

Verhoef, Erik T.; Rouwendal, Jan

**Working Paper**

## A Structural Model of Traffic Congestion

Tinbergen Institute Discussion Paper, No. 01-026/3

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* Verhoef, Erik T.; Rouwendal, Jan (2001) : A Structural Model of Traffic Congestion, Tinbergen Institute Discussion Paper, No. 01-026/3, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/85938>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



TI 2001-026/3

Tinbergen Institute Discussion Paper

# A Structural Model of Traffic Congestion

*Erik T. Verhoef*  
*Jan Rouwendal*

*Department of Spatial Economics, Faculty of Economics, Vrije Universiteit Amsterdam, and  
Tinbergen Institute.*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50

3062 PA Amsterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to [wdriessen@few.eur.nl](mailto:wdriessen@few.eur.nl).

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

# A STRUCTURAL MODEL OF TRAFFIC CONGESTION

## Endogenizing speed choice, traffic safety and time losses<sup>#</sup>

Erik T. Verhoef<sup>\*</sup> and Jan Rouwendal<sup>\*\*</sup>  
Department of Spatial Economics  
Free University Amsterdam  
De Boelelaan 1105  
1081 HV Amsterdam  
The Netherlands  
Phone: +31-20-4446094  
Fax: +31-20-4446004  
Email: everhoef@econ.vu.nl

This version: 17/10/03

**Key words:** Speed, Traffic congestion, Safety, Road pricing

**JEL codes:** R41, R48, D62

### **Abstract**

*Conventional economic models of traffic congestion assume that the relation between traffic flow and speed is a technical one. This paper develops a behavioural model of traffic congestion, in which drivers optimize their speeds by trading off time costs, expected accident costs and fuel costs. Since the presence of other drivers affects the latter two cost components and hence the Nash equilibrium speed, a 'behavioural' speed-flow relationship results for which external congestion costs include expected accident costs and fuel costs, in addition to the time costs considered in the conventional model. It is demonstrated that the latter in fact even cancel in the calculation of optimal congestion tolls. The overall welfare optimum in our model is found to be off the speed-flow function, and off the average and marginal cost functions derived from it in the conventional approach. This full optimum requires tolls to be either accompanied by speed policies, or to be set as a function of speed. Using an empirically calibrated numerical simulation model, we illustrate these qualitative findings, and attempt to assess their potential empirical relevance.*

<sup>#</sup> The authors thank Toon van der Hoorn and Piet Rietveld for stimulating discussions.

<sup>\*</sup> Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam. The research of Erik Verhoef has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

<sup>\*\*</sup> Jan Rouwendal is also at Wageningen Agricultural University, Department of Economics.



## 1. Introduction

Road traffic congestion is among the most widely studied externalities in the economic literature (for overviews, see Small, 1992; and Lindsey and Verhoef, 2000). This interest is motivated by the importance and immediate visibility of traffic congestion in most cities, and the growing interest of authorities in applying economic principles – *i.e.* road pricing – for its regulation (*e.g.* Small and Gómez-Ibáñez, 1998). Traffic congestion in reality is a complex phenomenon. It is the result of – and in turn affects – the dynamic behaviour of and interactions between many road users. For reasons of tractability, most economic models rely on greatly simplified representations of this complex reality. This is especially true for the conventional static economic model of traffic congestion, which uses cost curves derived directly from the speed-flow relation. Although well suited to illustrate the economic principles of traffic congestion and congestion pricing, it may miss out on other important aspects that may strongly affect the nature of optimal congestion tolls, and the welfare gains that can be achieved using these. This was illustrated convincingly for instance by Vickrey's (1969) dynamic model of traffic congestion, extended in various directions by Arnott, De Palma and Lindsey (1994, 1998) – the so-called bottleneck model.

Besides assuming stationary traffic conditions, however, the conventional model makes more potentially far-reaching simplifying assumptions. The one of concern in the present paper is that the underlying speed-flow function is a purely technical relation, that can be taken as given when evaluating the welfare properties of free-market and optimal equilibria. We develop a model for congested highway traffic that does not treat the speed-flow relation as a technical law, but instead as resulting from road users' optimizing behaviour. Specifically, we assume that road users choose an optimal speed given the situation on the road, so as to minimize generalized travel costs. This involves a trade-off between time costs and expected accident costs, which both vary with the speed chosen. Since accident risks, and individual drivers' responses to these through speed choice, are determined by the presence and behaviour of other drivers, the model endogenously generates traffic congestion as the result of individuals' optimizing behaviour. It thus integrates three aspects of travel behaviour that are often studied separately: speed choice, safety and congestion (moreover, also fuel costs are included in the numerical version of the model). Our approach thus incorporates the obvious but often ignored observation that people slow down in traffic congestion for a good reason, namely that otherwise accident risks would become excessive, and tries to assess the implications for congestion policies from an economic perspective.

Our analysis builds upon that of Rotemberg (1985), in which road users choose optimal gaps between their own and their leaders' cars, taking the leaders' speeds as given. In our model, road users instead choose a speed, taking the traffic density as given. Our analysis furthermore differs from Rotemberg's (1985) in that it derives the optimal 'flat' (speed-independent) toll and discusses the difference with congestion tolls as derived in conventional models, considers the first-best optimum of a joint toll and speed policy explicitly (Rotemberg, 1985, studies the analytics of optimal speed regulation alone), discusses the

relation between the behavioural model and observed speed-flow functions and shows how the latter would follow from the former, derives explicitly that the first-best optimum is off the conventional speed-flow function (which is implicit in Rotemberg's (1985) analysis), and presents an empirically calibrated numerical simulation model.

## **2. An analytic model**

### *2.1. Introductory remarks*

The conventional static economic model of traffic congestion is rooted in the works of Pigou (1920) and Knight (1924). The relevant average and marginal cost curves are derived from a 'technical' relationship between traffic flow and speed, analogous to the technical cost curves derived from production functions in the microeconomic theory of the firm. The speed-flow relationship is thus regarded as a reflection of what may be called 'traffic or congestion technology' (Small, 1992). This approach has the advantage that the speed-flow curve can be measured, and hence provides an empirical starting point for the economic analysis of congestion (see for instance Keeler and Small, 1977, or Boardman and Lave, 1977, for examples of empirical examinations of the speed flow relationship). Even though this approach is common, it may be questioned whether it is entirely adequate for the purpose of an economic analysis of congestion externalities. If the speed-flow relationship is not a technical relationship, but determined by driver behaviour instead, it may not be invariant to changes in the environment in which the drivers determine their behaviour. We develop a model to investigate whether this is the case, and if so, which economic policy and welfare implications arise from a 'behavioural' instead of a 'technical' speed-flow relation.

The following assumptions are made. We consider traffic on a single, uniform unidirectional road. Road users are homogeneous with respect to all generalized cost elements of a trip. Apart from enhancing tractability, this homogeneity assumption has the conceptual advantage that the results we obtain, including those relating to speed regulation, are valid also when road users have the same (desired and actual) speeds – the case for which these results are in fact more surprising than when heterogeneity in speeds would exist. Road users may differ with respect to their willingness to pay for making a trip, so that the aggregate (inverse) demand function is generally not perfectly inelastic. A stable demand function applies, meaning that with a constant equilibrium level of generalized costs, trips will be made at a constant rate. Only equilibria with stationary traffic conditions will be considered, in which speed, density and flow are constant over time and along the road. The choice for a static model is primarily motivated by its tractability, but also by our desire to compare the results to those of the conventional static model of traffic congestion.

If the (endogenously derived) generalized average cost function exhibits a backward-bending section, the demand function is assumed to be such that there exists an intersection with the 'normally congested' upward-sloping segment – as opposed to the backward-bending, 'hypercongested' segment (see Section 3 below; and also Verhoef, 1999, 2000). Based on the argumentation provided in these same two papers, we will furthermore ignore the possibility of ending up in a hypercongested equilibrium on our homogeneous road. As

we will not derive the dynamic instability of these hypercongested equilibria underlying this dismissal explicitly in this paper, we postulate it as an assumption here. The motivation is that we wish to concentrate on the insights that our model yields for the ‘normally’ congested segment of the cost function, the nature and analytics of which has been under economic debate far less than the hypercongested range (see for instance McDonald *et al.*, 1999).

Road users can make two decisions. The first is whether or not to use the road, which will be done when the equilibrium generalized cost of travel (including a toll) does not exceed the individual’s benefits from making a trip. Secondly, when using the road, an individual ( $i$ ) can choose at which speed ( $s^i$ ) to drive. Individual road users are atomistic, and take aggregate variables as given. These include the speeds chosen by all other drivers ( $S$  will denote the average speed of all drivers, with a zero variance in a symmetric Nash equilibrium), the flow of traffic ( $F$ ), and traffic density ( $k$ ). For density, this means that it is in fact assumed that frictionless overtaking is in principle possible, although it will of course not occur in a symmetric Nash equilibrium with equal speeds. Only with overtaking possible, a driver exhibiting Nash behaviour does not believe she could create a structurally lasting larger gap to the car in front (the leader in car-following terminology) by slowing down, and that she could thus affect (local) density. Starting in a stationary state, a driver  $i$  that slows down and chooses a speed  $s^i < S^{-1}$  (with  $S^{-1}$  denoting the equilibrium speed of all other drivers) expects she will be overtaken at a constant rate of  $k \cdot (S^{-1} - s^i)$  vehicles per unit of time, and therefore to experience a constant average local density of  $k$ , as well as time-averaged gaps to both the leader and the follower equal to  $1/k$ . Therefore, the individual does not believe she can affect  $k$  by changing speed (nor in any other way).<sup>1</sup>

Although we endogenize the trade off between travel times and expected accident costs in the determination of equilibrium speeds, we ignore that in a symmetric Nash equilibrium with equal speeds on a unidirectional road, accidents are strictly speaking impossible. The speeds we use should therefore in fact be interpreted as ‘average speeds’ (*i.e.*, over an individual’s entire trip). The ‘human factor’ that causes the actual fluctuations in this average speed during the trip, which in turn – with equal average speeds for all road users – increases the probability of a collision above the zero level, is not modelled explicitly.

---

<sup>1</sup> The assumption of road users taking density as given may seem inconsistent with car-following theory (as applied by, for instance, Verhoef, 2000), where drivers choose both a speed and a distance to the leader. We emphasize that our present model concerns stationary state equilibria with constant speeds only, whereas car-following theory is relevant primarily for the case where speeds vary during a trip. Note in particular that the car-following model of Verhoef (2000) predicts stationary states in which speeds are constant for the entire (or most of the) trip, with equal spacing between vehicles. Our model would concern constant-speed regimes of such a car-following model, and the assumption would mean that in such a regime, an individual user does not believe she can (and indeed could not) create a lasting lower density in his direct vicinity than that applies on average. Note also the difference between our model and Rotemberg’s (1985), in which each driver is assumed to choose a (constant) gap between him and his leader, taking his leader’s speed as given. As will become clear, our qualitative policy implications are nevertheless consistent with Rotemberg’s (1985).

## 2.2. Generalized costs and speed choice in a symmetric Nash equilibrium

To determine equilibrium speed choice, we now turn to the generalized costs of travel. We postpone the consideration of fuel costs to the empirical model of Section 3, and consider the case where an individual  $i$ 's total travel costs for a trip,  $c^i$ , consists of the sum of time costs  $c_{time}^i$  and expected accident costs  $c_{acc}^i$ . Because of our interest in stationary state traffic only, we will not develop a full-fledged dynamic model that could for instance describe drivers' changing speeds during their trip. Instead, we consider speed choice when drivers are forced to have a constant speed during their trip, which they can, however, freely choose based on perfect information on the traffic conditions that will be encountered.

