TI 2013-028/VIII
Tinbergen Institute Discussion Paper

Equilibrium at a Bottleneck when Long-Run and Short-Run Scheduling Preferences diverge

Stefanie Peer
Erik T. Verhoef

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

EQUILIBRIUM AT A BOTTLENECK WHEN LONG-RUN AND SHORT-RUN SCHEDULING PREFERENCES DIVERGE

Stefanie Peer*[1,2] and Erik T. Verhoef[1,2]

[1] *Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam*
[2] *Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam*

February 12, 2013

Abstract

We consider equilibrium and optimum use of a Vickrey road bottleneck, distinguishing between long-run and short-run scheduling preferences in an otherwise stylized scheduling model. The preference structure reflects that there is a distinction between the (exogenous) 'long-run preferred arrival time', which would be relevant if consumers were unconstrained in the scheduling of their activities, versus the 'short-run preferred arrival time', which is the result of an adaptation of travel routines in the face of constraints caused by, in particular, time-varying congestion levels. We characterize the unpriced equilibrium, the social optimum as well as second-best situations where the availability of the pricing instruments is restricted. All of them imply a dispersed distribution of short-run preferred arrival times. The extent of dispersion in the unpriced equilibrium, however, is higher than socially optimal.

*Corresponding author (E-mail: s.peer@vu.nl)

1 Introduction

In this paper, we introduce a distinction between long-run and short-run scheduling preferences in an equilibrium setting. We assume a standard bottleneck technology that is dynamic in nature. Since the original study by Vickrey (1969), this bottleneck model has become the workhorse model for the analysis of equilibrium and socially optimal timing of usage of congestible facilities (e.g. Arnott et al., 1990, 1993). Applied to traffic congestion, the basic idea is that drivers have common preferences to arrive at a certain place at a certain time, a typical example being the morning rush hour. If the road capacity is not sufficient to accommodate all drivers in such a way that all of them can arrive at their preferred arrival moment, a queue will form in front of the bottleneck. In equilibrium, drivers who arrive close to their preferred arrival time will spend a considerable amount of time queuing in front of the bottleneck, while those who decide to travel early or late in the peak will face lower travel times at the cost of arriving earlier or later than their preferred arrival time. The costs resulting from earliness and lateness with respect to the preferred arrival time are commonly referred to as schedule delay costs.

Unlike the standard bottleneck model, the model introduced in this paper distinguishes between long-run and short-run scheduling decisions. In the long run, commuters decide on their optimal arrival routines, while in the short run, they choose their optimal departure times subject to these arrival routines. Individuals are therefore less constrained in their long-run choices than in their short-run choices. More specifically, this implies that in the long run they are able to optimize their commuting routines, trading off the time-varying average congestion levels over time of the day against deviations from their "long-run preferred arrival time" (LRPAT). The latter is defined as the preferred arrival time they would have under uncongested conditions, and can be interpreted as a preference that is driven by external factors, which may for instance be biological (such as daylight (e.g. Weiss, 1996)) or institutional (such as positive temporal agglomeration forces at work (e.g. Henderson, 1981)). Also the number of working hours and scheduling restrictions arising from other activities may affect this long-run preferred arrival time (e.g. Jenelius et al., 2011; Zhang et al., 2005). In the short run, the travel routines that have been chosen in the long run are fixed, and the optimized arrival time from the long-run problem becomes the preferred moment of arrival, which we refer to as "short-run preferred arrival time" (SRPAT). Daily short-run departure time decisions are thus made in the face of the routines chosen in the long run, taking into account the bottleneck capacity on that particular day and the resulting time-varying congestion levels.

The distinction between short-run and long-run behavior becomes only relevant when consecutive peaks are not exact replicas. Otherwise, travelers are likely to end up in a less interesting corner solution where they either equate their LRPAT and SRPAT, or to choose the same departure time everyday that results in an arrival time identical to the SRPAT. To make the distinction useful, we consider a bottleneck with stochastic, day-specific capacity, such that a difference exists between the long-run problem of choosing the SRPAT considering expected travel times, versus the short-run problem of choosing the departure time when the capacity realization is known and the SRPAT is fixed. This reflects that typically more information becomes available in the short run (e.g. Chorus et al., 2006). Our assumption of a bottleneck capacity that varies between days for example represents situations where changes in road capacity persist over the entire day, for instance as a consequence of severe incidents, lane closures or adverse weather conditions. Comparable representations of capacity fluctuations in a bottleneck setting have also been used in previous studies (e.g. Lindsey, 1995; Arnott et al., 1996, 1999).

Earlier studies of bottleneck congestion did not distinguish between a long- and short-run dimension of scheduling. However, a recent paper by Peer et al. (2011) provides empirical evidence that short-run and long-run preferences, and as a consequence also the corresponding scheduling choices, may diverge. Their work suggests that drivers plan their routines to avoid congestion, driving a wedge between the LRPAT and SRPAT. Estimating a scheduling model that distinguishes explicitly between the long run and the short run, Peer et al. (2011) furthermore confirm the intuitive notion that the value of travel time is higher in the long run than in the short run, presumably because an incidental time gain can be used less effectively than a structural one. The opposite is true for the values of schedule delays early and late, which may well reflect that scheduling constraints are more binding in the short run than in the long run. They find that the long-run and short-run valuations differ substantially, by factors ranging between 2 to 5. Differences between short-run and long-run shadow prices are also present in the theoretical model introduced in this paper.

In this paper, we do not only characterize the unpriced equilibrium, but also first-best and second-best optima. We show that the first-best optimum can be achieved by levying first-best tolls upon passage of the bottleneck - hence, in the short run - while simultaneously using a long-run pricing instrument to affect the choice of the short-run preferred arrival time.[1] The application of both short-run and long-run pricing instruments may not always be feasible, for instance due to technical or political restrictions. We thus consider also second-best situations where either only short-run or long-run pricing instruments are available. We find that the unpriced equilibrium as well as the first- and

[1] It is quite straightforward to imagine how a short-run toll can be implemented in practice, namely by charging a (time-of-day- and capacity-dependent) toll at the entry of the bottleneck. The practical implementation of a long-run toll, however, is less straightforward, since the SRPAT can usually not be directly observed nor affected by policy makers. As a consequence, in reality, long-run tolls might have to be levied in a more indirect way. One example would be to use financial incentives to shift day-care and school starting times to off-peak hours, possibly resulting in off-peak travel routine choices of the parents. Also, for some groups of people such as public sector employees, the SRPAT can more easily be observed and therefore also be influenced by policy makers through pricing instruments.

second-best optima entail routine arrival times at work (the SRPATs) that are dispersed across drivers.

The welfare implications of dispersed work starting hours have been studied earlier, for instance by Henderson (1981) and Arnott (2007), assuming flow congestion, and by Fosgerau and Small (2010), assuming bottleneck congestion. In these papers, the equilibrium pattern of endogenous preferred arrival times at work is driven by positive agglomeration externalities that increase in the number of individuals who are simultaneously present at work. Our model shows that the consideration of equilibria with dispersed work starting hours does not necessarily require the presence of agglomeration economics, but may also follow from distinguishing between long-run and short-run scheduling decisions.

In the context of our analysis it is natural to define the long run as the time frame where travel routines are chosen. This may be different from other settings, in particular those that involve also residential or employment choices (e.g De Vany and Saving, 1982; Van Ommeren et al., 2000; Van Ommeren and Fosgerau, 2009). Our focus is motivated by our aim to analyze the distinction between long-run and short-run scheduling decisions in the framework of the standard bottleneck model. Using a model structure that is close to the standard formulation of the bottleneck model enables us to compare the equilibrium solutions of the model that distinguishes between long-run and short-run scheduling choices, to the solutions obtained in the standard model.

The structure of the paper is as follows. Section 2 introduces the bottleneck model that distinguishes between long-run and short-run scheduling decisions. Section 3 characterizes the unpriced equilibrium, while Section 4 discusses the first-best and Section 5 the second-best optima. Section 6 concludes. Various mathematical proofs are contained in the appendix of the paper.

2 The Model

2.1 Introduction

Applied to the morning peak hour, the standard bottleneck model assumes that every day a fixed number of N identical commuters travel from home to work. All of them have the same route, which includes the passage of a single bottleneck with a fixed capacity s. The capacity level s therefore denotes the maximum number of vehicles that can pass the bottleneck per unit of time. Drivers pass the bottleneck in the order of their departure time from home. If the departure rate from home exceeds s, a so-called 'vertical' queue grows, meaning that spill-back effects are ignored. Without loss of generality, the free-flow time required to travel from home to work is usually normalized to 0. Consequently, home-work travel times consist only of the queuing time in front of the bottleneck.

