal external cost of a higher traffic flow not only affects drivers at the same time, but also those driving later on. The congestion toll derived from deterministic models without accidents is therefore merely a lower bound for the efficient road toll, and particularly during the early phase of the rush hour, when traffic levels are rising, a higher toll is recommended.

References

Arnott, R., A. de Palma and R. Lindsey (1998): Recent Developements in the Bottleneck Model, in Button, K. and E. Verhoef (eds.), Road Pricing, Traffic Congestion and the Environment. Issues of Efficiency and Social Feasibility, Cheltenham.

Calthrop, E. and S. Proost (1998): Road Transport Externalities, Environmental and Resource Economics 11, 335–348.

Chu, X. (1995): Endogenous Trip Scheduling: The Henderson Approach Reformulated and Compared with the Vickrey Approach, Journal of Urban Economics 37, 324–343.

Coase, R. H. (1960): The Problem of Social Cost, Journal of Law and Economics 3, 1–44.

Dickerson, A., J. Peirson and R. Vickerman (2000): Road Accidents and Traffic Flows: An Econometric Investigation, Economica 67, 101–121.

Edlin, Aaron S. and P. Karaca-Mandic (2006): The Accident Externality from Driving, Journal of Political Economy 114, 931–955.

FHWA (Federal Highway Administration) (2005): Congestion Mitigation, http://www.fhwa.dot.gov/congestion/congest2.htm (last visited 04/15/2005).

Grenzeback, L. R. and C. E. Woodle(1992): The True Costs of Highway Congestion, ITE Journal 62, 16–20.

Hau, T. D., 1992, Economic Fundamentals of Road Pricing. A Diagrammatic Analysis, The World Bank Policy Research Working Papers 1971.

Henderson, J.V. (1974): Road Congestion. A Reconsideration of Pricing Theory, Journal of Urban Economics 1, 346–365.

Henderson, J.V. (1977): Economic Theory and the Cities, New York.

Hensher, David A. (2006): Integrating Accident and Travel Delay Externalities in an Urban Speed Reduction Context, Transport Reviews 26, 521–534.

Jansson, J. O. (1994): Accident externality charges, Journal of Transport Economics and Policy 38, 31–43.

Kwon, J. and P. Varaiya (2005): The congestion pie: Delay from collisions, potential ramp metering gain, and excess demand, Prepared for Transportation Research Board 84th Annual Meeting, Washington.

Lindberg, G. (1999): Calculating Transport Accident Costs. Final Report Of The Expert Advisors To The High Level Group On Infrastructure Charging, http://europa.eu.int/comm/transport/infr-charging/library/crash-cost.pdf (last visited 28/01/2006).

Lindberg, G. (2001): Traffic Insurance and Accident Externality Charges, Journal of Transport Economics and Policy 35, 399–416.

Lindsey, R. and E. Verhoef (2000): Congestion Modeling, in: D. A. Hensher and K. J. Button, eds., Handbook of Transport Modelling, Amsterdam, 377–397.

Newbery, D. M. (1988): Road User Charges in Britain, The Economic Journal 98, 161–176.

Noland, R. B. and M. A. Quddus (2005): Congestion and safety: A spatial analysis of London, Transportation Research Part A 39, 737–754.

Parry, I. W. H. (2004): Comparing alternative policies to reduce traffic accidents, Journal of Urban Economics 56, 346–368.

Peirson, J., I. Skinner and R. Vickerman (1998): The Microeconomic Analysis of the External Costs of Road Accidents, Economica 65, 429–440.

Pigou, A. C. (1920): The Economics of Welfare, London.

Pöppel-Decker, M., A. Schepers and I. Komann (2003): Grundlagen streckenbezogener Unfallanalysen auf Bundesautobahnen, Berichte der Bundesanstalt für Straßenwesen M 153, Bergisch Gladbach.

Puls, T. (2004): Verkehrsinfrastruktur in Deutschland. Bestand – Methoden – Finanzierung, IW-Umwelt-Service Themen 2/2003, Köln.

Ross, S. M.(1983): Stochastic Processes, New York et al..

Roth, G. (1996): Roads in a Market Economy, Avebury, Aldershot.

Schade, J., and B. Schlag (2003): Acceptability of urban transport pricing strategies, Transportation Research Part F: Traffic Psychology and Behaviour 6, 45–61.

Schrage, A. (2005): Straßenmaut und Verkehrsstaus, Tübingen.

Shefer, D. and P. Rietveld (1997): Congestion and Safety on Highways: Towards an Analytical Model, Urban Studies 34, 679–692.

Skabardonis, A., P. P. Varaiya and K. F. Petty (2002): Measuring recurrent and non-recurrent traffic congestion, Presentation at the 82nd Transportation Research Board Meeting 2003, Washington.

Small, K. A. (1982): The Scheduling of Consumer Activities: Work Trips, The American Economic Review 72, 467–479.

Vickrey, W. (1968): Automobile Accidents, Tort Law, Exernalities, and Insurance: An Economist's Critique, Law and Contemporary Problems 33, 464–487.