The time costs depend exclusively on the speed chosen by the individual,  $s^i$ . The expected accident costs depend on more factors. For a given density of vehicles, both the probability of an accident and its severity can be expected to depend on one's own speed  $s^i$  and on the average speed of all other drivers on the road,  $S^{-i}$ . Furthermore, before having established that the Nash equilibrium will be symmetric, it is reasonable to assume that also the variance of all other drivers' speeds,  $\text{var}(S^{-i})$ , would matter when speed differences increase the probability of an accident. Finally, given the prevailing speeds, the probability of an accident is likely to depend on the average density of vehicles on the road,  $k$ , as  $k$  is a measure for both the number of potential 'partners' for a collision, and for the number of cars over which a driver should divide attention (we ignore the effects of any possible dispersion of  $k$  along the road upon travel costs). Note that the expected accident cost need not be zero if  $k=0$ : also collisions with fixed objects alongside the road may be possible. We thus obtain:

$$c^i = c_{time}^i(s^i) + c_{acc}^i(s^i; S^{-i}, \text{var}(S^{-i}), k) \quad (1)$$

We assume that  $c_{time}^i$  and  $c_{acc}^i$  are twice differentiable in all arguments. We will now make a number of assumptions stipulating how  $c^i$  depends on its various arguments.

We assume that travel time cost  $c_{time}^i$  is a decreasing convex function of  $s^i$ , and the associated curve gets infinitely steep when  $s^i$  approaches zero and becomes flat when  $s^i$  gets high:<sup>2</sup>

$$\frac{\partial c_{time}^i}{\partial s^i} < 0 \quad (2a)$$

$$\frac{\partial^2 c_{time}^i}{\partial s^{i2}} > 0 \quad (2b)$$

$$\left. \frac{\partial c_{time}^i}{\partial s^i} \right|_{s^i \downarrow 0} = -\infty \quad (2c)$$

$$\left. \frac{\partial c_{time}^i}{\partial s^i} \right|_{s^i \rightarrow \infty} = 0 \quad (2d)$$

---

<sup>2</sup> For convenience we ignore that a physical upper limit on  $s^i$  would theoretically be given by the speed of light.

It is assumed that the expected accident cost  $c_{acc}^i$  is convex in  $s^i$ , and that there always exists a range of  $s^i$  for which  $c_{acc}^i$  increases in  $s^i$  (the latter assumption is plausible for all  $s^i$  when  $k=0$ , and otherwise only when driver  $i$  does not drive too slow compared to  $S^{-i}$ , in which case speeding up might decrease expected accident costs by reducing speed differences):

$$\exists s^i : \frac{\partial c_{acc}^i}{\partial s^i} > 0 \quad (2e)$$

$$\frac{\partial^2 c_{acc}^i}{\partial s^{i^2}} > 0 \quad (2f)$$

It is now straightforward to characterize a driver's speed choice. Minimization of (1) with respect to  $s^i$  implies that, given the levels of the aggregate variables, the marginal impact of a speed adjustment on time cost should be balanced by that on accident cost:

$$\frac{\partial c^i}{\partial s^i} = \frac{\partial c_{time}^i}{\partial s^i} + \frac{\partial c_{acc}^i}{\partial s^i} = 0 \quad (3)$$

which requires that  $\partial c_{acc}^i / \partial s^i > 0$  at the optimum speed. Assumptions (2a)-(2f) imply that there exists a unique positive solution to (3) for given values of the aggregate variables when  $\partial c_{acc}^i / \partial s^i < \infty$  for  $s^i = 0$ , as we will assume. This means that we discard zero-speed equilibria. All densities  $k$  to be considered in the sequel are therefore below a certain possible jam density  $k^{jam}$  that would possibly make a zero speed, and infinite travel time costs, optimal.

The next question involves existence and uniqueness of a Nash equilibrium in the model. The first thing to observe is that, because drivers are identical and atomistic, and hence face the same values of the aggregate arguments in their cost function, any Nash equilibrium must be symmetric: all drivers choose the same speed. This means that we can ignore situations in which  $\text{var}(S^{-i})$  is positive when looking for such equilibria. We therefore ignore the term  $\text{var}(S^{-i})$  in what follows, and replace equation (1) by:

$$c^i = c_{time}^i(s^i) + c_{acc}^i(s^i; S^{-i}, k) \quad (4)$$

To prove existence and uniqueness of a Nash equilibrium, we introduce two additional assumptions that concern the effect of a change in the driver's own speed on accident costs in situations when the speed difference with all other drivers remains constant. A change in the driver's own speed implies a change in the speed difference with all others. The effect of the change in speed difference on expected accident cost may be positive (if the speed difference increases) or negative (if it decreases). However, it seems reasonable to assume that if we remove this effect (by keeping the speed difference constant), accident cost will still be an increasing and convex function of the driver's own speed. This can be stated formally as follows:<sup>3</sup>

---

<sup>3</sup> If the changes in  $s^i$  and  $S^{-i}$  are equal:  $ds^i = dS^{-i} = dS$ . Taking the total differential of accident cost and imposing the condition that  $dc_{acc}/dS$  must be positive gives (2g). Next, (2h) follows from assuming that  $dc_{acc}/dS$  increases in the 'joint' speed.

$$\frac{\partial c_{acc}^i}{\partial s^i} + \frac{\partial c_{acc}^i}{\partial S^{-i}} > 0 \quad (2g)$$

$$\frac{\partial}{\partial s^i} \left( \frac{\partial c_{acc}^i}{\partial s^i} + \frac{\partial c_{acc}^i}{\partial S^{-i}} \right) > 0 \quad (2h)$$

It is shown in the Appendix that these additional assumptions guarantee the existence of a unique Nash equilibrium. As explained, in a Nash equilibrium all drivers must have the same ‘joint’ speed  $S$ . We will now show that the equilibrium speed is the one that minimizes travel cost (4) over  $S (= s^i = S^{-i})$  (still treating  $k$  as given), provide the mild assumption is added that the expected accident cost for a driver with a given speed  $s^i$  is minimized over  $S^{-i}$  when speed differences are minimized and  $S^{-i} = s^i$  holds. Individual cost minimization implies that the cost function will be minimized over  $s^i$ , but of course not necessarily over the joint speed  $S$ . To demonstrate that this will be the case, we define a function  $c(S, k)$  as follows:

$$c(S, k) = c^i(s^i = S, S^{-i} = S, k) \quad (5)$$

A symmetric Nash equilibrium must be a point on this curve, and we are now interested in the question of whether the minimum of  $c(S, k)$  is such an equilibrium for a given  $k$ , and whether it is a unique Nash equilibrium. First, assumptions (2g) and (2h), together with (2a)-(2d) ensure that  $c(S, k)$  has a unique minimum with respect to  $S$  for a given  $k$ . In this minimum:

$$\frac{\partial c}{\partial S} = \frac{\partial c^i}{\partial s^i} + \frac{\partial c^i}{\partial S^{-i}} = 0 \quad (6)$$

Individual cost minimisation implies that in a Nash equilibrium,  $\partial c^i / \partial s^i = 0$  (compare (3)). For (6) to be valid in equilibrium, we must therefore also have  $\partial c^i / \partial S^{-i} = 0$ . This is the case if accident costs, for a given  $k$  and  $s^i$  and considered as a function of  $S^{-i}$ , is minimized if speed differences are minimized and all other drivers set their speeds equal to  $s^i$ . This seems a reasonable condition, and it can be stated formally as an additional assumption on the properties of the cost function (4):

$$\frac{\partial c_{acc}^i}{\partial S^{-i}} = 0 \text{ if } S^{-i} = s^i \quad (2i)$$

$$\frac{\partial^2 c_{acc}^i}{\partial S^{-i^2}} > 0 \quad (2j)$$

Figure 1 illustrates the unique symmetric Nash equilibrium that is obtained under assumptions (2i) and (2j). The figure shows in bold the  $c(S, k)$  from equation (5) as a function of  $S$  for a given  $k$ . Furthermore, it shows the individual driver’s cost function  $c^i(s^i; S^{-i}, k)$  as a function of  $s^i$  for three values of  $S^{-i}$  (still for the same, given  $k$ ):  $S^{eq}(k)$ , which is defined as the joint speed  $S$  that minimizes  $c(S, k)$ ;  $S^L$ , which is lower than  $S^{eq}(k)$ ; and  $S^H$ , which is higher than  $S^{eq}(k)$ . Finally, the figure identifies the individual driver’s optimal response to each of these values of  $S^{-i}$ ,  $s^*(S^{-i}, k)$ , as the speed  $s^i$  that minimizes  $c^i(s^i; S^{-i}, k)$ .

It follows from (2i) and (2j) that for a given  $S^{-i}$  and  $k$ ,  $c^i(s^i; S^{-i}, k)$  is at least as large as  $c(S = s^i, k) = c^i(s^i; S^{-i} = s^i, k)$ , and that the values are equal only when  $S^{-i} = s^i$ . In

other words: a driver choosing a speed  $s^i$  under a density  $k$  will experience the lowest possible costs only when other drivers also drive at a speed  $S^{-i} = s^i$ . As a consequence, for any given  $S^{-i}$  and  $k$ , the curve  $c^i(s^i; S^{-i}, k)$  as a function of  $s^i$  will touch the function  $c(S, k)$  at  $S = S^{-i}$  without crossing it.

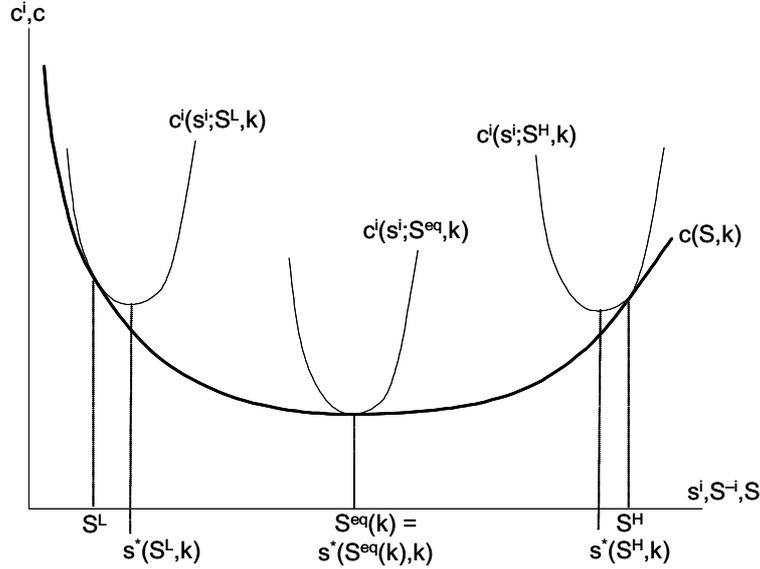


Figure 1. A symmetric Nash equilibrium for speed choice

Figure 1 illustrates why, as a result,  $S^{eq}(k)$  is the unique symmetric Nash equilibrium. First, it is a Nash equilibrium because the strict convexity of  $c(S, k)$  with respect to  $S$ , combined with equations (2i)-(2j), implies that defecting would certainly raise an individual's  $c^i$ .<sup>4</sup> The individual's optimal response  $s^*(S^{eq}(k), k)$  is therefore  $S^{eq}(k)$ . Secondly, it is symmetric because this is true for all drivers. Third, it is unique because both  $c^i$  is assumed to be twice differentiable. As explained, equations (2i) and (2j) imply that  $c^i(s^i; S^{-i}, k)$  must be tangent to  $c(S^{-i}, k)$  for  $s^i = S^{-i}$ , which implies that for  $S^{-i} \neq S^{eq}(k)$ ,  $c^i(s^i; S^{-i}, k)$  then obtains its minimum for  $s^i > S^{-i}$  when  $S^{-i} < S^{eq}(k)$ , and for  $s^i < S^{-i}$  when  $S^{-i} > S^{eq}(k)$ . In other words,  $s^*(S^L, k) > S^L$  when  $S^L$  is lower than  $S^{eq}(k)$ , and  $s^*(S^H, k) < S^H$  when  $S^H$  is higher than  $S^{eq}(k)$ , as illustrated in Figure 1. For a density  $k$ ,  $S^{eq}(k)$  is thus the unique, symmetric Nash equilibrium.