When deciding on their optimal departure time, commuters trade off costs of travel delays and costs of schedule delays. The former indicate the costs associated with travel time losses, while the latter are defined as the costs of earliness and lateness of the actual arrival time relative to the preferred arrival time. The costs of a trip through the bottleneck then depend linearly on travel times and schedule delays, whereby schedule delay costs assume a piecewise linear function, which allows the costs attached to being one minute early to differ from the costs that result from arriving one minute late. The unit costs associated with travel delays, schedule delay early and schedule delay late are denoted by α, β and γ, respectively. The unit cost parameters as well as the preferred arrival time are often assumed identical across drivers (e.g. Arnott et al., 1990, 1993).

In the bottleneck model that is introduced in this paper, long-run and short-run scheduling decisions are distinguished. More specifically, long-run decisions reflect choices of travel routines, whereas short-run decisions represent departure time choices. Just as in the standard bottleneck model, drivers trade off travel delay and schedule delay costs both in the long run and in the short run. The distinction requires two adaptations to the standard formulation of the bottleneck model.

First, as argued in the introduction of this paper, a distinction between short-run and long-run scheduling decisions is not useful if days are exact replicas. Therefore, in our model, the bottleneck capacity can assume $i = 1, \ldots, J$ possible discrete values, each of which is realized with probability p_i. The regarding capacity levels are then denoted by s_i. Clearly, it must hold that $\sum_{i=1}^{J} p_i$ equals 1. As in the models of Lindsey (1995) and Arnott et al. (1996, 1999), these capacity levels are day-specific, and therefore do not vary during a given day. We assume that in the long run commuters only know these probabilities, while in the short run (hence, before deciding on their departure time on a specific day) they know the actual realization of the bottleneck capacity.

Second, we distinguish between two different preferred arrival times: the "long-run preferred arrival time" (LRPAT) and the "short-run preferred arrival time" (SRPAT). The LRPAT is exogenously given, and is assumed to be identical across all drivers. The SRPAT, in contrast, is endogenous and may differ across drivers. It represents the preferred arrival routine. We introduce a function $Z(t)$ that describes the cumulative distribution of SRPATs over time of the day. The corresponding density function is denoted by $\dot{Z}(t)$.

The relation between the LRPAT and the SRPAT is established in the long-run model, where drivers choose their SRPAT as a function of their LRPAT. Drivers will choose a SRPAT that differs from their LRPAT if the long-run scheduling costs - due to deviating from the LRPAT- are counterbalanced by lower costs due to shorter travel delays (in the long and/or short run) or schedule delays (in the short run). In reality, this may for instance translate to the situation where a commuter with a LRPAT at 9:00 chooses a routine arrival time at work (his SRPAT) at 7:00, in order to avoid lengthy average travel times, when these are higher at 9:00 than at 7:00. Given the traffic conditions on a specific day, he may choose a departure time that results in an arrival time different from the SRPAT. For example, on a day with low road capacity, she may depart from home such that she arrives at work already at 6:30.

2.2 Short-run scheduling decision

Short-run decisions are analogous to the decisions represented in the standard bottleneck model. Therefore, when we will later on compare the results of the bottleneck model that distinguishes between long-run and short-run scheduling decisions with the results of the standard bottleneck model, we can refer to the results of the latter by using the results presented in this section.

The short-run costs of passing the bottleneck for a driver with a SRPAT equal to t, $C^{SR}(t, s_i)$, consist of travel delay ('queuing') costs, $C_T^{SR}(t, s_i)$, and schedule delay ('scheduling') costs, $C_{SD}^{SR}(t, s_i)$:

$$C^{SR}(t, s_i) = C_T^{SR}(t, s_i) + C_{SD}^{SR}(t, s_i) \tag{1}$$

Note that in contrast to earlier studies where t, besides being the time index, usually denotes the timing of the departure time decisions, t is used here as a short-hand for the SRPAT. We adopt this notation for all cost, price and toll functions for which t serves as an argument. This renders it easier to integrate short-run and long-run scheduling decisions using a common notation. Moreover, we add s_i as function argument - not only to the cost functions but also to departure and arrival rates, and the starting and end time of the queue - in order to emphasize the capacity-dependency of the short-run equilibrium.

While not added as an argument explicitly, short-run costs depend on the cumulative distribution of SRPATs, $Z(t)$; specifically, on the relation between the density function of the SRPATs, $\dot{Z}(t)$, and the bottleneck capacity s_i. So, unless the density of SRPATs, $\dot{Z}(t)$, is smaller than (or equal to) s_i for all time instances t between the earliest SRPAT, t_l, and the latest SRPAT, $t_{l'}$, the equilibrium outcome will entail queuing. In the following analysis, we distinguish these two cases: $\dot{Z}(t) > s_i$ (Case 1) and $\dot{Z}(t) \leq s_i$ (Case 2) (for all $t_l \leq t \leq t_{l'}$). Other cases will only be discussed briefly, as they turn out to be irrelevant in the analysis of our model.

Case 1: Density of SRPATs exceeds s_i for all $t_l \leq t \leq t_{l'}$

If $Z(t)$ is consistently steeper than the cumulative arrivals at work, $A(t, s_i)$, and therefore intersects $A(t, s_i)$ only once (at a moment in time that will be denoted by t^*), Hendrickson and Kocur (1981) showed that, regardless of the exact shape of $Z(t)$, the cumulative departures from home, $D(t, s_i)$, will be such that there is only one time interval where the queue in front of the bottleneck will grow ($[t_q(s_i), t^*]$), and another one where the queue will dissipate ($[t^*, t_{q'}(s_i)]$); $t_q(s_i)$ thus denotes the start of the peak, and $t_{q'}(s_i)$ the end of it. Between these two time instances, the bottleneck operates at its maximum capacity. Hence, all drivers except the first and last one departing experience queuing. The following conditions need to be satisfied in equilibrium:

$$A(t^*, s_i) = Z(t^*) \tag{2a}$$

$$t_{q'}(s_i) - t_q(s_i) = \frac{N}{s_i} \tag{2b}$$

$$D(t_q(s_i), s_i) = A(t_q(s_i), s_i) = Z(t_q(s_i)) = 0 \tag{2c}$$

$$D(t_{q'}(s_i), s_i) = A(t_{q'}(s_i), s_i) = Z(t_{q'}(s_i)) = N \tag{2d}$$

Eq. 2a provides the definition of t^* as the intersection point of $Z(t)$ and $A(t, s_I)$, and Eq. 2b defines the duration of the peak as the ratio between N and s_i. Eq. 2c states that at the time the queue starts to form $t_q(s_i)$, no driver has yet passed the bottleneck, and all drivers have a SRPAT equal or later than $t_q(s_i)$. At the other end, Eq. 2d states that at the time the queue has disappeared, $t_{q'}(s_i)$, all drivers N must have passed the bottleneck, and none of them has a SRPAT later than $t_{q'}(s_i)$.

One of the properties that has been shown to hold in the standard model, and that is therefore also valid in the short-run model here, is that in the case when drivers have different preferred arrival times (and are identical otherwise), the equilibrium order of departure is undetermined (Lindsey, 2004; Smith, 1979; Daganzo, 1985). This is a direct consequence of the linear formulation of the cost function.[2] In all subsequent analyses, we make the assumption that drivers pass the bottleneck in order of increasing SRPAT. Although this equilibrium is not unique, it is equivalent to other equilibria in terms of costs, both in the aggregate and for every driver individually.

Schedule delays for a driver with a SRPAT equal to t are then defined as the difference between t and the actual arrival time. The latter is given by $t_q(s_i) + Z(t)/s_i$, because the bottleneck is active since $t_q(s_i)$ and drivers are assumed to arrive in order of their SRPAT. All drivers with a SRPAT between t_l and t^* arrive early, while all with a SRPAT between t^* and $t_{l'}$ arrive late. Depending on whether a driver arrives early or late, the unit costs of β and γ, respectively, are relevant, and the schedule delay costs, $C^{SR}_{SD}(t, s_i)$, can therefore be expressed as follows:

$$C^{SR}_{SD}(t, s_i) = \begin{cases} \beta \left(t - t_q(s_i) - \frac{Z(t)}{s_i} \right) & \text{if } t_l < t \leq t^* \\ \gamma \left(-t + t_q(s_i) + \frac{Z(t)}{s_i} \right) & \text{if } t^* < t < t_{l'} \end{cases} \tag{3}$$

[2] Due to the linearity of the cost function, drivers are indifferent between arrival times in the interval $[t_q(s_i), t^*]$, and in the interval $[t^*, t_{q'}(s_i)]$.