Vickrey, W. (1969): Congestion Theory and Transport Investment, The American Economic Review 59, 251–260.

Vitaliano, D. F., and J. Held (1991): Road accident external effects: An empirical assessment, Applied Economics 23, 373–378.

Walters, A. A. (1961): The Theory and Measurement of Private and Social Cost of Highway Congestion, Econometrica 29, 676–699.

A Derivation of (14)

The value-function in (13) is decomposed into the deterministic social cost of travel over a short time interval Δ, for which $f(t)$ is approximately constant and $\kappa(t)$ is known with certainty, and the expected cost for the rest of the rush hour. This yields the following Bellman-equation:

$$V(t, A(t), \kappa(t)) = \min_f [f(t)c(t)\Delta + E\{V(t+\Delta, A(t+\Delta), \kappa(t+\Delta))\}] \quad (A.1)$$

The cumulative number of departures at time $t+\Delta$ is determined via equation (11): $A(t+\Delta) = A(t) + f(t)\,\Delta$. The change in capacity is stochastic. With Δ small enough to assume $g(\kappa)$ constant, $\kappa(t+\Delta)$ can be approximated as

$$\kappa(t+\Delta) = \kappa(t) + \mathrm{d}\kappa(z)\Delta = \kappa(t) + g(\kappa(t))\Delta - \sigma\kappa dq\Delta\ . \quad (A.2)$$

Uncertainty about $V(t+\Delta)$ arises only from future capacity, therefore the expected value can be written as a probability-weighted average of the value function in case the next few drivers cause no accident or do cause one:

$$\begin{aligned} E\{V(t+\Delta)\} = (1-\rho(f)\,\Delta)\cdot V(t+\Delta,\, A+f\,\Delta,\, \kappa+g\,\Delta) \quad + \\ \rho(f)\,\Delta \ \cdot\ V(t+\Delta,\, A+f\,\Delta,\, \kappa(1-\sigma)+g\,\Delta)\ , \end{aligned} \quad (A.3)$$

where (A.2) was used for $\kappa(t+\Delta)$. This is plugged into (A.1) and $V(t)$ subtracted from both sides of the equation. Dividing by Δ and taking limits for $\Delta \to 0$, the following is obtained:

$$0 = \min_f [fc + V_t + V_A f + V_\kappa g + \rho(f)[V(t, A, \kappa(1-\sigma)) - V(t, A, \kappa)]]\ . \quad (A.4)$$

The optimal traffic flow at time t minimizes the sum of instantaneous cost $f(t)\cdot c(t)$ and the expected change in V during an infinitesimal amount of time. By differentiating the expression in brackets with respect to f, (14) is derived as the first order condition.

B Derivation of (17)

Inserting the optimal traffic volume $f^*(t, A(t), \kappa(t))$ as implicitly defined by (14), equation (A.1) becomes an identity.

$$V(t, A(t), \kappa(t)) \equiv f^*(t)c(t, f^*(t), \kappa(t))\Delta + E\left\{V(t+\Delta, A(t+\Delta), \kappa(t+\Delta))\right\} \quad \text{(B.1)}$$

Differentiate with respect to $A(t)$ to obtain

$$\begin{aligned} V_{A(t)}(t) = {} & (c + f c_f)\mid_{f^*} \frac{\partial f^*}{\partial A(t)}\Delta + \\ & [-\rho_f \Delta V(\kappa + g\Delta) + \rho_f \Delta V(\kappa(1-\sigma) + g\Delta)] \frac{\partial f^*}{\partial A(t)} + \\ & \Big[(1-\rho\Delta)V_{A(t+\Delta)}(\kappa + g\Delta) + \rho\Delta V_{A(t+\Delta)}(\kappa(1-\sigma) + g\Delta)\Big] \underbrace{\left(1 + \frac{\partial f^*}{\partial A(t)}\Delta\right)}_{\partial A(t+\Delta)/\partial A(t)} . \end{aligned} \quad \text{(B.2)}$$

Using

$$E\left\{\frac{\mathrm{d}\, V_A}{\mathrm{d}t}\Delta\right\} = E\{V_{A(t+\Delta)}(t+\Delta)\} - V_{A(t)}(t) \ , \quad \text{(B.3)}$$

equation (B.2) can be rearranged to find

$$\begin{aligned} E\left\{\frac{\mathrm{d}(-V_A)}{\mathrm{d}\, t}\Delta\right\} = {} & \Big[(c + f c_f)\mid_{f^*} \Delta + \\ & -\rho_f \Delta V(\kappa - g\Delta) + \rho_f \Delta V(\kappa(1-\sigma) + g\Delta) + \\ & (1-\rho\Delta)V_{A(t+\Delta)}(\kappa - g\Delta)\Delta + \rho\Delta V_{A(t+\Delta)}(\kappa(1-\sigma) + g\Delta)\Delta \Big] \frac{\partial f^*}{\partial A(t)} . \end{aligned} \quad \text{(B.4)}$$

Dividing by Δ and finally taking limits for $\Delta \to 0$ results in equation (17).