As a consequence, the equilibrium speed for a given density  $k$  can be found by minimizing  $c(S, k)$  in (5) with respect to the joint speed  $S$  and hence occurs when (6) holds:

$$\frac{\partial c(S, k)}{\partial S} = \frac{\partial c_{time}}{\partial S} + \frac{\partial c_{acc}}{\partial S} = 0 \quad (7a)$$

<sup>4</sup> Note that the Appendix proves existence of a unique Nash equilibrium without using equations (2i) and (2j). Figure 1 illustrates that when adding these assumptions, the existence of a unique symmetric Nash equilibrium in fact requires (2g) and (2h) to hold only for  $s^i = S^{-i}$ , which is of course less restrictive than requiring these inequalities to hold also when speed differences exist. We have, however, not included this consideration in stating (2g) and (2h), because these conditions appear plausible also when  $s^i \neq S^{-i}$ , and because the proof in the Appendix requires them to hold also when speed differences exist.

where we have decomposed the function  $c$  into parts referring to time and accident cost, suppressing the superscript  $i$  in order to indicate that we consider situations in which all speeds are equal. Nash equilibrium cost levels can be found by substituting the equilibrium speed  $S^{eq}(k)$  into the cost function  $c(S,k)$ :

$$c(S^{eq}(k), k) \text{ with } S^{eq}(k) = \arg \min_S c(S, k) \quad (7b)$$

From a modelling perspective, a practical benefit of using  $c(S,k)$  is that, to find Nash equilibria for different densities, there is no need to set up a many-actor model that describes how an individual's generalized cost level depends on her own and all other drivers' speed choices. It is sufficient to consider a generalized cost function  $c(S,k)$  as in (5), and minimize it with respect to  $S$ .

It is important to emphasize that the use of the cost functions (5) and (7b) should not be mistaken to imply that it is assumed that drivers coordinate their speed choices – which would of course run counter the interpretation of a Nash equilibrium. Instead, the minimization of  $c$  with respect to a joint speed  $S$  to identify the Nash equilibrium speed choice is just a computational procedure, made possible by the additional assumptions (2i) and (2j), to find the equilibrium speed that comes about in a game that satisfies standard Nash assumptions, and in which individual drivers therefore take all other drivers' speeds, as well as traffic density and flow, as given. Individuals set their individual speeds optimally through minimization of  $c^i(s^i; S^{-i}, k)$  in equation (4). The aggregate result of this behaviour is that  $c(S,k)$  in equation (5) will be minimized with respect to  $S$ .

For an analytical model, it is of course convenient to work with a deterministic model that has a unique function  $S^{eq}(k)$ . At the same time, the scatter observed in empirical speed-flow relations (*e.g.* Small, 1992, pp. 64–66) suggests that there is no reason to believe that in reality, only one equilibrium speed should correspond with a given density. It is therefore of some interest to point out that if we relax some of the assumptions that cause the equilibrium to be unique in our model, one would generally expect the other symmetric Nash equilibria also to exist only for speeds  $S$  around the minimum of a cost function as in (5). The assumptions that cause the equilibrium to be unique would therefore appear not to affect the (welfare) results in any significant way. For example, multiple equilibria might exist if (2h) does not hold, and  $c(S,k)$  has no unique minimum, but a flat minimum segment instead. In that case, there appears to be no pitfalls in interpreting the unique equilibrium speed  $S^{eq}(k)$  that we consider, as the expected value of the candidate equilibria that would then arise. Next, multiple symmetric Nash equilibria would also exist if  $c^i(s^i; S^{-i}, k)$  in equation (4) is not twice differentiable in all its arguments, but would be kinked and minimized also for some  $s^i = S^{-i} \neq S^{eq}(k)$ .  $S^{eq}(k)$  (defined according to (7b)) then remains a symmetric equilibrium, but so would be some speeds for which the middle term in (7a) is either positive or negative. The expected value of the possible equilibrium speeds may then, in reality, still be close (or even equal) to  $S^{eq}(k)$  as just defined, because the aforementioned scatter-plots suggest that equilibrium speeds for a given density only occur within a certain compact range, which includes  $S^{eq}(k)$  if (2g) and (2h) remain valid; and because there is no reason to expect that

$S^{eq}(k)$  would be near either extreme of this range. Also note that, when (2g) and (2h) hold,  $S^{eq}(k)$  requires fewer assumptions to qualify as a Nash equilibrium than other speeds do:  $S^{eq}(k)$  would be a Nash equilibrium independent of whether  $c^i(s^i; S^{-i}, k)$  is kinked at  $S^{-i}$ , whereas the other speeds do require a kink.

For analytical convenience, we will proceed under the assumptions (2a)-(2j) that guarantee a unique symmetric Nash equilibrium, that is a unique minimum of (5).

Finally, two remarks have to be made with respect to the list (2a)-(2j) of assumptions on  $c^i$ . The first one is that they provide a set of intuitively plausible assumptions that are sufficient to derive our results, but that some of them may be relaxed. For instance, assumption (2i) is only used in order to ensure the existence of a minimum of  $c(S, k)$  and all our results remain unchanged when it does not hold for situations in which there are speed differences. Some other assumptions may be relaxed as well. Second, there are other plausible characteristics of this function that have not been included in the list because we did not need them for our present purposes, for instance that it is increasing in  $\text{var}(S^{-i})$  and  $k$ .

### 2.3. *A behavioural speed-flow function and its policy implications*

We are now ready to derive a ‘behavioural’ speed-flow function. First we derive under which additional assumption on the generalized cost function  $c(S, k)$  in equation (4),  $S^{eq}(k)$  will fall if  $k$  rises, as observed in reality for sufficiently high densities. When starting in an equilibrium with (7a) satisfied, raising  $k$  would make the middle term of (7a)  $\partial c / \partial S$  positive if  $\partial^2 c_{acc} / \partial S \partial k > 0$ , as is plausible for sufficiently high densities (when the cross-derivative is zero, *e.g.* at low densities, no change in speed will be induced). To restore equality to zero,  $S$  must then fall (recall that  $\partial^2 c_{time} / \partial S^2 > 0$  and that (2h) implies  $\partial^2 c_{acc} / \partial S^2 > 0$ ).

This negative relation between  $k$  and  $S^{eq}(k)$  is consistent with the ‘fundamental diagram of traffic congestion’. As in standard expositions, we may next use the identity (8):

$$F \equiv S^{eq}(k) \cdot k \tag{8}$$

to derive a behavioural speed-flow function. As stated, we will only consider the non-hypercongested segment of this function; *i.e.*, that part for which the elasticity of  $S^{eq}(k)$  with respect to  $k$  is (in absolute terms) less than unity. For this part of the speed-flow function, a higher flow is associated with a lower equilibrium speed and a higher density.

We can now derive some important implications from using a behavioural rather than a technical speed-flow function.

#### *Marginal external costs with respect to traffic flow with Nash equilibrium speeds*

First we consider the marginal external costs with respect to flow when drivers are free to choose speed and Nash equilibria as described above result. For this purpose, it is convenient to denote the speed and density that are, according to (8), consistent with a given flow  $F$  as  $S(F)$  and  $k(F)$ . The total costs as a function of traffic flow,  $TC(F)$ , can then be found by multiplying the Nash equilibrium generalized costs of (7b) – with  $S(F)$  and  $k(F)$  substituted – with  $F$ :

$$TC(F) = F \cdot (c_{time}(S(F)) + c_{acc}(S(F), k(F))) \quad (9a)$$

Equation (9a) implies marginal costs  $mc(F)$  equal to:

$$mc(F) = c_{time}(\cdot) + c_{acc}(\cdot) + F \cdot \frac{\partial c_{time}}{\partial S} \cdot \frac{\partial S(F)}{\partial F} + F \cdot \frac{\partial c_{acc}}{\partial S} \cdot \frac{\partial S(F)}{\partial F} + F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{\partial k(F)}{\partial F} \quad (9b)$$

and hence marginal external costs  $mec(F)$  equal to:

$$mec(F) = F \cdot \frac{\partial c_{time}}{\partial S} \cdot \frac{\partial S(F)}{\partial F} + F \cdot \frac{\partial c_{acc}}{\partial S} \cdot \frac{\partial S(F)}{\partial F} + F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{\partial k(F)}{\partial F} \quad (9c)$$

These marginal external costs consist of three terms. The first (positive) term reflects the change in total travel time costs due to a speed change. The second (negative) term represents the change in total expected accident costs owing to a speed change. The third (positive term) gives the change in total expected accident costs due to a density change.

One way of grouping these three terms is therefore into marginal external time costs (the first term) and marginal external accident costs (the second and third term). The total marginal external costs are the sum of these, and so should be the optimal congestion toll. The conventional analysis of optimal congestion pricing considers the first of these three terms – the marginal external time costs – only. This clearly implies a non-optimal congestion toll when expected accident costs are relevant congestion costs, too, and explain why drivers slow down when traffic gets heavier. Indeed, the treatment of the speed-flow relation as a technical relation obscures the simple fact that people slow down in congestion for a good reason, namely to reduce expected accident costs. The associated economic analysis of congestion pricing consequently overlooks the fact that observed speed flow functions results from drivers' trade-offs between expected accident costs and time costs, and hence convey information on the marginal relevance of the external costs of both these types. This is of course not to suggest that external accident costs have been neglected altogether in studies on optimal transport pricing – but the logical connection with external time costs through drivers' speed choices in congested traffic, as reflected through the shape of the speed-flow function, has been (a review of economic studies on the relation between traffic flow and marginal accident costs is provided by Peirson, Skinner and Vickerman, 1998).

Before explaining why it is relevant to consider this connection, we first note that it is not entirely inconceivable that a toll based only on the first term in (9c) may sometimes lead to overpricing, rather than underpricing. That is, because the second term in (9c) is negative and the third positive, their sum need not be positive, and marginal external costs may fall when traffic flow increases. In general, however, one would expect both marginal external time and accident costs to increase in  $F$ : at a higher flow, it will typically not be considered optimal to slow down so much that the optimal accident risk associated with any lower flow level is exactly reached. A congestion toll accounting for marginal time costs only would then be an underestimation. In general, such a toll can be qualified as an erroneous estimation (typically an overestimation, possibly an underestimation) – unless we have the special case where road users have one particular level of  $c_{acc}(\cdot)$  they desire, regardless of the situation on

the road (*i.e.*, there would be a perfectly inelastic demand for traffic safety, a possibility that has become known as *homeostasis* in the traffic safety literature (Wilde, 1982).

A closer look at (9c) shows why it may be important to consider the mutual interdependence between marginal external time and expected accident costs explicitly. A second way of grouping the three terms is namely into marginal external costs that arise via the induced change in speed (the first two terms), and those that arise through the induced change in density (the third). Equation (7a) shows that these first two terms will in fact cancel as a result of optimizing speed choice, so that equation (9c) can be simplified as:

$$mec(F) = F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{\partial k(F)}{\partial F} \quad (9d)$$

(Figure 7 below will illustrate this for our numerical model). The conventional analysis therefore not only considers only a part of external congestion costs (only those through time losses), but ironically also a part that – according to our model – cancels against another term in the calculation of optimal congestion tolls. It is important to emphasize the root of this difference. The conventional model, assuming a technical speed-flow relation, captures the fact that an increasing flow will lead to a lower speed, and infers that as a result the time costs for other drivers increase. The first term in (9c) reflects precisely this effect. However, our behavioural model incorporates the idea that the equilibrium speed at a given flow is such that a marginal change in an individual's speed will leave her travel costs unaltered; both if the other drivers do not make a similar a speed adjustment, and if they do (the symmetric Nash equilibrium speed choice occurs at the minimum of  $c^i$  in equation (4) and at the minimum of  $c$  in equation (5)). Therefore, the induced change in the joint speed  $S$ , following a marginal increase in flow, has a zero total impact on other drivers' total travel costs. As a result, marginal external accident costs should not just be included as equally relevant congestion costs as time costs are. It is even the case that the conventional marginal external congestion costs through time losses cancel against reduced marginal external accident costs as resulting from the same speed reduction in the calculation of total marginal external costs. The result is that these total marginal external costs can be written in terms of accident costs alone.