It can be shown that not only in the case when all drivers share the same preferred arrival time, but also when the preferred arrival times are dispersed in time, the equilibrium departure rate, $\dot{D}(t, s_i)$, is such that the marginal benefits of shifting one's departure time closer to the preferred arrival time are exactly offset by an increase in queuing costs of the same size (Hendrickson and Kocur, 1981). This is a necessary condition for a driver not to have an incentive to marginally adjust the travel moment. The equilibrium departure rates for early and late arrivals, again as a function of the SRPAT t, can thus be expressed as follows:[3]

$$\dot{D}(t, s_i) = \begin{cases} s_i \frac{\alpha}{\alpha-\beta} & \text{if } t_l < t \leq t^* \\ s_i \frac{\alpha}{\alpha+\gamma} & \text{if } t^* < t < t_{l'} \end{cases} \tag{4}$$

Travel times are defined as the difference between arrival and departure times[4]The corresponding cost, $C_T^{SR}(t)$, are obtained by multiplying travel times by parameter α:

$$C_T^{SR}(t, s_i) = \begin{cases} \alpha \left(\frac{\beta}{\alpha} \frac{Z(t)}{s_i} \right) = \beta \frac{Z(t)}{s_i} & \text{if } t_l < t \leq t^* \\ \alpha \left(\frac{\gamma}{\alpha} \frac{N-Z(t)}{s_i} \right) = \gamma \frac{N-Z(t)}{s_i} & \text{if } t^* < t < t_{l'} \end{cases} \tag{5}$$

From the conditions given in Eq. 2b–2d, the equilibrium travel delay and queuing costs (Eqs. 3 and 5) and the fact that the first driver and the last driver must face equal costs in equilibrium[5], we can derive the following equilibrium results for the relative share of drivers who arrive before their SRPAT, θ (hence, $Z(t^*) = \theta N$), as well as the start and the end time of the queue.

$$\theta = \frac{\gamma}{\beta + \gamma}, \quad t_q(s_i) = t^* - \theta \frac{N}{s_i} \quad \text{and} \quad t_{q'}(s_i) = t^* + (1 - \theta) \frac{N}{s_i} \tag{6}$$

As a next step, we can then derive total travel delay and scheduling costs in equilibrium, which we denote by $TC_T^{SR}(s_i)$ and $TC_{SD}^{SR}(s_i)$:

$$TC_T^{SR}(s_i) = \frac{\delta}{2} \frac{N^2}{s_i}, \quad \text{where} \quad \delta = \frac{\beta\gamma}{\beta + \gamma} \tag{7}$$

[3] We assume that $\alpha > \beta$, which is in accordance with empirical findings (e.g. Small, 1982). Without this assumption, cost equality among equal drivers can only be established if a mass departure of drivers takes place at $t_q(s_i)$.

[4] Departure times can be derived by solving the equations $Z(t) = \dot{D}(t, s_i)(t - t_q(s_i))$ (if $t_l < t \leq t^*$) and $Z(t) = N - \dot{D}(t, s_i)(t_{q'(s_i)} - t)$ (if $t^* < t < t_{l'}$) with respect to t. It therefore follows that the departure times for a driver with a SRPAT at t are given by $t_q(s_i) + \frac{\alpha-\beta}{\alpha} \frac{Z(t)}{s_i}$ and $t_{q'}(s_i) - \frac{\alpha+\gamma}{\alpha} \frac{N-Z(t)}{s_i}$. Subtracting them from the respective arrival times $t_q(s_i) + \frac{Z(t)}{s}$ and $t_{q'}(s_i) - \frac{N-Z(t)}{s}$ results in the travel times given in Eq. 5.

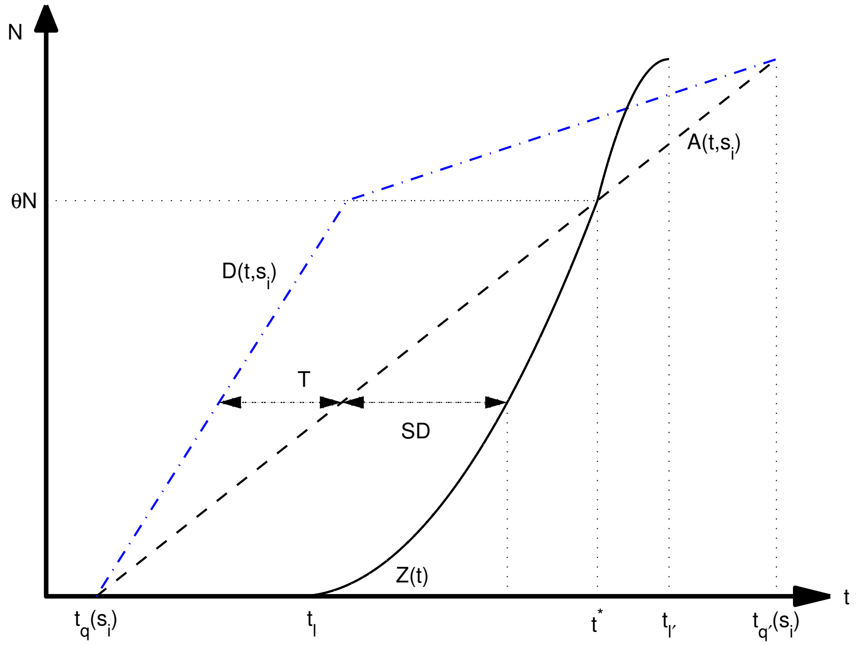
[5] This follows from the rationale that the driver with a SRPAT equal to t^* must be willing to exchange with both the first and the last driver (see Footnote 2).

$$TC_{SD}^{SR}(s_i) = \int_{t_l}^{t_{l'}} C_{SD}^{SR}(t, s_i)\dot{Z}(t)dt \quad \leq \quad \frac{\delta}{2}\frac{N^2}{s_i} \tag{8}$$

We cannot give a closed-form analytical expression for total scheduling costs $TC_{SD}^{SR}(s_i)$, since we have not explicitly defined a distribution of SRPATs, $Z(t)$. However, we can use the setting where all drivers have equal SRPATs as a benchmark. Then, all drivers face equal costs in equilibrium, and total travel delay and scheduling costs are equal. For obvious reasons, scheduling costs are at their maximum in that case. For any setting with dispersed preferred arrival times, total scheduling costs will thus be lower than travel delay costs.

Figure 1 provides an example of a bottleneck where the bottleneck operates at full capacity throughout the peak, showing cumulative departures and arrivals in equilibrium. Travel times are then given by the horizontal difference between cumulative departures and arrivals; and schedule delays by the horizontal difference between the cumulative arrivals and the cumulative distribution of SRPATs, $Z(t)$.

Figure 1: Standard bottleneck model



Case 2: Density of SRPATs is smaller than (equal to) s_i for all $t_l \leq t \leq t_{l'}$

The bottleneck congestion technology implies that total costs become 0 if the density of SRPATs, $\dot{Z}(t)$ is below the bottleneck capacity s_i for all time instances between t_l and $t_{l'}$. Both queuing and scheduling costs are equal to 0 then, as each driver is able to arrive at this SRPAT without queuing. It must therefore hold that:

$$\begin{aligned} D(t, s_i) = A(t, s_i) = Z(t) \\ C_T^{SR}(t, s_i) = C_{SD}^{SR}(t, s_i) = 0 \end{aligned} \tag{9}$$

Other cases

Besides the equilibria that entail congestion throughout the entire peak, or no congestion at all, one can also imagine equilibria where the queue does not start with the first driver but only after some drivers have arrived under uncongested conditions. This is the case if $A(t, s_i)$ and $Z(t)$ intersect multiple times. While the start and end of the queue will change in such a setting, the optimal departure rates for drivers who depart and arrive under congested conditions (see Eq. 4) still remain valid also in this case (e.g. Newell, 1987).

2.3 Long-run scheduling decision

Drivers decide on their travel routine by minimizing overall costs, $EC(t)$.[6] Their long-run choices thus determine the distribution of SRPATs, $Z(t)$. The overall costs consist of long-run costs as well as (equilibrium) short-run costs. In accordance with empirical findings (e.g. Peer et al., 2011), the long-run values of travel time and schedule delay early and late may differ from the corresponding short-run valuations. The long-run values are denoted by $a\alpha, b\beta, c\gamma$, respectively, where a, b, c thus reflect the ratios of long-run and short-run costs.

In the long run, only the probability distribution of capacity realizations is known, rather than a deterministic single capacity realization. Since we assume that drivers perceive the probability distribution of capacities correctly, long-run travel delay costs are a function of expected short-run travel times. More specifically, long-run travel delay costs, $C_T^{LR}(t)$, differ from expected short-run travel delay costs $EC_T^{SR}(t)$ only by parameter a. The overall travel delay costs that determine the choice of the SRPAT, $C_T(t)$ are then equal to:

$$C_T(t) = EC_T^{SR}(t) + C_T^{LR}(t) = (1 + a)EC_T^{SR}(t) \tag{10}$$

[6] Just as in the short-run cost functions, we do not explicitly add $Z(t)$ as a function argument, in order to keep the notation simple.

On the basis of intuition and empirical estimates obtained by Peer et al. (2011) and Tseng et al. (2011), which show that the costs related to one hour of queuing are higher in the long-run model (taking into account overall costs) than in the short-run model, one might expect $\alpha > 0$.

Note that here, as well as in the rest of the paper, we define the expectation operator attached to any capacity-dependent (and hence, short-run) function $f(s_i)$, Ef, as the weighted average across the capacity levels s_i $(i = 1, \dots, J)$, with the weights being given by p_i:

$$Ef = \sum_{i}^{J} p_i f(s_i) \tag{11}$$

Next, we define the costs related to long-run schedule delays, $C_{SD}^{LR}(t)$, as the deviations of the SRPAT t from the LRPAT, evaluated by $b\beta$ or $c\gamma$, depending on whether they concern earliness or lateness (with respect to the LRPAT). For the sake of notational convenience we set the LRPAT, which is assumed to be identical across drivers, at 0. Based on the results obtained by Börjesson (2009), Börjesson et al. (2012), and Peer et al. (2011) and the intuition that delays are less costly if they are known far in advance (and thus allow for adjustments in one's schedule), it is expected that $0 < b < 1$ and $0 < c < 1$.

$$C_{SD}^{LR}(t) = \begin{cases} -tb\beta & \text{if } t_l < t \leq 0 \\ tc\gamma & \text{if } 0 < t < t_{l'} \end{cases} \tag{12}$$

Finally the overall cost function can be stated as follows:

$$EC(t) = C_T(t) + C_{SD}^{SR}(t) + C_{SD}^{LR}(t) \tag{13}$$

2.4 Further assumptions

In order to maintain a simple model structure, we assume in the subsequent analysis of the unpriced equilibrium and the social optima that only two different realizations of capacity levels are possible. The lower capacity state, denoted by s_{min}, occurs with probability p, and the higher state, s_{max}, occurs with probability $1 - p$. Moreover, the analyses below assume that b equals c. Drivers thus attach the same value to long-run schedule delays relative to short-run schedule delays for both earliness and lateness, which is a rather realistic assumption judging by the estimates in Peer et al. (2011). We denote this common scheduling parameter by g. Moreover, we focus on the parameter range of $p < g < 1$, which leads to the most insightful solutions, and is consistent with empirical findings of $g < 1$. Outside this range, mostly corner solutions arise.[7] Finally,

[7] For instance, if g was larger than 1, the social optimum would imply that each driver has a SRPAT equal to his LRPAT. On the other side, if g was smaller than p, the social optimum would entail a density

for reasons that will become clear in the next section of the paper that discusses the unpriced equilibrium (see in particular Footnote 9), the parameters are set such that the following inequality holds: $a < \frac{g}{p} - 1 < a\frac{s_{max}}{s_{min}}$.

Some figures will be added as an illustration of the analytical results in the following sections. These assume the following parameter values: $N = 1000, s_{min} = 10/\text{min}, s_{max} = 20/\text{min}, \alpha = 10$ Euro/h$, \beta = 5$ Euro/h$, \gamma = 15$ Euro/h$, a = 0.4, g = 0.8, p = 0.5$.[8] The long-run costs of travel delay are thus 14 Euro/h and the long-run schedule delay costs 12 Euro/h. The duration of the peak (N/s_i) will then be 100 min in the s_{min} state, and 50 min in the s_{max} state (if the bottleneck operates at its maximum capacity throughout). The short-run unit cost parameters have been chosen such that the usual relation of $\beta < \alpha < \gamma$ holds (e.g. Small, 1982). For the long-run unit cost parameters, the values are specified in a rather conservative way, understating the differences between long-run and short-run values by factors 2–5 that were found by Peer et al. (2011). If the differences were assumed larger in the theoretical model, again corner solutions would be obtained for many instances.

3 Unpriced equilibrium

In the unpriced equilibrium, each driver choses the SRPAT in an attempt to minimize the sum of expected short-run and long-run costs of traveling through the bottleneck, $EC^E(t)$ (Eq. 13). Since drivers are identical in their valuations of travel time and schedule delays, and have the same LRPAT, they must face equal values of $EC^E(t)$. The cost equality condition is therefore satisfied if the derivative of the expected costs in Eq. 13 with respect to the SRPAT t, $\mathrm{d}EC^E(t)/\mathrm{d}t$, equals 0. From the resulting differential equation we can obtain an expression for the equilibrium density of SRPATs, $\dot{Z}^E(t)$ (see Section A.1 for the derivations). We find that in equilibrium, queuing only occurs in the s_{min} state, and is absent in the s_{max} state. The equilibrium density of SRPATs in equilibrium, $\dot{Z}^E(t)$, is then given by:

$$\dot{Z}^E(t) = \frac{g - p}{ap} s_{min} \tag{14}$$

Eq. 14 shows that $\dot{Z}^E(t)$ is constant, implying that the SRPATs are uniformly distributed. Moreover, $\dot{Z}^E(t)$ is proportional to s_{min}, which is a natural result as s_{min} determines short-run scheduling and queuing costs, while these costs are equal to 0 in the s_{max} state (where no queuing takes place). A similar reasoning holds for the finding that $\dot{Z}^E(t)$ is a decreasing function of p: The higher the probability that the s_{min} state occurs, the flatter and therefore closer to s_{min} $\dot{Z}^E(t)$ will be. Moreover, $\dot{Z}^E(t)$ is increasing in

of SRPATs equal to the low capacity state, and therefore no queuing even in the absence of tolling. The underlying argumentation can be found in Section 4.2.

[8] Note that $a < \frac{g}{p} - 1 < a\frac{s_{max}}{s_{min}}$ holds: $0.4 < 0.6 < 0.8$.

g and decreasing in a. Naturally, relatively high long-run scheduling costs (i.e. a high g) lead to a steeper $Z^E(t)$, as deviations from the LRPAT become more costly. Higher long-run travel delay costs (i.e. a high a), on the other hand, render a relatively flat $\dot{Z}^E(t)$ necessary in order to compensate for the high travel delay costs for drivers with a SRPAT close to their LRPAT.

For the cost equality condition to be satisfied, $Z^E(t)$ must then intersect cumulative arrivals at the LRPAT (i.e. $t^* = 0$) both in the s_{min} as well as in the s_{max} state, resulting in θN drivers who have a SRPAT that is earlier than their LRPAT, while the remaining $(1-\theta)N$ drivers have a SRPAT that is later than their LRPAT (Section A.1). This means that drivers with a SRPAT that is earlier than their LRPAT always arrive early (in the s_{min} state) or on time (in the s_{max} state), while all drivers with a SRPAT later than their LRPAT either arrive late (in the s_{min} state) or on time (again in the s_{max} state). This result is closely related to the finding in the standard bottleneck framework that θN drivers arrive before their preferred arrival time (Eq. 6). It follows directly from $Z^E(0) = \theta N$ and the linearity of $Z^E(t)$ that the timing of the earliest SRPAT, t_l^E, and the latest SRPAT, $t_{l'}^E$, must be equal to the inverse of $\dot{Z}^E(t)$ times $-\theta N$ and $(1-\theta)N$, respectively.[9]

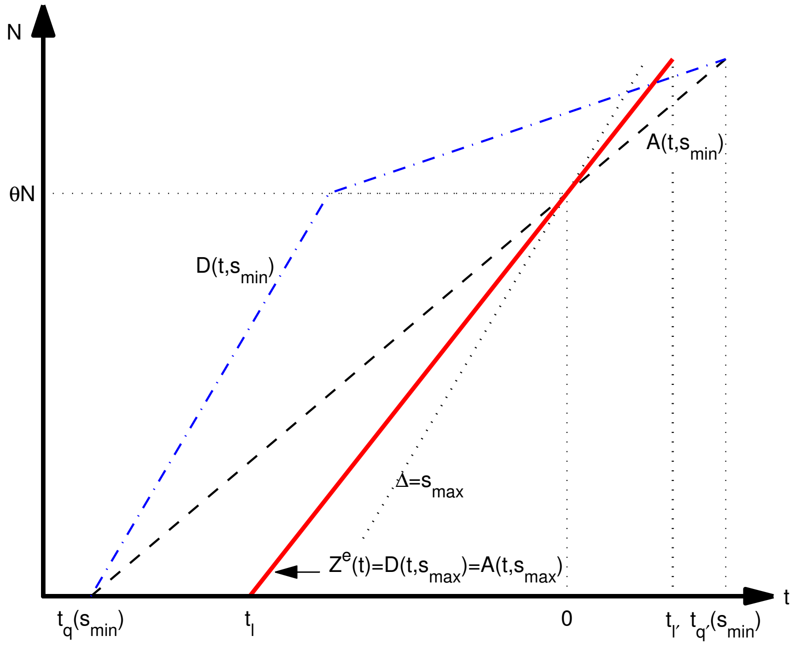
$$t_l^E = -\theta N \frac{ap}{(g-p)s_{min}} \quad \text{and} \quad t_{l'}^E = (1-\theta)N \frac{ap}{(g-p)s_{min}} \tag{15}$$

It is easy to show that drivers do not have an incentive to shift their SRPAT when the equilibrium density of SRPATs as derived above as well as the corresponding t_l^E and $t_{l'}^E$ prevail. For instance, the driver with the earliest SRPAT (who also departs first and hence does not face queuing) has no incentive to choose an earlier SRPAT (a shift size denoted by Δ), as this would cause him to lose $g\beta\Delta$ from moving away from the LRPAT, while gaining only $p\beta\Delta$ for decreasing the short run schedule delay costs. Since we assumed g to be larger than p, losses would prevail. If he moved his SRPAT to a later moment in time, his costs cannot decrease due to the cost equality condition under which $Z^E(t)$ has been derived. Figure 2 shows an example of an equilibrium situation, with queueing in the s_{min} and no queuing in the s_{max} state.