### *The full optimum*

One might be tempted to conclude that a congestion toll set equal to the marginal external costs in equations (9c) and (9d) would be sufficient to achieve the model's full optimum. This, however, is not the case. To derive the full optimum, we first define social surplus as the difference between total benefits – given by the area below the inverse aggregate demand function,  $D(F)$  – and total travel costs. The latter are simply defined as  $F \cdot c(S, k)$ , with  $c(S, k)$  defined according to (5). It is therefore not imposed that the speed chosen should be the Nash equilibrium speed. In contrast, the welfare function will be optimized with respect to two variables out of the triplet  $F$ ,  $S$  and  $k$  – with the third following from the fundamental identity  $F \equiv S \cdot k$ . It is immaterial which of these three is substituted out of the welfare function, and we choose to maximize welfare with respect to  $F$  and  $S$ . This makes the first-order conditions

correspond directly to the optimal use of two policy instruments that we are interested in, namely a toll  $\tau$  (to control the flow) and a speed policy, stipulating the speed to be chosen. (Note that the assumptions on the cost function (1) are such that there is no welfare gain to be reaped from setting different speeds for different users, so there is no loss from working with a prescribed joint speed  $S$ ). If speed policies are binding, an equilibrium condition equivalent to (7a) of course no longer holds, as speed cannot be chosen freely by individuals. The social optimization problem then becomes:

$$\begin{aligned} \text{Max}_{F,S} W &= \int_0^F D(f)df - F \cdot \left( c_{time}(S) + c_{acc}(S, \frac{F}{S}) \right) \\ \text{s.t. : } D(F) &= \tau + c_{time}(S) + c_{acc}(S, \frac{F}{S}) \end{aligned} \quad (10)$$

The first order conditions can be written as:

$$\frac{\partial W}{\partial F} = D(F) - c_{time} - c_{acc} - F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{1}{S} = 0 \quad \Leftrightarrow \quad \tau = F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{1}{S} \quad (11a)$$

$$\frac{\partial W}{\partial S} = -F \cdot \frac{\partial c_{time}}{\partial S} - F \cdot \frac{\partial c_{acc}}{\partial S} - F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{-F}{S^2} = 0 \quad (11b)$$

in which we have re-introduced  $k$  to facilitate comparability with earlier expressions.

Equation (11a) in fact implies a simple Pigouvian tax rule equal to marginal external costs with respect to flow – consistent with what is found in the conventional model. Specifically, because  $c_{time}$  is independent of  $k$ , and using  $F \equiv S \cdot k$ ,  $\tau$  in equation (11a) can be rewritten in its familiar form as:

$$\tau = F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{1}{S} = F \cdot \frac{\partial c_{acc}}{\partial k} \cdot \frac{\partial k}{\partial F} = F \cdot \frac{\partial c}{\partial k} \cdot \frac{\partial k}{\partial F} = F \cdot \frac{\partial c}{\partial F} \quad (12a)$$

Note, however, that – as under free speed choice – again only accident costs matter for the marginal external costs with respect to flow, and again only induced changes via a change in density, not in speed (note the similarity between (9d) and (12a)). An interesting implication of this latter aspect is that, unlike in the conventional model, we may also rewrite  $\tau$  (again using  $F \equiv S \cdot k$ ) as:

$$\tau = k \cdot \frac{\partial c_{acc}}{\partial k} = k \cdot \frac{\partial c}{\partial k} \quad (12b)$$

Equation (12b) does not hold true in the conventional model, in which the value of (12b) would be zero because only speed  $S$ , not density  $k$ , determines the cost of travel.

Next, (11b) shows how the regulator should set the speed  $S$  optimally. Without speed policies, the Nash equilibrium speed  $S^{eq}$  would be such that the sum of the first two terms in the middle expression in (11b) equals zero (compare (7a)). This implies that under a free speed choice, (11b) will not be satisfied – unless there is no marginal impact of density on expected accident costs and  $\partial c_{acc}/\partial k=0$ . However, when  $\partial c_{acc}/\partial k>0$ , a freely chosen speed would lead to the middle expression in (11b) being positive, and equal to  $F^2/S^2 \cdot \partial c_{acc}/\partial k = k^2 \cdot \partial c_{acc}/\partial k$ . Raising  $S$  would then be necessary to satisfy the optimality condition (11b). A

higher  $S$  would in the first place increase the value of  $\partial c_{time}/\partial S + \partial c_{acc}/\partial S$  (because  $\partial^2 c_{time}/\partial S^2 > 0$  and  $\partial^2 c_{acc}/\partial S^2 > 0$ ), so that the first two terms in the middle expression in (11b) become negative. A higher  $S$  would furthermore reduce the value of the positive third term (which equals  $F^2/S^2 \cdot \partial c_{acc}/\partial k$ ).

We can thus conclude that – consistent with Rotemberg (1985) – the optimal speed  $S^*$  must exceed the Nash equilibrium speed in order to satisfy (11b). To obtain the full welfare optimum, the regulator would thus have to use a combination of a ‘flat’ (speed-independent) toll as in (12ab) and a compulsory speed as implied by (11b). Note that, as  $\partial c/\partial S > 0$  in the optimum, a *minimum* speed restriction could be used, too. The motivation for actively using speed policies, alongside tolling, lies in the fact that the Nash equilibrium speed choice is characterized by a zero partial derivative of generalized costs  $c$  with respect to  $S$  (as in (6)), while optimal speed choice requires a zero total derivative of generalized costs  $c$  with respect to  $S$  (as in (11b)). The difference is caused by the indirect effect of a speed change, via a change in density  $k$  when keeping flow  $F$  fixed at its optimal level, upon these generalized costs. Atomistic road users will ignore this effect.

Speed policies as described above will not be a very practical policy for city traffic, where vehicles often decelerate and accelerate, and stationary traffic conditions seldom apply. Speed regulation may appear more feasible for (long) highways (for which also the conventional model, to which our results are contrasted, is much more appropriate). Speed regulation could be enforced by manual or electronic monitoring. Besides direct speed regulation, tolls could also be set dependent on the speed chosen, so as to decentralize the choice of the optimal speed. Especially when sophisticated electronic tolling techniques are used, this may have a great advantage in terms of monitoring and compliance. As a third option, the regulator may choose to affect density and flow instead of speed and flow, and obtain the same full optimum (letting  $S$  follow from the identity  $F \equiv k \cdot S$ ). Flow could then be controlled using the toll, as above, and density by prescribing a minimum gap. (Rotemberg, 1985, considered the derivation of such an optimal gap, but for ‘gap policies’ in isolation).

Our analysis thus shows that when a behavioural representation of traffic congestion is used, the full optimum can be achieved only when both the traffic flow and the speed at which people drive are affected simultaneously. An important consequence is that the optimum combination of flow and speed are off the empirical speed-flow curve (representing Nash equilibria), and off the average and marginal cost functions that the conventional analysis derives from it. The optimum therefore by definition cannot be found in the conventional analysis, which treats the speed-flow function as an exogenous, technical relationship.

The result that the optimum speed is above the speed that would be freely chosen warrants a few final comments, not in the least place because it may run counter primary intuition, and seems to be in contradiction with practical policy making, where speed limits invariably involve *maxima*, not *minima* as suggested here. In the first place, speed limits in practice are typically motivated by safety considerations in uncongested situations, and sometimes by environmental concerns – which are absent in our model, and that might indeed change the results. Our result concerns congested traffic. Moreover, the type of congestion

considered is steady-state flow congestion, not bottleneck congestion. In other words, our result does certainly not imply that we would advise traffic regulators to increase the speed at which car drivers should approach a queue behind a fixed (tunnel) or temporary (accident) bottleneck. Our result also does not imply that we advise to set a minimum speed limit above the prevailing speed after congestion has already become severe. Instead, the combination of a toll and a lower speed limit only makes sense if evaluated from the steady state perspective, *i.e.*, both should be known beforehand, and of course if flow congestion is present on the road considered. Only then can the beneficial effect of a higher speed on lower density and hence lower expected accident costs be expected to be reaped.

### 3. An empirically calibrated simulation model

The question of course arises what the empirical relevance of the above findings might be. To the best of our knowledge, no traffic model is currently available that could be used to answer this question, so a numerical model was developed for this purpose. Especially for accident risks as a function of speed and density, we could not find any specifications in the literature that were useful for our purposes<sup>5</sup>, and an admittedly simple specification was chosen and calibrated using available empirical evidence. If only for this reason, we emphasize the rather speculative character of our numerical model, and stress that it only serves to illustrate the qualitative points made above in an empirical context. Given the potential importance of the qualitative conclusions obtained above, further empirical research seems warranted.

This section describes the structure and calibration of our numerical model. Policy evaluations are described in Sections 4 and 5 below.

A few introductory remarks are in order. For the purpose of calibration, we found it convenient to work – without loss of generality – with a normalized density  $\kappa$ , which can be defined as the ratio between the number of cars per unit of road space and the number of cars that can be packed on one lane over a unit length, *cap* (*cap* is therefore the maximum capacity in terms of density, not in terms of flow as it is usually defined). When  $k$  approaches the so-called jam capacity  $k^{jam}$ ,  $\kappa$  approaches unity. For a road with  $L$  lanes and a length  $l$ , traffic density depends as follows on the number of vehicles present on the road ( $n$ ):

$$k = \frac{n}{l \cdot L} \quad (13a)$$

while  $\kappa$  is defined as:

$$\kappa = \frac{n}{l \cdot L \cdot cap} \quad (13b)$$

The fundamental identity becomes:

$$F \equiv S \cdot k = S \cdot \kappa \cdot cap \quad (14)$$

---

<sup>5</sup> Two recent contributions worth mentioning are Dickerson, Peirson and Vickerman (2000), who study – for different types of road – the relation between aggregate flow and accident rates; and McCarthy (2001), who reviews the literature on speed limits and highway safety. Also these studies do not provide disaggregated empirical estimations of accident risks as a function of density and speed, as needed for our model.

Finally, we aim to calibrate the generalized cost function of equation (5) (with fuel costs added), and not the individual's cost function of equation (1), and hence will only consider the 'joint' speed  $S$  as an argument – not the individual's speed  $s^i$ .

### 3.1. Time costs

The time costs  $c_{time}$  experienced by a driver are simply assumed to be the product of the constant value of time,  $a_{time}$ , and the travel time,  $l/S$ :

$$c_{time} = a_{time} \cdot \frac{l}{S} \quad (15)$$

( $l$  denotes the length of the road). If  $c_{time}$  were the only relevant costs, as assumed in the conventional analysis, the optimal speed implied by (15) would be infinite.

### 3.2. Fuel costs

Compared with the models presented in Section 2, we add fuel as a third cost component. Also fuel costs typically vary with driving speed. Two types of approaches to determine fuel use as a function of speed can be distinguished in the literature.

A first approach concerns fuel use as a function of constant driving speeds, as measured for instance in laboratory experiments. The three main effects of speed on fuel use per kilometre distinguished in such studies include the fuel used per unit of time to keep the engine running, implying an inverse relation with speed; a constant; and a quadratic term which reflects increased aerodynamic and mechanical friction at higher speeds. The three together imply a U-shaped relation (e.g. Rouwendal, 1996; Pronk *et al.*, 1993):

$$c_{fuel} = l \cdot P_{fuel} \cdot \left( \frac{a_{fuel}}{S} + b_{fuel} + d_{fuel} \cdot S^2 \right) \quad (16)$$

where  $P_{fuel}$  is the fuel price, and  $a_{fuel}$ ,  $b_{fuel}$ , and  $d_{fuel}$  are parameters. Equation (16) implies that fuel consumption per kilometre will reach a minimum at some positive, finite speed – the optimal speed if only fuel costs mattered.

A second strand of studies instead focuses on fuel use in actual traffic conditions (Fwa and Ang, 1992; OECD, 1982). Fwa and Ang (1992) give a detailed overview of such studies, all of which use a simple inverse function that implies a cost function of the following type:

$$c_{fuel} = l \cdot P_{fuel} \cdot \frac{a_{fuel}^{\#}}{S} \quad (17)$$

where  $a_{fuel}^{\#}$  is a parameter different from  $a_{fuel}$  above.