Finally, we can specify the expected total costs faced by the commuters in equilibrium, ETC^E. We can split them up into costs related to overall travel times (both long- and short-run), ETC_T^E, short-run schedule delay costs, $ETC_{SD}^{E,SR}$, and long-run schedule

[9] Note that here as well as in the further analysis, we focus on the case when the parameters of the model are such that $s_{min} < \dot{Z}^E(t) < s_{max}$. It is easy to show that this inequality holds if the parameters are chosen such that $a < \frac{g}{p} - 1 < a\frac{s_{max}}{s_{min}}$.

Figure 2: Unpriced equilibrium



delay costs, $TC_{SD}^{E,LR}$ (see Section A.2 for the analytical derivations):

$$\underbrace{\delta\frac{N^2}{2}p\frac{1+a}{s_{min}}}_{ETC_T^E}+\underbrace{\delta\frac{N^2}{2}\frac{p}{s_{min}}\left(1+a-\frac{ap}{(g-p)s_{min}}\right)}_{ETC_{SD}^{E,SR}}+ \tag{16}$$

$$\underbrace{\delta\frac{N^2}{2}\frac{p}{s_{min}}\frac{ap}{(g-p)s_{min}}}_{TC_{SD}^{E,LR}}=\underbrace{\delta N^2p\frac{1+a}{s_{min}}}_{ETC^E}$$

We find that half of total costs are due to travel delay costs. Also this outcome is closely related to the regarding expression in the standard bottleneck model, for which the same result holds if all drivers have the same preferred arrival time. The regarding total costs are then given by $\delta N^2/(2s_i)$ (see Eqs. 7 and 8). It is straightforward to show that ETC^E converges to the solution obtained in the standard bottleneck framework if a goes to 0, and p goes to 1. The proportionality of the cost function with respect to p can be attributed to the finding that drivers incur costs only if s_{min} is realized, which happens with probability p. The factor $(1+a)$, on the other hand, is due to the additional long-run travel delay costs, which are not present in the standard bottleneck model. The second half of total costs consists of scheduling costs. The relative shares being attributed to short-run and long-run scheduling costs are dependent on the inverse of $\dot{Z}^E(t)$. Clearly, the steeper $Z^E(t)$, the smaller are the long-run schedule delay costs.

4 Social optima

4.1 Introduction

The unpriced equilibrium as derived in the previous section is not an efficient outcome, because it entails queuing when capacity is low, and the average private costs that drivers face for traveling through the bottleneck are only half the marginal social costs that they cause. These average private costs can be derived by dividing total costs in equilibrium (ETC^E, see Eq. 16) by the number of drivers N, while the marginal cost are defined as the derivative of ETC^E with respect to N:

$$\underbrace{\delta N \frac{p}{s_{min}}(1+a)}_{ETC^E/N} < \underbrace{\delta 2N \frac{p}{s_{min}}(1+a)}_{\mathrm{d}ETC^E/\mathrm{d}N} \tag{17}$$

The difference between marginal social costs and private costs is referred to as marginal external congestion cost, which arises because drivers do not internalize the costs they impose on other drivers by contributing to overall congestion. The social (first-best) optimum can be achieved if all drivers face the marginal social costs that result from their scheduling decisions. Pigou (1920) was the first to show that the social optimum can be decentralized by applying tolls that are equal to the marginal external congestion cost. In the standard bottleneck model (with $Z(t) > s_i$ for all $t_l \leq t \leq t_{l'}$), this can be achieved by levying a time-varying toll, which follows exactly the pattern of travel delay costs in the unpriced equilibrium. The first-best toll is thus 0 for the first and the last driver, and largest ($\beta\theta N/s_i$) for the driver who arrives exactly at his preferred arrival time (the θNth driver). In the standard bottleneck model, drivers will then arrive at the same time as in the unpriced equilibrium; however, without facing any queuing delay, while the bottleneck will operate at its capacity throughout the peak. Travel delays are thus always a deadweight loss in the bottleneck model, as they can be reduced without increasing scheduling costs.

In our model that distinguishes between long-run and short-run scheduling decisions, two types of pricing instruments are conceivable: long-run tolls and short-run tolls. Both can vary freely over the time of the day. The former would vary with the choice of the SRPAT, and the latter with the choice of departure time. We will consider the first-best situation where both instruments are available, as well as second-best optima, where only one of these two pricing instruments is available, for instance for political or technical reasons.

Short-run tolls are levied at the bottleneck. They depend on both the realized bottleneck capacity as well as the departure time chosen by a specific driver (for the realized capacity state), and are denoted by $\tau^{SR}(t, s_i)$. We differentiate between two different forms of short-run pricing. The first one are the conventional *first-best short-run tolls*, which are by definition equal to the marginal external costs. These tolls are therefore only relevant if congestion would occur without the application of the toll; otherwise they are equal to 0. First-best short-run tolls can be determined without the regulator knowing about the underlying long-run choice process that gives rise to the distribution of SRPATs. The reason is that just as queuing costs, first-best short-run tolls are independent of $Z(t)$ as long as it holds that $Z(t) > s_i$ for all time instances between t_l and $t_{l'}$.

In addition to the first-best short-run tolls, we define a second form of short-run tolling, which we refer to as *complementary short-run tolls.* We use this label to refer to tolls that are levied on days where the capacity of the bottleneck is high enough such that no queuing would occur in the absence of tolls. We will show that under specific conditions it is welfare-improving to levy such tolls in addition to first-best short-run tolls, since they can be used to affect the long-run choice of the SRPAT such that schedule delay costs are minimized. To set the complementary short-run tolls optimally, the regulator must therefore be aware of the long-run choice problem of the drivers.

In contrast to the short-run tolls, long-run tolls are independent of the bottleneck capacity, and only depend on a driver's SRPAT. They are denoted by $\tau^{LR}(t)$. The interpretation behind such a long-run pricing instrument is that the regulator levies a tax on the choice of the routine work starting time (i.e. the SRPAT).[10]

Finally, the expected price function $EP(t)$ can be defined. For the first-best optimum it consists of the overall costs $EC(t)$ (Eq. 13) as well as expected short-run and long-run tolls:

$$EP(t) = EC(t) + E\tau^{SR}(t) + \tau^{LR}(t) \tag{18}$$

This price function can be adjusted easily for the second-best optima, leaving out one of the pricing instruments. Similar to the cost equality condition in the unpriced equilibrium and for the same reasons (drivers share a common LRPAT and attach identical values to reductions in travel delays and schedule delays), expected prices must be equal across drivers if first- and second-best optima are decentralized. We will show that the social optima derived in the following sections of the paper again imply a uniform distribution of SRPATs. Consequently, the relative share of drivers who have a SRPAT earlier than their LRPAT must again be equal to θ, and the timing of the earliest SRPAT and the latest SRPAT are given by $t_l = -\theta N/\dot{Z}(t)$ and $t_{l'} = (1-\theta)N/\dot{Z}(t)$, respectively.

[10] We do not worry here about the realism of such a tax (see also Footnote 1); what is of interest to us is the question of how it would be set if it were available.

4.2 First-best optimum

Specification of the optimum

In this section, we will first characterize the optimum, and then derive the tolls required to achieve it. Since it is feasible to levy first-best short-run tolls, the optimum entails no queuing. It is easy to see that scheduling costs are minimized if the density of SRPATs, $\dot{Z}^F(t)$, is equal to the higher capacity state s_{max}. Starting from that, a $\dot{Z}^F(t)$ below s_{max} would induce a decrease in aggregate short-run schedule delays (evaluated at $p\beta$ and $p\gamma$ per unit of time adjustment of the SRPAT)[11] and an increase in long-run schedule delays (evaluated at $g\beta$ and $g\gamma$). These changes in short-run and long-run aggregate schedule delays are equally big, but since we assumed that $p < g$, the value of the decrease in short-run scheduling costs does not outweigh the increase in long-run scheduling costs. At the same time, an increase in $\dot{Z}^F(t)$ above s_{max} would induce a decrease in aggregate long-run schedule delays (again evaluated at $g\beta$ and $g\gamma$), and an increase in aggregate short-run schedule delays of the same size (now evaluated at β, because short-run schedule delays would then result for both capacity states). Since $g < 1$ is assumed, the decrease in long-run scheduling costs does not outweigh the increase in short-run scheduling costs. The socially optimal density of SRPATs is therefore equal to s_{max}, and therefore *higher* than in the unpriced equilibrium:

$$\dot{Z}^F(t) = s_{max} \tag{19}$$

A higher concentration of SRPATs in the optimum than in the no-toll equilibrium may seem counterintuitive, given the standard notion that optimal pricing would lead to a more dispersed traffic pattern over the day. The intuition behind the results is that the stronger concentration of SRPATs is combined with an elimination of queuing. The optimal concentration of SRPATs therefore results from a trade-off between scheduling costs only; the free-market concentration adds a desire to avoid the peak because of travel delay costs on top of these schedule delay components.

The expected total (social) costs corresponding to the first-best optimum, ETC^F, consisting of short-run and long-run scheduling costs, $ETC_{SD}^{F,SR}$ and $TC_{SD}^{F,LR}$, are then given in Eq. 20. They can be derived in a similar way as the total costs in the unpriced equilibrium, the derivations of which are shown in Section A.2, with the only difference being that in the unpriced equilibrium queuing costs are 0 and $\dot{Z}(t)$ is equal to s_{max} (instead of $\dot{Z}^E(t)$).[12]

[11] This is true if $s_{min} \leq \dot{Z}^F(t)$. But any decrease of $\dot{Z}^F(t)$ below s_{min} is inefficient for obvious reasons, inducing unnecessarily high long-run scheduling costs, while not decreasing short-run schedule delay any further.

[12] It can be easily shown that the social optimum entails lower costs than the unpriced equilibrium: $ETC^F < ETC^E \Leftrightarrow \frac{g}{p} - 1 < (1+2a)\frac{s_{max}}{s_{min}}$. Since we assumed $\frac{g}{p} - 1 < a\frac{s_{max}}{s_{min}}$ (see Footnote 9) and $a > 0$, $\frac{g}{p} - 1 < (1+2a)\frac{s_{max}}{s_{min}}$ holds too.

$$\underbrace{\delta\frac{N^2}{2}p\left(\frac{1}{s_{min}}-\frac{1}{s_{max}}\right)}_{ETC_{SD}^{F,SR}}+\underbrace{\delta\frac{N^2}{2}\frac{g}{s_{max}}}_{TC_{SD}^{F,LR}}=\underbrace{\delta\frac{N^2}{2}\left(\frac{p}{s_{min}}+\frac{g-p}{s_{max}}\right)}_{ETC^F} \tag{20}$$

Short-run scheduling costs are increasing in p. This is an intuitive outcome since short-run scheduling costs only arise in the s_{min} state, which occurs with probability p. Moreover, short-run scheduling costs decrease in s_{min} and increase in s_{max}. This outcome is not unexpected either. If s_{min} increases and thus becomes closer to s_{max}, each driver is able to arrive closer to his SRPAT (distributed with density s_{max}), decreasing short-run scheduling costs. An increase in s_{max} leads to exactly the opposite result. It can furthermore be shown that the costs derived for the first-best optimum again approach the corresponding costs for the standard bottleneck case, if parameters are set accordingly. So, if the long-run scheduling parameter g is assumed equal to 0, and $\dot{Z}^F(t)$ is set equal to infinity (indicating that all drivers have equal SRPATs), long-run scheduling costs approach 0 and total scheduling costs approach the scheduling costs found in the standard model for the case that all drivers have an identical preferred arrival time: $\delta N^2/2s_i$ (Eq. 8).

Total (social) costs in the first-best case are less than half of total costs in the unpriced equilibrium.[13] Besides the elimination of both short-run and long-run queuing costs, also the sum of short and long-run scheduling costs is lower in the first-best optimum than in the unpriced equilibrium. The reason for this result, which is different from the solution found for the standard bottleneck case (where scheduling costs are equal in the unpriced and the first-best optimum), is that the dispersion of SRPATs in the unpriced equilibrium is higher than socially optimal ($\dot{Z}^E(t) < \dot{Z}^F(t)$).

Tolls

If s_{max} applies, first-best short-run tolls will be equal to 0 for all time instances. Since $Z^F(t)$ is equal to s_{max}, also without toll no queuing would occur in this case. The first-best short-run tolls for the s_{min} state are set such that the queuing that occurs under unpriced conditions is eliminated. As in the standard bottleneck model, this entails that the tolls are equal to short-run travel delay costs (Eq. 5). Consequently, the departure rate becomes equal to the capacity of the bottleneck (in this case s_{min}) and all queuing

[13] Given that the parameter values are chosen such that $(g-p)/(ap) < s_{max}/s_{min}$ (see Footnote 9 for the underlying rationale).

disappears:

$$\tau^{F,SR}(t, s_{min}) = \begin{cases} \beta \frac{Z^F(t)}{s_{min}} = \beta \left(\frac{\theta N}{s_{min}} + t \frac{s_{max}}{s_{min}} \right) & \text{if } t_l < t \leq 0 \\ \gamma \frac{N - Z^F(t)}{s_{min}} = \gamma \left(\frac{(1-\theta)N}{s_{min}} - t \frac{s_{max}}{s_{min}} \right) & \text{if } 0 < t < t_{l'} \end{cases} \tag{21}$$
$$\tau^{F,SR}(t, s_{max}) = 0$$

However, if only first-best short-run tolls were levied, drivers would not choose for the socially optimal density of SRPATs, i.e. $\dot{Z}^F(t) = s_{max}$, since the expected price of traveling through the bottleneck would then differ across drivers, rendering the equilibrium under tolls as in Eq. 21 inefficient (see Section A.3 for the derivations). In particular, if $\dot{Z}^F(t)$ were equal to s_{max}, the driver with the SRPAT equal to LRPAT would face the lowest expected price, while the drivers with the earliest and latest SRPAT, respectively, would face the highest one. A long-run toll is thus applied in addition to the short-run tolls to bridge this gap, and hence to reach full efficiency. The optimal long-run toll, $\tau^{F,LR}(t)$, assumes the following shape (note that g is assumed to be larger than p, meaning that $\tau^{F,LR}(t)$ will always be positive):

$$\tau^{F,LR}(t) = \begin{cases} (g-p)\beta \frac{Z^F(t)}{s_{max}} = (g-p)\beta \left(\frac{\theta N}{s_{max}} + t \right) & \text{if } t_l < t \leq 0 \\ (g-p)\gamma \frac{N - Z^F(t)}{s_{max}} = (g-p)\gamma \left(\frac{(1-\theta)N}{s_{max}} - t \right) & \text{if } 0 < t < t_{l'} \end{cases} \tag{22}$$

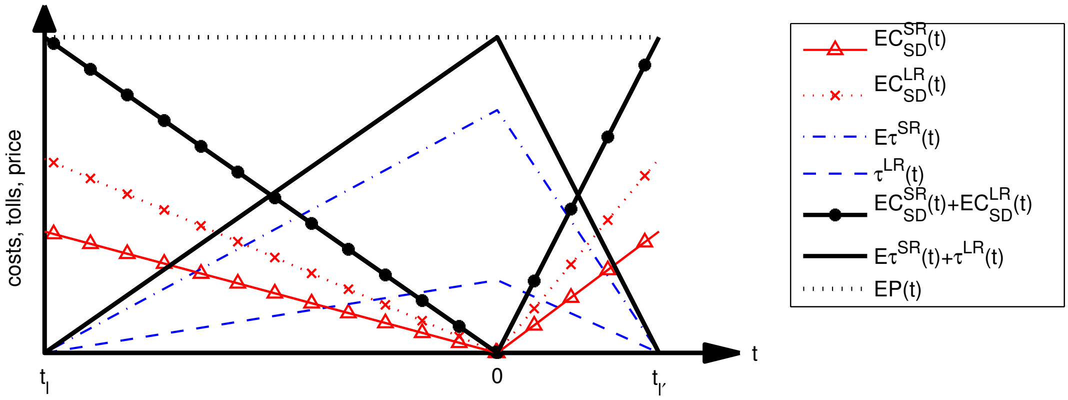
Figure 3 provides the graphical intuition for why both short-run and long-run pricing instruments are required to reach the full optimum. Each instrument by itself is insufficient to equalize the price across across drivers. Only if both of them are used, the social optimum can be decentralized as the sum of scheduling costs and tolls is equal for all drivers, both in the short and the long run.

5 Second-best optima

As discussed above, we consider two second-best optima, both of which are characterized by the availability of only one type of pricing instrument: either a long-run or a short-run pricing instrument. In contrast to other possible second-best situations that are frequently considered in the literature, we assume that the available instrument is not restricted in its form as long as its short or long-run character, respectively, is not altered (e.g. the long-run toll cannot become capacity-specific).[14]

[14] An overview of alternative second-best optima, for instance involving the restriction that the toll cannot be varied freely over time of the day, can be found in Small and Verhoef (2007, pp. 137–148).