The formulation in (16) ignores increased fuel use at lower speeds as caused by congestion, and the one in (17) that at higher speeds, aerodynamic friction will become more important and raises fuel use. Neither is therefore directly suitable for our analysis. Also a straightforward combination of the two – for instance the compromise of inserting an estimate or approximation for  $a_{fuel}^{\#}$  in (16) – is not satisfactory, as our model should reflect that increased fuel use at congestion-induced lower speeds is not caused primarily by the low

(average) speed itself, but is instead due to the fluctuations in this speed, which in turn result from the higher occupation of the road. To reflect these considerations, we use a formulation with (16) as the basis, but that is extended with a term that approaches unity when density approaches zero, and that is otherwise increasing in density and in speed. This reflects that the number of times a driver would have to adjust speed to avoid a collision increases both in the joint speed, given the density, and in the density, given the joint speed:

$$c_{fuel} = l \cdot P_{fuel} \cdot \left(1 + \delta_{fuel} \cdot S^{\beta_{fuel}} \cdot \kappa^{\alpha_{fuel}}\right) \cdot \left(\frac{a_{fuel}}{S} + b_{fuel} + d_{fuel} \cdot S^2\right) \quad (18)$$

where  $\alpha_{fuel}$ ,  $\beta_{fuel}$  and  $\delta_{fuel}$  are parameters. As will be discussed in Section 3.4, these parameters were calibrated such that  $c_{fuel}$  in (18) approaches an empirical estimate for (17) when congestion gets more severe, while it approaches (16) when the road is nearly empty.

### 3.3. Expected accident costs

The expected accident costs are probably the most complex to model among the three generalized cost components we distinguish, at least when aiming for the simplest possible specification for each component. Unfortunately, we were unable to find any empirical studies presenting disaggregated specifications of expected accident costs, or even risks, as a function of speed and density. Hence we developed a specification of our own, in which we consider one type of accidents only, namely those involving other vehicles (collisions with fixed objects alongside the road are ignored). On a unidirectional road where drivers have equal average speeds, this should be interpreted as head-tail collisions resulting from fluctuations in this average speed. In our analysis,  $c_{acc}$  represents the share of accident costs that will affect a driver's behaviour. It should thus be interpreted as the non-insured share of accident costs, typically the immaterial costs of pain and suffering and the value of life multiplied by the probability on a fatal accident.<sup>6</sup>

A first element governing the expected costs of accidents involves its probability,  $p_{acc}$ . We assume that this probability is 0 when the joint speed is zero, and otherwise increases more than proportionally with speed given the density, to reflect a more than proportional impact of reaction time and breaking distance. In addition, the probability increases more than proportionally with the density given the joint speed, as a greater occupation urges a driver to divide attention among more fellow road users. Moreover, these other users behave less predictable with a greater road occupation, because they too have to divide attention among more fellow road users. This leads us to the following specification:

$$p_{acc} = l \cdot \left(\delta_{acc} \cdot S^{\beta_{acc}} \cdot \kappa^{\alpha_{acc}}\right) \quad (19)$$

where  $\delta_{acc}$ ,  $\alpha_{acc}$  and  $\beta_{acc}$  are parameters, the latter two greater than 1.

Next, the expected costs of an accident, in case one indeed actually happens,  $c_{acac}$ , will generally also depend on speed (despite the actual occurrence, we still speak of 'expected'

---

<sup>6</sup> Note that we ignore both the welfare and the behavioural effects of insured accident costs and insurance premiums.

costs to reflect that the severity of a given accident for a given speed and density may still be stochastic). In general, the expected costs of a given accident can be expected to increase more than proportionally with the actual speed difference between the two cars involved, due to the greater impact of a collision. In the context of our model, the next question is to what extent implicit expected actual speed differences during the instant of a collision would vary with the explicit, identical, average speeds. We simply assume that these absolute speed fluctuations are a fixed technical function of  $S$ , which we admit is a simplification, but one that does not affect our results in any fundamental way. In addition, we assume that any accident, also one at a speed marginally above zero, will involve fixed costs  $c_{acac,f}$ , which reflect the emotional shock of hitting another vehicle, the administrative burden, time losses, *etc.* Under these assumptions, the costs of an actual accident may take the following form:

$$c_{acac} = c_{acac,f} + \delta_{acac} \cdot S^{\beta_{acac}} \quad (20)$$

The overall expected accident costs can then be written as:

$$c_{acc} = p_{acc} \cdot c_{acac} \quad (21)$$

### 3.4. Calibration of the numerical model

Sections 3.1-3.3 introduced the specific cost functions underlying our numerical model. The simplicity of the functions used was motivated by the absence of alternative models, in combination with our desire to keep the model as simple as possible, given the complicated issues studied with it. Insofar as it is capable of reproducing observed facts related to speed choice, traffic congestion, fuel use and traffic safety, it does so using probably the simplest possible underlying model, which has the advantage of minimizing the model's black-box character as well as the number of parameters that need to be calibrated.

The model has 17 parameters. Some of these could be chosen freely. In particular, we consider a single lane ( $L=1$ ) of one kilometre long ( $l=1$ ), which means that all results to be presented are standardized to a per-kilometre, per-lane basis (the unit of time will be 1 hour). Other parameters could be determined directly, using available estimates. A third category had to be determined indirectly, by assessing their impact on the implied speed-flow curve (Figure 2) and fuel-use curve (Figure 3). Table 1 gives an account of the parameter values used, and the motivation. There were fewer parameters to be determined indirectly than goals specified for those parameters. This means that given the structure of the model, there was not much freedom left for simultaneous changes in any pair of parameters and still obtain a speed-flow curve and a fuel-use curve similar to those depicted in Figures 2 and 3.

Figure 2 shows the speed-flow curve generated by the numerical model using the parameter values in Table 1. It was obtained by gradually increasing density from 1 to the maximum of 250, and combining the associated values of  $S$  and  $F (=k \cdot S)$ .

Parameter	Value	Motivation	
		Determined directly	Determined indirectly <sup>a</sup>
$\alpha_{time}$	16.2 <sup>b</sup>	'Official' average value of time for road traffic in the Netherlands (AVV, 1998)	
$P_{fuel}$	2.25 <sup>b</sup>	Average gasoline price per litre in The Netherlands, 1999-2000	
$a_{fuel}$	1	Based on Pronk <i>et al.</i> (1993) <sup>c</sup>	
$b_{fuel}$	0.0175	Based on Pronk <i>et al.</i> (1993) <sup>c</sup>	
$d_{fuel}$	$2.6 \cdot 10^{-6}$	Based on Pronk <i>et al.</i> (1993) <sup>c</sup>	
$\delta_{fuel}$	0.03		Set such that given the choice of $\alpha_{fuel}$ and $\beta_{fuel}$ , curve D in Figure 3 is sufficiently close to B between 40 and 100 km/hr
$\alpha_{fuel}$	3		<ol style="list-style-type: none"> <li>To reflect an anonymous expert's opinion that for speeds between 60 and 100 km/hr, the two opposing effects of speed on fuel use as reflected by curves A and B in Figure 3 will more or less cancel out and a flat curve D should result</li> <li>To obtain a curve D in Figure 3 sufficiently close to B for lower speeds</li> </ol>
$\beta_{fuel}$	2		As above
$\delta_{acc}$	$3 \cdot 10^{-6}$		<ol style="list-style-type: none"> <li>To obtain a risk level at a speed of 114 km/hr of <math>3 \cdot 10^{-9}</math>, well below the average risk level of <math>4 \cdot 10^{-8}</math> (per vehicle-kilometre) for head-tail collisions on Dutch highways, for which the same average speed applies (over a year; <a href="http://avvisn0.rws-avv.nl/cgi-bin/wdbcgiv/avv/AVV.home">http://avvisn0.rws-avv.nl/cgi-bin/wdbcgiv/avv/AVV.home</a>)<sup>d</sup></li> <li>To obtain a maximum flow level between 2000 and 2500 vehicles per hour per lane (Figure 2), in line with empirical findings (<i>e.g.</i> Small, 1992, Figure 3.4)</li> </ol>
$\alpha_{acc}$	5		To obtain a plausible curvature of the speed-flow function (Figure 2): <ol style="list-style-type: none"> <li>a rather flat segment up to a flow of 1500 and a strongly concave backward-bending shape beyond that point</li> <li>a maximum flow for a speed of around 60 km/hr</li> </ol>
$\beta_{acc}$	2		As above
$c_{acc,f}$	5000		Best guess of the fixed costs of an accident; <i>i.e.</i> , the willingness to pay to avoid even the least serious possible accident
$\delta_{acc}$	95000/120		Given the choice of $\alpha_{acc}$ and $\beta_{acc}$ , this value implies that the expected costs (non-insured, typically immaterial) of an accident at an average speed of 120 km/hr is Dfl 100 000
$\beta_{acc}$	1		Assumption that the expected speed difference during a collision rises less than proportional with joint speed, which exactly offsets the more than proportional increase of accident costs with expected speed difference. If the latter were quadratic, to reflect the law of kinetic energy, our assumption would be that the expected speed difference increases linearly with the square root of a constant times $S$
$l$	1	Unity (1 km) by assumption	
$L$	1	Unity (one lane) by assumption	
$cap$	250	$l$ (1 km) divided by the average length of a passenger car (4 meters)	

## Notes

- The motivations in this column give the *primary target(s)* for which the parameter was used. Evidently, as 'everything affects everything' in this model, we caution against the possible suggestion that parameters could be set independently of each other so as to realize the given target directly.
- The exchange rate of the Dutch Guilder early 2001 was Dfl 2.20  $\approx$  € 1  $\approx$  \$ 0.95.
- Pronk *et al.* proposed somewhat different values of  $a_{fuel} \approx 1.014$ ,  $b_{fuel} \approx 0.015$  and  $d_{fuel} \approx 3.049 \cdot 10^{-6}$  (see the curve A in Figure 3). At the given fuel price of Dfl 2.25, this would in absence of any other road users and hence accident risks lead to an unrealistically low free-flow speed of 110 km/hr. By lowering  $d_{fuel}$  and adjusting  $b_{fuel}$ , we obtained a somewhat flatter right-hand side of the fuel use curve (see curve C in Figure 3), implying a free-flow speed of 116 km/hr.
- The factor 8 difference is our best guess of the effects of the two facts that in reality, traffic is heterogeneous, and that the empirical figure is the average of congested and uncongested situations. In The Netherlands, in 1996 some 1750 head-tail collisions involving passenger cars occurred on highways (SWOV, personal communication), for a vehicle kilometrage of 47 billion (AVV, 2000)

Table 1: Calibrated parameter values (cost-side of the model) and motivation

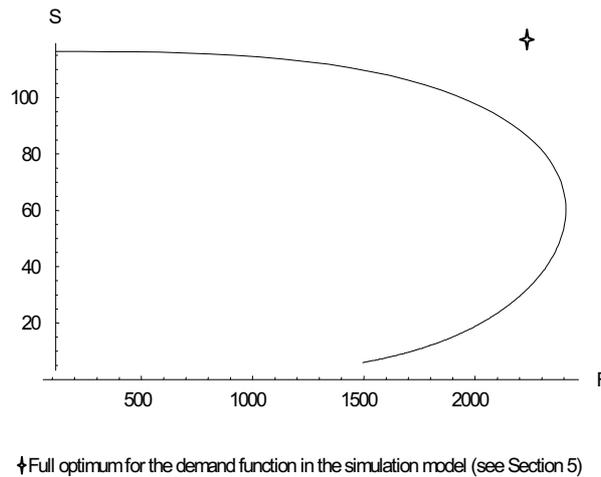


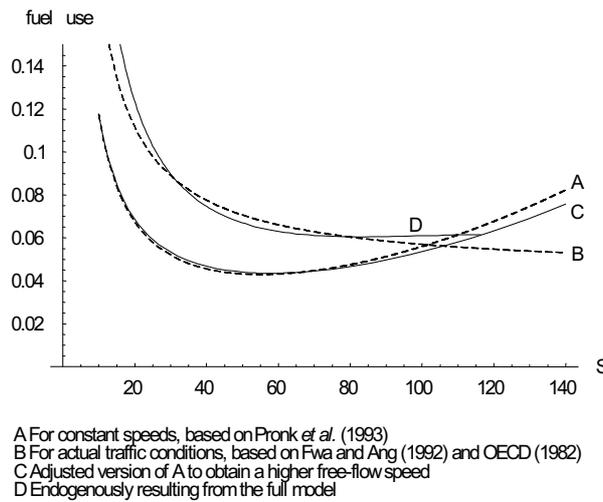
Figure 2. The speed-flow curve generated by the numerical model, and the full optimum for the assumed demand function

Note that even for a fully packed road, our model suggests a positive speed of 6 km/hr, so that road users will then choose not to incur infinite time costs by staying on the road forever. For all points shown, generalized costs are minimized with respect to  $S$  (taking  $k$  as given) as required for a symmetric Nash equilibrium (see Section 2.2); see also Figure 4 below. The curve has the familiar backward-bending shape, and a maximum flow of 2408 vehicles per hour at a speed of 60 km/hr, which makes it similar to speed-flow curves as observed in reality (e.g. Small, 1992, Figure 3.4).