Figure 3: First-best tolls and schedule delay costs



5.1 Short-run toll only

We first consider the second-best situation when only short-run tolls are available. Again these can be used to fully eliminate queuing. As we argued for the first-best optimum, the remaining costs, the sum of (expected) short-run and long-run scheduling costs, is minimized if the SRPATs are distributed with density s_{max} (Eq. 19). We will show that $Z^S = s_{max}$ can be achieved by introducing a complementary short-run toll in the s_{max} state, leading to the same welfare level as in the first-best optimum (Eq. 20):

$$\dot{Z}^S(t) = \dot{Z}^F(t) = s_{max} \quad \text{and} \quad ETC^S = ECT^F \tag{23}$$

Tolls

As in the first-best optimum, first-best short-run tolls are levied in the s_{min} state (Eq. 21). Moreover, a complementary short-run toll is introduced in the s_{max} state in order to affect the choice of the SRPATs such that $Z^S(t)$ becomes equal to s_{max}, maximizing social welfare. The complementary short-run toll must be set such that the expected price in this second-best situation becomes equal to the expected price that drivers face in the first-best optimum. This can be attained by replacing the long-run toll of the first-best equilibrium by an appropriate combination of short-run tolls. Since the short-run toll in the state with s_{min} should be set exactly such that it eliminates queuing while keeping the departure rate at s_{min}, we can only use the toll in the state with s_{max} for this purpose. Because the s_{max} state occurs with probability $1-p$, $\tau^S(t, s_{max})$ should be equal to the contribution of the long-run toll in the first-best optimum, $\tau^{F,LR}(t)$, divided by $1-p$:

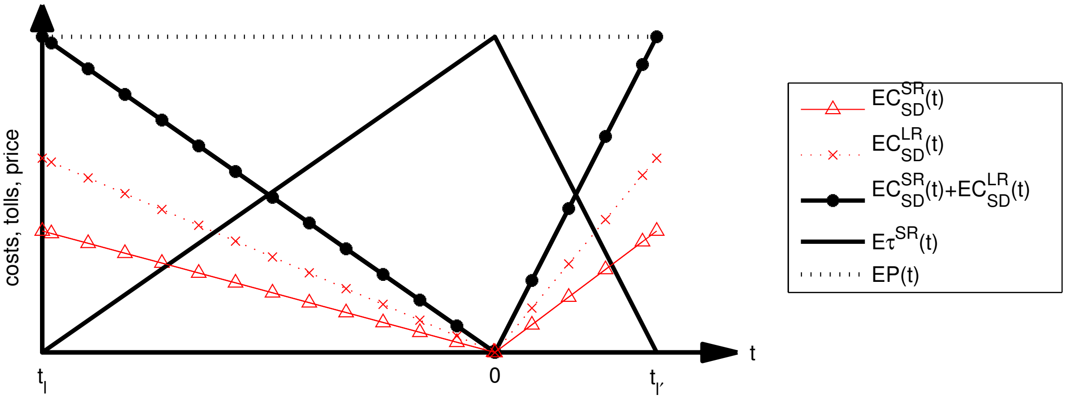
$$\tau^{S}(t, s_{min}) = \tau^{F,SR}(t, s_{min}) \tag{24}$$

$$\tau^{S}(t, s_{max}) = \frac{1}{1-p}\tau^{F,LR}(t) \tag{25}$$

Recalling from Eq. 22 that the slopes of $\tau^{F,LR}$ are $(g-p)\beta$ and $-(g-p)\gamma$, the toll in Eq. 25 will be consistent with the short-run optimum with s_{max} of every driver arriving at her SRPAT, as long as $g < 1$. This is true by assumption, reflecting that the unit cost of schedule delay cannot be smaller in the short run, when there is less flexibility, than in the long run.

Figure 4 gives a graphical overview of the tolls and schedule delay costs in this second-best optimum. Long-run and short-run schedule delay costs are the same as in Figure 3. This holds true also for the sum of schedule delays and tolls. However, unlike in Figure 3, the sum of tolls consists only of the weighted average of short-run tolls rather than both long-run and short-run tolls.

Figure 4: Tolls and schedule delays if only short-run tolls are available



5.2 Long-run toll only

Specification of the optimum

If long-run tolls are the only pricing instrument available to the regulator, queuing can only then be eliminated fully, if the density of SRPATs, $\dot{Z}(t)^L$, is equal or smaller than s_{min}. Also, short-run scheduling costs are then equal to 0, as each driver can arrive at his SRPAT. However, the downside is that this low density of SRPATs induces high total long-run scheduling costs. A higher $\dot{Z}(t)^L$, on the other hand, decreases long-run scheduling costs, but leads to queuing, and to higher short-run scheduling costs. In fact, we find that for the parameter ranges under consideration, these trade-offs yield two possible local second-best optima. The parameter values determine which of these is the most efficient one. The first optimum has $\dot{Z}(t)^L$ equal to s_{min} (Case 1), while the second one has it equal to s_{max} (Case 2). Both possible optima are corner solutions, with the density of SRPATs in the unpriced equilibrium, $\dot{Z}^E(t)$, being located between.

The existence of two local optima arises from the discontinuity in total travel delay cost at $\dot{Z}(t) = s_{min}$, where all queuing disappears while travel delay cost in state s_{min} would be independent of $\dot{Z}(t)$ as long as it exceeds s_{min}. For any constant $\dot{Z}(t)$ between s_{min} and s_{max}, it is easy to see that total schedule delay cost increase in $\dot{Z}(t)$, as they decrease in $N/\dot{Z}(t)$ by a marginal amount $N\frac{1}{2}(g-p)(\theta\beta + (1-\theta)\gamma)$. This suggests $\dot{Z}(t) = s_{max}$ would be optimal. But it is the said discontinuity that may make $\dot{Z}(t) = s_{min}$ optimal as well.

Case 1: $\dot{Z}(t)^L$ equals s_{min}

In this case, drivers do not face any queuing nor short-run delays. Total costs TC^L thus consist only of total long-run schedule delay costs TC^L_{SD} that arise for $\dot{Z}(t)^L = s_{min}$. These can be obtained by integrating long-run scheduling costs (Eq. 12) across all drivers, starting from the driver with the earliest SRPAT, $t_l = -\theta N/s_{min}$, and ending at the driver with the latest $t_{l'} = (1-\theta)N/s_{min}$.

$$TC^L_{SD} = TC^L = \delta\frac{N^2}{2}\frac{g}{s_{min}} \tag{26}$$

Case 2: $\dot{Z}(t)^L$ equals s_{max}

If $\dot{Z}(t)^L$ is equal to s_{max}, drivers face queuing as well as short-run schedule delays if s_{min} but not not if s_{max} is realized. Expected total costs, ETC^L, are thus composed of expected (short-run and long-run) queuing costs, ETC^L_T, short-run schedule delay costs, $ETC^{L,SR}_{SD}$, and long-run schedule delay costs $TC^{L,LR}_{SD}$. Note that the expected total costs differ from the total costs in the first-best optimum (Eq. 20) only by the queuing costs, which are absent in the first-best optimum:

$$\underbrace{\delta\frac{N^2}{2}p\frac{(1+a)}{s_{min}}}_{ETC_T^L}+\underbrace{\delta\frac{N^2}{2}p\left(\frac{1}{s_{min}}-\frac{1}{s_{max}}\right)}_{ETC_{SD}^{L,SR}}+ \tag{27}$$

$$\underbrace{\delta\frac{N^2}{2}\frac{g}{s_{max}}}_{TC_{SD}^{L,LR}}=\underbrace{\delta\frac{N^2}{2}\left(\frac{g-p}{s_{max}}+p\frac{2+a}{s_{min}}\right)}_{ETC^L}$$

Clearly, if the parameters have been chosen such that TC^L is lower for Case 1, the second-best optimum entails that $\dot{Z}(t)^L$ equals s_{min}. In the opposite case, hence when total costs are lower for Case 2, $\dot{Z}(t)^L = s_{max}$ applies.

From the comparison of cases we find that Eq. 26 is smaller than Eq. 27 so that the long-run policy seems to eliminate queuing when a is sufficiently large (i.e. the long-run costs of travel delays should be sufficiently high), g sufficiently small (i.e. the long-run schedule delay costs should be sufficiently low), and p sufficiently large. The exact condition is $\frac{g}{p} < \frac{(2+a)s_{max}-s_{min}}{s_{max}-s_{min}}$.

Tolls

Finally, we can derive the long-run pricing instruments required to achieve the second-best distribution of SRPATs.