The lower segment of the curve corresponds to hypercongestion. Hypercongestion is often observed in reality. Verhoef (1999, 2000) claims that although such equilibria can exist as stationary states in a model without downstream (fixed or moving) bottlenecks, they are dynamically unstable in the sense that there is no path from any other stationary state that would lead to a queue-free hypercongested stationary state.<sup>7</sup> However, it is the first part of the previous sentence – hypercongested equilibria *can* exist as stationary states – that makes the views expressed in those papers consistent with the speed-flow curve as depicted in Figure 2.

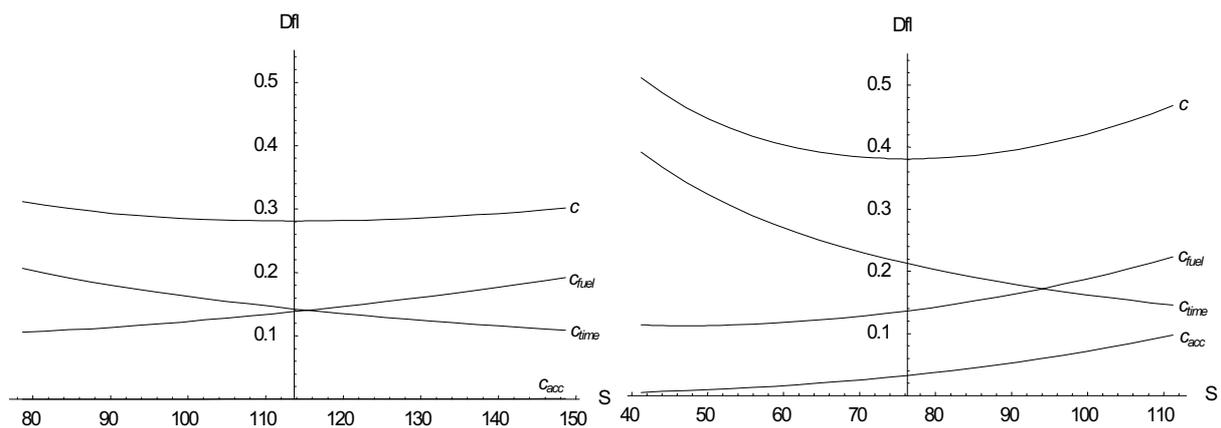
Next, Figure 3 shows fuel use as a function of  $S$ . The dashed curves A and B show the empirical estimates of equations (16) and (17) that were used as the basis for the calibration. Curve A gives (16) as presented by Pronk *et al.* (1993), which was adjusted slightly to the solid curve C for reasons outlined in Table 1. Curve B gives (17) as estimated for Dutch highways, which was found in Fwa and Ang (1992) quoting OECD (1982). Curve D gives the equilibrium fuel use function as given in (18) for our model, calculated for the combinations of  $S$  and  $k$  underlying the speed-flow curve in Figure 2. As we attempted, the curve approaches curve C for freely flowing traffic, and gets close to curve B for lower speeds. Note that curve D in Figure 3 is defined up to the free-flow speed of 116 km/hr, only, but that equation (18) allows the determination of fuel use also for higher speeds (as in Section 5).

<sup>7</sup> According to this view, observed hypercongestion would thus always have to result from a downstream bottleneck (for a similar view, see also Daganzo, Cassidy and Bertini, 1999).



*Figure 3. Fuel use as a function of equilibrium speed (litres/km): sources and endogenously determined function*

Finally, Figure 4 shows the generalized costs  $c$ , and its three components, as a function of  $S$  for equilibria corresponding with the average speed on Dutch highways (around 114 km/hr; left panel) and the base-case equilibrium that will be used in Sections 4 and 5 below (around 76 km/hr; right panel). The upper curves depict generalized costs as a function of  $S$  (for a given density  $k$ ), and show that these are indeed minimized for the equilibrium speed (in the centre of both diagrams). Furthermore, the diagrams show that expected accident costs are practically insignificant for the higher speed equilibrium in the left panel (having an equilibrium value of Dfl 0.0004 per kilometre), whereas they become significantly higher for the lower speed (base-case) equilibrium in the right panel (Dfl 0.03 in the equilibrium).<sup>8</sup>



*Figure 4. Generalized costs per kilometre for the average speed on Dutch highways (left panel) and the base-case of Sections 4 and 5 (right panel) as a function of joint speed  $S$  (keeping density  $k$  fixed); equilibrium speed at the intersection of the axes*

<sup>8</sup> The exchange rate of the Dutch Guilder in 2002 was Dfl 2.20  $\approx$  € 1  $\approx$  \$ 1.

Time costs have the expected shape. Fuel costs are rising more sharply than what might be expected on the basis of Figure 3. The reason is that Figure 3 shows equilibrium levels of fuel use, while Figure 4 shows the effects of a change in  $S$  keeping density fixed. Finally, as expected, generalized costs are higher in the right panel (Dfl 0.28 *versus* Dfl 0.38). The associated curves may be flatter than one may have anticipated. However, recall that we consider costs as a function of the joint speed  $S$  in these figures, *not* of an individual's speed  $s^i$  keeping others' speeds fixed.

This concludes the discussion of the calibration of the cost side of our numerical model. It turned out that we were able to base some parameter values on direct estimates (such as value of time and fuel prices), and choose the remaining ones such that the model produces a speed-flow function and a fuel consumption function similar to those observed in reality. Also the main endogenous variables (risk levels, range of speeds for the speed-flow curve, the maximum flow and the speed at which it occurs) are conform empirical evidence. However, we re-emphasize the model's illustrative character, implying that the main results as discussed in Sections 4 and 5 below should be treated with sufficient care.

We conclude this section by specifying the demand function to be used. A simple linear inverse demand function is assumed to apply, which has – consistent with the steady-state character of the model – flow  $F$  as the argument:

$$D = \delta - \alpha \cdot F \tag{22}$$

With  $\delta=1.32$  and  $\alpha=0.0004$ , the equilibrium depicted in the right panel was obtained, for which an equilibrium demand elasticity of  $-0.4$  applies.

#### 4. 'Flat' tolls: the performance of a 'naïve' regulator

The potential relevance of using a behavioural approach can in the first place be assessed by comparing the 'true' welfare optimum with flat tolls – according to the behavioural model, so assuming that that is the correct model – with the performance of what we will call a 'naïve' regulator, who is not aware of the underlying behavioural model and uses the conventional procedure for setting congestion tolls.

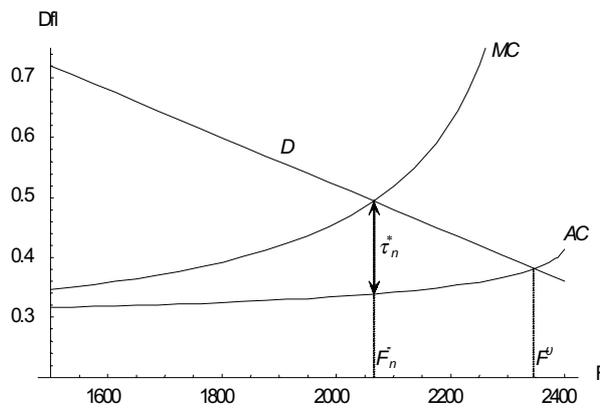


Figure 5. The market diagram according to the 'naïve' regulator

Specifically, we assume that the naïve regulator sets the congestion toll he expects to be optimal, based on a correct knowledge of the speed-flow curve, the value of time, and the demand relation. To do so, he manipulates the upper segment of the speed-flow curve in Figure 2 into an average cost curve  $AC$  by using  $c_{time} = \alpha_{time} \cdot l/S$ . A marginal cost curve  $MC$  is next obtained according to  $MC = AC + \partial AC / \partial F$ , and a toll  $\tau_n$  can be found as the difference between these two in the ‘optimum’ where  $D=MC$  holds. Figure 5 shows the resulting market diagram.<sup>9</sup>

In the non-intervention equilibrium, a flow of  $F^0=2347$  and a speed of 76 km/hr applies. The naïve regulator identifies an ‘optimum’ with a flow of  $F_n^*=2065$  and a speed of 95 km/hr, which he expects can be realized using a toll  $\tau_n^*$  equal to Dfl 0.155 per kilometre (Table 2 in Section 5 summarizes the key features of the various equilibria and optima considered in Sections 4 and 5). When actually applying this toll in the true model, a flow of 2107 and a speed of 93 km/hr will in fact result. The difference is due to changes in marginal expected accident costs and fuel costs, which are not anticipated by the naïve regulator. For the sake of the argument, however, we assume that the naïve regulator ignores the implied discrepancies with his prediction. The question we wish to address – for reasons that will become apparent below – is namely to what extent the *predicted* naïve toll and its welfare effects would deviate from the truly optimal flat toll and its efficiency gains.

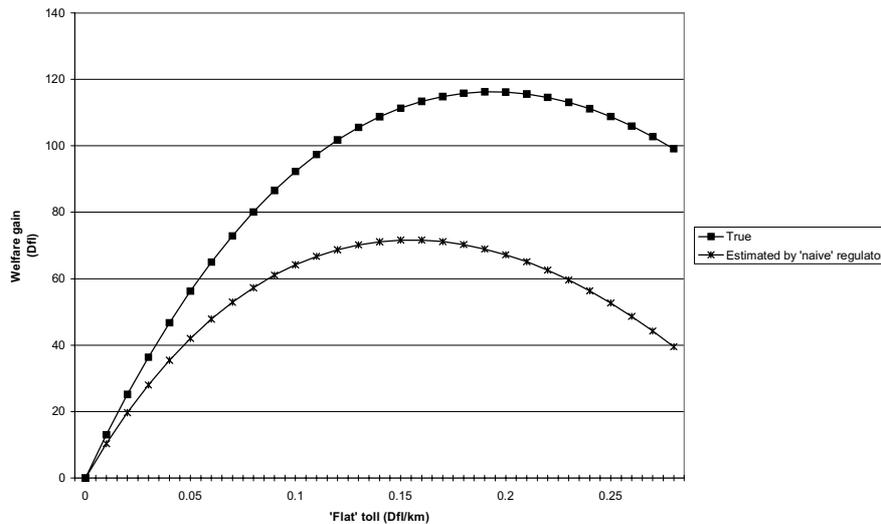


Figure 6. True welfare gains and welfare gains as estimated by a ‘naïve’ regulator as a function of ‘flat’ (speed-independent) tolls

To answer this question, Figure 6 shows the true welfare gains, and those as predicted by the naïve regulator, as a function of the ‘flat’ (speed-independent) toll  $\tau$ . The latter obtains an

<sup>9</sup> Note that we assume that the naïve regulator treats all costs other than time costs in the non-intervention equilibrium as constant average costs, which has led to an upward shift of the  $AC$  and  $MC$  to values above time costs alone. Otherwise, the cost function  $AC$  and demand function  $D$  could not produce the observed non-intervention flow  $F^0$ .

optimum at  $\tau=0.155$ , consistent with Figure 5 above, for which a welfare gain of Dfl 72 (per hour per kilometre highway) is predicted.<sup>10</sup> The upper curve shows that the truly optimal ‘flat’ (*i.e.*, speed-independent) toll amounts to  $\tau_{flat}^*=0.195$ , yielding a welfare gain of Dfl 116.

Figure 6 conveys two important messages. First, as predicted in Section 2.3, the naïve regulator would set the flat toll too low. In our example,  $\tau_{flat}$  is some 25% higher than  $\tau_n^*$ . However, as shown by the upper curve, the welfare implications of this ‘mistake’ are, relatively speaking, much smaller. Due to the concave shape of the true welfare gains as a function of the flat toll,  $\tau=0.155$  leads to a gain of Dfl 112: some 96% of the maximum gains that can be achieved with a flat toll, and much higher – nearly 60% – than what the naïve regulator anticipates and indeed will ever be aware of. Secondly, the truly achievable welfare gains of congestion pricing are much higher than anticipated by the naïve regulator: Dfl 116 *versus* Dfl 72, which means more than 60% higher.

The conclusion is therefore that our numerical model suggests that the main efficiency losses that can be expected from using the conventional instead of a behavioural model to determine congestion tolls would *not* be primarily due to the setting of an erroneous toll level. The concave shape of welfare gains as a function of toll levels can be expected to be a general phenomenon, as the absolute difference between marginal benefits and marginal social costs will generally decrease in the absolute deviation from the optimal flow. As a result, any x% downward deviation from the optimal toll will typically lead to a less than x% loss in welfare gains compared to the true optimum (with  $0 < x < 100$ ). Instead, our model suggests that a more important source of inefficiency would be due to the underestimation of welfare gains of congestion tolling – both optimal, and using the naïve toll. As a result, congestion tolls may often not be put into practice at all, in which case of course neither the expected nor the actual welfare gains from ‘naïve’ tolling will ever be realized.

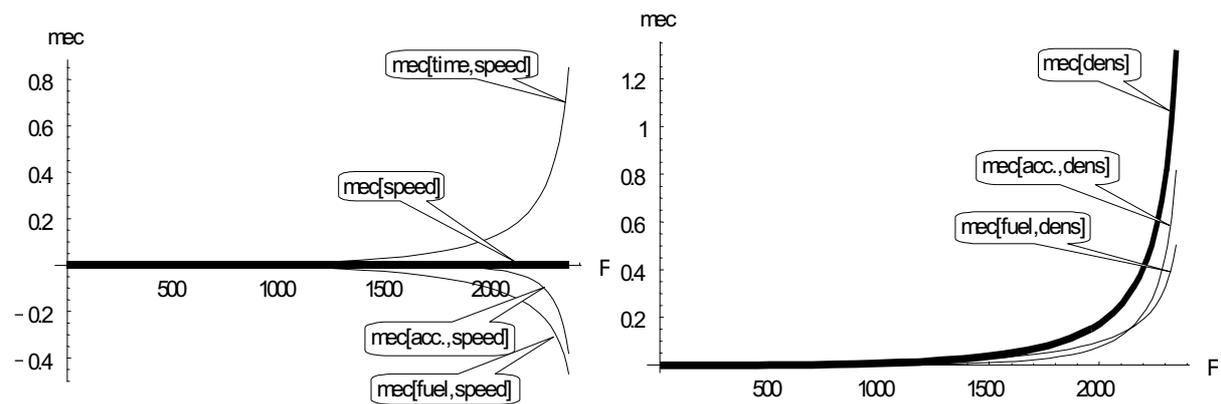


Figure 7. The five marginal external cost components under Nash equilibrium speed choice: via induced changes in speed (left panel) and density (right panel)

Finally, we use the numerical model to illustrate the result that under Nash equilibrium speed choice, the conventional marginal external costs with respect to flow (due to time losses)

<sup>10</sup> All tolls reported are determined at a Dfl 0.005 precision level.

cancel against other marginal external cost components related to induced speed changes. The left panel in Figure 7 shows the relevant three marginal external cost components as a function of traffic flow. The upper curve,  $mec[\text{time}, \text{speed}]$ , shows the conventional marginal external costs, resulting from travel time losses due to a lower speed (the notation for the other marginal external cost components follows the same logic). It has the expected shape. However, speed reductions following an increase in flow reduce the marginal external fuel and accident costs, as shown by the other two curves, and the sum of these three – shown by the bold  $mec[\text{speed}]$  line – is indeed equal to zero, consistent with our discussion of (9c) and (9d) in Section 2.3. The total marginal external costs are therefore equal to the sum of those components that are related to induced changes in density, shown in the right panel. Their sum, given by the bold  $mec[\text{dens}]$  curve, therefore corresponds with the ‘true’ marginal external costs. It was verified that at the optimal flow under ‘flat’ pricing ( $F = 2031$ ), these marginal external costs are indeed equal to the required toll level (Dfl 0.1936) (note that for this calculation, we used a greater precision for the optimal toll than elsewhere).

## 5. Toll- and speed policies: the full optimum

The numerical model also allows us to study the claim that the full optimum requires simultaneous interventions with respect to flow and speed in an empirical context. To that end, Figure 8 displays welfare gains as a function of (flat) toll levels and minimum speed limits, the latter shown as speeds in km/hr above the speed that would be chosen freely in an equilibrium with the flat toll alone (this speed is given in parentheses along the toll axis). Since the limits will be binding, actual speeds will be equal to the minimum limit specified, and the policy of minimum speed limits is in fact equivalent to one of exactly prescribed speeds. The ‘front arch’, along the toll axis, reproduces the upper curve in Figure 6.

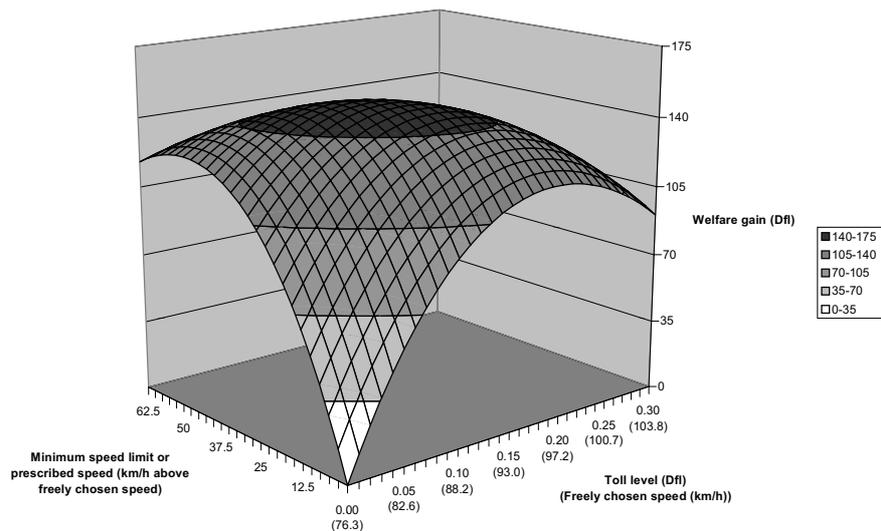


Figure 7. Welfare gains as a function of tolls and minimum speed limits

The figure shows that by using these two instruments together, further welfare gains can be realized compared to the use of flat tolls alone: Dfl 150 *versus* Dfl 116, an increase with nearly 30%. The full optimum involves a significantly lower optimal toll  $\tau^*_{opt}$  than flat tolls in isolation:  $\tau^*_{opt}=0.12$  *versus*  $\tau^*_{flat}=0.195$ ; or only just above 60%. At the same time, drivers are requested to drive considerably faster than the 90 km/hr that would apply in an equilibrium with a flat toll of Dfl 0.12 alone: 121 km/hr – which is even above the free flow speed of 116 km/hr applying in our numerical model. As a result, the full optimum (marked with a star in Figure 2) is off the speed-flow curve. Incidentally, it occurs at a speed that implies an average (time) cost level even below the free-flow (time) costs as considered by a naïve regulator (the nearly horizontal segment of *AC* in Figure 5), which nicely illustrates in a different way the impossibility of identifying this full optimum using the conventional procedures.

	Non-intervention	Flat tolls by naïve regulator		Optimal flat tolls	Prescribed speeds	Full optimum: tolls and speeds
		Believed	True			
Flow	2347	2066	2107	2028	2479	2220
Speed	76	95	93	97	124	121
Costs:						
$c$	0.381	0.339	0.322	0.314	0.328	0.312
$c_{fuel}$	0.136	–	0.137	0.137	0.182	0.167
$c_{time}$	0.212	0.170	0.173	0.167	0.131	0.134
$c_{acc}$	0.032	–	0.012	0.009	0.016	0.009
Toll	0	0.155	0.155	0.195	0	0.120
Welfare gain	0	72	112	116	127	150
% of first-best	0%	100% (believed) 48% (actual)	75%	77%	85%	100%

Table 2. Key characteristics of the various equilibria considered

As described in Section 2.3, the additional welfare gains are due to the beneficial impact that a higher speed, for a given flow leading to a lower density, has on expected accident costs and – in the numerical model – fuel costs. Note that, when comparing flat tolling alone and the full optimum, these gains are for the numerical model nearly fully ‘absorbed’ by the higher flow that becomes possible at nearly the same level of generalized costs, and that in turn is made possible as a market equilibrium due to the lower toll level applying.

Speed policies thus appear a strong instrument in our numerical model. This is in the first place illustrated by the rather fierce use of the instrument when combined with tolls, and by the fact that doing so allows optimized flows to go up from 2028 to 2220 while keeping average generalized costs nearly unchanged at Dfl 0.31 per kilometre (see Table 2; note the interesting but intuitively plausible aspect that whereas  $c_{acc}$  has remained virtually unchanged, the magnitudes of  $c_{time}$  and  $c_{fuel}$  have moved in opposite directions between both optima).

This same point is illustrated perhaps even more dramatically by the fact that for our numerical model, speed policies alone perform better than flat tolling alone – an unexpected result for an economic model. Whereas the optimal flat toll in isolation achieves 77% of the welfare gains of the optimal combination of policies, speed policies alone realize a gain of Dfl

127, or 85%. Clearly, this result is due to the specific cost functions used in the numerical model, and may be reversed to the more comfortable conclusion (at least for economists) that tolls alone are more efficient than speed policies alone, when different functions are used. Different results could of course also be obtained when environmental costs were included, or when a different demand elasticity were used. The important point here, however, is that our numerical model suggests that speed policies *may* lead to substantial welfare benefits for roads suffering from flow congestion, and may therefore deserve more attention in the design of congestion policies.

Finally, we noted in Section 2.3 that instead of using speed restrictions and tolls together, tolls could also be set dependent on the speed chosen in order to decentralize the choice of the optimal speed. In the optimum, the absolute *level* of the toll should then be equal to the value as given above, while the *slope* of the ‘toll gradient’ (the toll as a function of an individual’s speed) in the optimum should induce each road user to choose the optimum speed. Specifically, the optimum speed should become cost-minimizing when a road users takes into account the effect of speed chosen on the toll to be paid.

Given our consideration of symmetric equilibria and joint speeds  $S$  only, we are in fact unable to derive an analytical expression for the optimal slope of the gradient,  $\tau'(s^i)$ , for our current model specification. This slope should reflect the marginal costs for all other users resulting from a marginal change in speed by one individual user in the optimum – which can not be determined unless the effect of speed differences on generalized costs is modelled explicitly. Likewise, an individual will trade off  $\tau'(s^i)$  against the private costs of a marginal change in his own speed  $s$ , assuming all others’ speeds fixed. This too cannot be analyzed without an explicit modelling of the effect of speed differences on generalized costs. This all reflects that an individual considering a deviation from an equilibrium speed will not assume that all other road users will make the same deviation simultaneously.

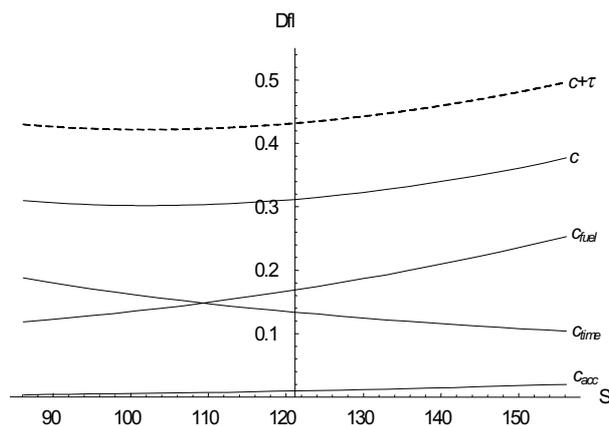


Figure 8. Generalized costs per kilometre for the full optimum as a function of joint speed  $S$  (keeping density  $k$  fixed); optimal speed at the intersection of the axes

We can nevertheless provide some insight into the numerical value of  $\tau'(s^i)$  in the optimum of our simulation model. Figure 8 shows that in this optimum, contrary to what was seen in

Figure 4,  $c$  is *not* minimized with respect to the joint speed  $S$  (keeping  $k$  fixed at its optimal level!) – which is why speed regulation is necessary in the first place. If  $c'(S)$  were taken as the basis for determining the optimal slope of the toll gradient, a value of around  $-0.001$  would be found as the optimal value of  $\tau'(S)$ : every 10 km/hr speed reduction in the joint speed  $S$  would have to lead to an increase in toll with Dfl 0.01 per kilometre driven. (This value was determined by evaluating the derivative of  $c$  in Figure 8 with respect to  $S$ , which is not shown graphically). This would lead  $c(S)+\tau(S)$  to obtain a minimum at the optimal speed of 121 km/hr. Depending on the magnitude of  $c'(s^i)$  relative to  $c'(S)$ , the policy variable of true interest,  $\tau'(s^i)$ , may be smaller than, equal to, or larger than  $\tau'(S)$  thus derived.

## 6. Conclusion

This paper presented a behavioural model of highway traffic congestion, which does not treat the speed-flow curve relation as technical relationship, but instead as resulting from individual's cost-minimizing speed choices instead. The model incorporates the basic but often ignored fact that people slow down in traffic congestion for a good reason, namely that otherwise accident risks would become excessive. Since accident risks depend on the presence of other users, congestion externalities therefore not only include time costs, but also expected accident costs (and fuel costs, as in our numerical model).

A number of conclusions stand out. A first is that congestion tolls suggested by cost curves derived from a speed-flow function in the conventional way (by multiplying the inverse of speed by the value of time for each flow level to obtain the average cost function) are typically not optimal when other externalities cause drivers to slow down when traffic flow increases. Secondly, this conventional toll in fact vanishes with Nash equilibrium speed choice in our model, because the extent to which drivers slow down and choose to incur higher time costs follows from a trade off with reduced expected accident and fuel costs, as depending on speed – which makes the derivative of travel costs with respect to speed zero in equilibrium. The third conclusion is that the full welfare optimum in our model occurs off the speed-flow function, and off the average and marginal cost functions as derived from it in the conventional analysis. The first result obtains whenever lower speeds with increased flow are not the result of some exogenous technical law, but instead are chosen consciously to avoid a part of the higher accident and fuel costs that would result from sticking to the original speed after an increase in road use. The second result rests on the Nash equilibrium speed choice occurring not only where the individual's travel costs are minimized with respect to her own speed, but also to the 'joint' speed. The third result requires individual drivers to treat aggregate variables as equilibrium flow and density as given. All requisites seem plausible.

The full optimum in our model requires either the combination of flat (speed-independent) tolls with speed restrictions – but perhaps surprisingly, a *minimum* rather than a *maximum* speed limit would then have to be set – or tolls to be dependent on the speed chosen. This is consistent with the finding of Rotemberg (1985). The optimum speed exceeds the Nash equilibrium speed because a higher speed, via a lower density (for a given, optimized flow), reduces accident risks. Similar results would be obtained if the regulator

could control flow and density, instead of flow and speed. Conventional economic analyses have focused on the control of one out of these three variables only, typically flow.

Our numerical model could reproduce speed-flow and fuel-use functions similar to what is observed in reality, using – where possible – parameter estimates taken directly from other sources. It suggests that the main efficiency losses that can be expected from using the conventional reduced-form representation instead of a behavioural model would *not* be primarily due to the setting of erroneous tolls. Instead, a more important source of inefficiency would be due to the underestimation of welfare gains of congestion tolling using the conventional approach, and the not unlikely consequence that congestion tolls would not be put into practice at all. The model further suggests that speed policies indeed may be a strong instrument in the regulation of flow congestion. Surprisingly, speed policies alone even outperformed tolling alone in terms of efficiency gains. This conclusion becomes even more noteworthy when realizing that it was obtained in a model in which road users are identical, and choose the same speed in all equilibria considered. Speed policies could have been expected to be relevant only for situations where a significant dispersion in speeds chosen exists.

All in all, it seems worthwhile to investigate the trade-offs made by individual drivers when choosing a speed in far greater (empirical) detail than has been done to date, so that a more reliable empirical specification can be used to further investigate the empirical relevance of our main qualitative findings. The numerical model used here only suggests that this relevance *could* be great; it remains to be seen whether it indeed *will* be great.

## References

- Arnott, R., A. de Palma and R. Lindsey (1993) "A structural model of peak-period congestion: a traffic bottleneck with elastic demand" *American Economic Review* **83** (1) 161-179.
- Arnott, R., A. de Palma and R. Lindsey (1998) "Recent developments in the bottleneck model". In: K.J. Button and E.T. Verhoef (eds.) (1998) *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* Edward Elgar, Cheltenham, pp. 79-110.
- AVV (1998) *Advies Inzake Reistijdwaardering van Personen (Advice on Value of Travel Time for Passengers; in Dutch)* Adviesdienst Verkeer en Vervoer (AVV), Rotterdam.
- AVV (2000) *Verkeersgegevens: Jaarrapport 1999 (Traffic Statistics: Year Report 1999; in Dutch)* Adviesdienst Verkeer en Vervoer (AVV), Rotterdam.
- Boardman, A.E. and L.B. Lave (1977) "Highway congestion and congestion tolls" *Journal of Urban Economics* **4** 340-359.
- Daganzo, C.F., M.J. Cassidy and R.L. Bertini (1999) "Possible explanations of phase transitions in highway traffic" *Transportation Research*, **33A**, 365-379.
- Dickerson, A., J. Peirson and R. Vickerman (2000) "Road accidents and traffic flows: an econometric investigation" *Economica* **67** 101-121.
- Fwa, T.F. and B.W. Ang (1992) "Estimating automobile fuel consumption in urban traffic" *Transportation Research Record* **1366** 3-10.
- Keeler, T.E. and K.A. Small (1977) "Optimal peak-load pricing, investment and service levels on urban express ways" *Journal of Political Economy* **85** 1-25.
- Knight, F. (1924) "Some fallacies in the interpretation of social costs" *Quarterly Journal of Economics*, **38**(4), 582-606.
- Lindsey, C.R. and E.T. Verhoef (2000) "Traffic congestion and congestion pricing" Discussion paper TI 2000-101/3, Tinbergen Institute, Amsterdam-Rotterdam.
- McCarthy, P. (2001) "Effects of speed limits on speed distribution and highway safety: a survey of recent literature" *Transport Reviews* **21** 31-50.
- McDonald, J.F., E.L. d'Ouille and L.N. Liu (1999). *Economics of Urban Highway Congestion and Pricing. Transportation Research, Economics and Policy*. Kluwer, Dordrecht.
- OECD (1982) *Automobile Fuel Consumption in Actual Traffic Conditions* Organization for Economic Cooperation and Development (OECD), Paris.
- 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) *Wealth and Welfare* Macmillan, London.
- Pronk (1993) *Forecasting Air Pollution by Traffic Simulation* Netherlands Economic Institute (NEI), Rotterdam.
- Rotemberg, J.J. (1985) "The efficiency of equilibrium traffic flows" *Journal of Public Economics* **26** 191-205.
- Rouwendal (1996) "An economic analysis of fuel use per kilometre by private cars" *Journal of Transport Economics and Policy* **30** 3-14.
- Small, K.A. (1992). *Urban Transportation Economics. Fundamentals of Pure and Applied Economics*. Harwood, Chur.
- Small, K.A. and J.A. Gómez-Ibáñez (1998) "Road pricing for congestion management: the transition from theory to policy" In: K.J. Button and E.T. Verhoef (eds.) (1998) *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* Edward Elgar, Cheltenham, pp. 213-246.
- Verhoef, E.T. (1999) "Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing" *Regional Science and Urban Economics* **29** 341-369.
- Verhoef, E.T. (2000) "An integrated dynamic model of road traffic congestion based on simple car-following theory: exploring hypercongestion" *Journal of Urban Economics* **49** 505-542.
- Vickrey, W.S. (1969) "Congestion theory and transport investment" *American Economic Review* **59** (Papers and Proceedings) 251-260.
- Wilde, G.J.S. (1982) "The theory of risk homeostasis: Implications for safety and health" *Risk Analysis* **2** 209-225.

### Appendix: Existence and uniqueness of a Nash equilibrium

It is observed in the main text of the paper that (2a)-(2f) guarantee existence of a unique positive solution to the individual driver's optimisation problem and that in a Nash equilibrium the speed of all drivers must be identical. We denote this individual equilibrium speed as  $s^*(S,k)$ . It is a continuous and differentiable function of  $S$ . In order to show that a unique Nash equilibrium exists, we have to establish that there is a unique speed,  $S^{eq}$ , for which  $S^{eq}=s^*(S^{eq},k)$ . In order to do this, we take a closer look at the function  $s^*(S,k)$ .

Start by observing that  $s^*$  is always positive and finite. Assumption (2a) excludes the possibility that there is a zero-speed equilibrium for individual drivers. This holds also for  $S=0$ , and hence  $S^{eq}=0$  cannot be the equilibrium and  $s^*(0,k)$  must be positive.

Next, consider  $ds^*/dS$ . Given that  $s^*$  is positive and finite for  $S=0$  and that it is a continuous function of  $S$ , we can be sure that there is a unique speed for which  $s^*=S$  if the  $ds^*/dS$  is always less than 1.<sup>11</sup> From (3) we derive:

$$\frac{ds^*}{dS} = -\frac{\frac{\partial^2 c_{acc}}{\partial s \partial S}}{\frac{\partial^2 c_{time}}{\partial s^2} + \frac{\partial^2 c_{acc}}{\partial s^2}} \quad (A1)$$

The denominator of the ratio on the right hand side of this equation is positive by the convexity assumptions (2b) and (2f). The numerator gives the effect of an increase in the speed of all other drivers on the marginal expected accident cost. If it is positive,  $ds^*/dS$  is negative and hence smaller than 1.

Assumption (4b) can be elaborated as follows:

$$\frac{\partial^2 c_{acc}}{\partial s^2} + \frac{\partial^2 c_{acc}}{\partial S \partial s} > 0 \quad (A2)$$

and this implies:

$$-\frac{\partial^2 c_{acc}}{\partial S \partial s} / \frac{\partial^2 c_{acc}}{\partial s^2} < 1. \quad (A3)$$

Using (A1), it is easy to verify that this ensures that  $ds^*/dS < 1$ . Hence there exists a unique Nash equilibrium.

---

<sup>11</sup> The assumption that  $ds^*/dS$  is smaller than 1 is sufficient for the existence of a unique value  $S^{eq}$  if both  $s$  and  $S$  are restricted to a finite interval  $[0, s^{max}]$ , with  $s^{max}$  a maximum possible speed. This condition is obviously satisfied in reality and therefore not stated explicitly. (If the two speeds are not restricted to a finite interval, a sufficient condition for existence and uniqueness would be that  $ds^*/dS$  is smaller than  $1-\varepsilon$  for some strictly positive  $\varepsilon$  and reasonable conditions that guarantee this could be introduced.)