Case 1: $\dot{Z}(t)^L$ equals s_{min}

If $\dot{Z}(t)^L$ equals s_{min}, drivers only face long-run schedule delay costs. The corresponding toll $\tau^L(t)$, which ensures that all drivers face equal (expected) prices, must therefore be highest for the driver who has a SRPAT equal to the LRPAT, and thus long-run schedule delay costs of 0. They decrease linearly towards the first and the last driver at a rate equal to the increase in long-run schedule delay costs: $g\beta$ and $g\gamma$, respectively.

$$\tau^L(t)=\begin{cases} g\beta\frac{Z^L(t)}{s_{min}}=g\beta\left(\frac{\theta N}{s_{min}}+t\right) & \text{if } t_l < t \leq 0 \\ g\gamma\frac{N-Z^L(t)}{s_{min}}=g\gamma\left(\frac{(1-\theta)N}{s_{min}}-t\right) & \text{if } 0 < t < t_{l'} \end{cases} \tag{28}$$

Case 2: $\dot{Z}(t)^L$ equals s_{max}

For the case that $\dot{Z}(t)^L$ is equal to s_{max}, the expected costs, $EC(t)$, are highest for the driver who has a SRPAT equal to the LRPAT. The corresponding toll, $\tau^L(t)$, which again ensures that all drivers face equal prices, is therefore *lowest* for this driver. Note that this unusual result of a toll that is highest for the drivers that depart first and last, respectively, is driven by the fact that in the model that distinguishes long-run and short-run scheduling decisions, long-run travel delay costs are evaluated at $a\alpha$ rather than α, rendering it relatively more costly to have long travel times (and hence to have a SRPAT close to the LRPAT). The tolls can be derived by taking the derivative of the cost function, based on the assumption that $\dot{Z}(t)^L$ equals s_{max} (see Eq. 31 in Section A.1). The toll that provides for price equality among drivers must then be equal to (-1) times the derivative:

$$\tau^L(t) = \begin{cases} t\beta \left(g - p - \frac{aps_{max}}{s_{min}}\right) & \text{if } t_l < t \leq 0 \\ t\gamma \left(\frac{aps_{max}}{s_{min}} - g + p\right) & \text{if } 0 < t < t_{l'} \end{cases} \tag{29}$$

Note that from the assumption that $Z^E(t)$ is smaller than s_{max} and hence $g - p < aps_{max}/s_{min}$ (see Footnote 9 for an explanation), it follows that $\tau^L(t)$ is always positive for all $t_l \leq t \leq t_{l'}$.

6 Conclusions

In this paper, we develop a bottleneck model that distinguishes between long-run decisions on travel routines and short-run decisions on departure times, with an application to the morning commute. The bottleneck capacity varies between days, and can either assume a high or low capacity state. We assume that in the long run only the distribution of capacities is known by the drivers, whereas in the short run they are informed about the exact realization of the bottleneck capacity on a specific day. Our model incorporates the intuitive notion that, in the face of congestion, people may change their schedules such that the desired arrival time at work deviates from what would be the most desired moment if congestion would not exist.

We show that in the unpriced equilibrium, routine arrival times at work, which we refer to as short-run preferred arrival times (SRPATs) and which are chosen by the drivers in the long run, are uniformly distributed in time, and therefore different from the long-run preferred arrival time (LRPAT), which is identical for all drivers by assumption. Congestion occurs only in the low capacity state, whereas it is absent in the high capacity state.

We also characterize first- and second-best optima, the latter being defined by a limited availability of pricing instruments. We examine how these can be decentralized by applying short-run and long-run tolls. While short-run tolls are used to affect departure time choices, long-run instruments are used to affect the choice of the routine arrival time at work. Both instruments have in common that they can vary by time of the day. However, while short-run tolls depend depend on the bottleneck capacity realized on a specific day, long-run tolls are capacity-independent.

We show that just as in the unpriced equilibrium, the first-best optimum implies a uniform distribution of SRPATs. However, the extent of dispersion is lower than in the unpriced equilibrium. This is surprising, as conventional wisdom tells us that a greater dispersion of desired arrival times (work start times), would be desirable if congestion exists. The first-best optimum can be reached by simultaneously applying first-best short-run and long-run. First-best short-run tolls as standalone pricing instrument are thus insufficient for reaching the socially efficient outcome. We find that the same level of welfare as in the first-best optimum can be attained if only short-run tolls are feasible. However, this second-best situation requires that, in addition to the first-best short-run toll that is applied in the low-capacity state, tolls are levied also on days when the high-capacity state is realized, and thus on days when - even without tolling - no congestion would occur; we refer to these tolls as complementary short-run tolls. Moreover, we investigate the case when only long-run tolls are feasible, and find that the social optimum can no longer be reached under this restriction.

Also, in the second-best optimum, it may be true that it is desirable to achieve a greater rather than smaller concentration of desired arrival times. The intuition is that a marginal change in the concentration of desired arrival times usually does not reduce travel delay costs (except for the discontinuity where all queues suddenly disappear - in our model, for a density of desired arrival times equal to the lowest capacity s_{min}). For higher densities, travel cost fall if that density is increased, which is due to the benefit from having SRPATs closer to the LRPAT exceeding the probability-weighted short-run schedule delay costs (if the ratio of long-run and short-run schedule delay values exceeds the probability that the lower capacity state occurs). This long-run schedule delay cost, associated with changing daily schedules and desired arrival times in the face of congestion, is not accounted for in typical analyses that propose spreading of work start times.

In this paper, we focus on developing a model that maintains a structure similar to the one of the standard bottleneck model, in particular to the version established by Arnott et al. (1990). In follow-up research we will investigate whether the main results of this paper still hold if an alternative congestion technology, in particular dynamic flow congestion, is assumed. Moreover, future work might focus on relaxing the rather restrictive assumptions on the distribution of bottleneck capacities and the extent of information available to drivers that are used in this paper. On a more general level, future research may also focus on alternative definitions of the long run. For instance, it would be interesting to extend the model such that it captures also decisions that concern the even longer run such as locational and job choices.

A Proofs

A.1 Derivation of $\dot{Z}^E(t)$

As stated in Section 2, drivers choose their SRPAT t by minimizing expected costs $EC^E(t)$. Since drivers are identical regarding their LRPAT and their valuations of time and schedule delays, costs must be equal across drivers. Clearly, the costs depend on whether queuing takes place in both capacity states, or only in the s_{min} state.[15] We find that in equilibrium the latter is true, and queuing is therefore absent in the s_{max} state.[16] The costs function, $EC^E(t)$, can then be determined using the results obtained for queuing costs (Eqs. 5 and 10), short-run schedule delay costs (Eq. 3) and long-run schedule delay costs (Eq. 12). We first derive $\dot{Z}^E(t)$ for the case of a driver who faces schedule delays early both in the short run (i.e. in the s_{min} state) and in the long run. Later we will argue, that in equilibrium a driver will either face earliness both in the short and the long run, or lateness both in the short and the long run. Hence, no combinations of earliness in one time dimension and lateness in the other time dimension are part of the equilibrium solution. The expected costs for earliness in both time dimensions are then given by:

$$EC^E(t) = \underbrace{(1+a)p\beta\frac{Z^E(t)}{s_{min}}}_{EC_T(t)} + \underbrace{p\beta\left(t - t_q(s_{min}) - \frac{Z^E(t)}{s_{min}}\right)}_{EC_{SD}^{SR}(t)} + \underbrace{g\beta(-t)}_{C_{SD}^{LR}(t)} \tag{30}$$

The equilibrium starting time of the peak in the s_{min} state, $t_q(s_{min}) = t^* - \theta N/s_{min}$ (Eq. 6), is a function of $Z^E(t)$, since t^* defines the moment when $A(t, s_{min})$ and $Z^E(t)$ intersect. So, an explicit expression for $t_q(s_{min})$ in Eq. 30, as a function of $Z^E(t)$, is only feasible if the functional form of $Z^E(t)$ is known. Given that the cost function is linear, it is a natural guess that linearity would also hold for $Z^E(t)$, and $\dot{Z}^E(t)$ would thus be a constant. If that is the case, t^* must be equal to the LRPAT(i.e. 0). This can most easily be demonstrated by comparing the costs of the driver with the earliest and the one with the latest SRPAT. Both of them will face equal short-run scheduling costs as

[15] It is straightforward to show that no queuing in either state cannot be an equilibrium solution. In the absence of queuing costs, all drivers would have an incentive to minimize their long-run scheduling costs by choosing a SRPAT equal to their LRPAT, and then depart at their SRPAT=LRPAT, in turn, inducing queuing.

[16] If queuing occurred in both states, the resulting equilibrium density of SRPATs, $\dot{Z}^E(t)$, that is consistent with the cost equality condition shows to be negative ($-\frac{(1-d)s_{min}s_{max}}{a(p(s_{max}-s_{min})+s_{min})} < 0$). Since for obvious reasons $\dot{Z}^E(t)$ cannot be negative, we discard this solution, and focus on the case when queuing only occurs in the s_{min} state. $\dot{Z}^E(t)$ for the case of queuing in both capacity states can be derived by performing the same computations as for the case that queuing occurs only in the s_{min} state, but then adding queuing and short-run schedule delay costs for the s_{max} state.

a consequence of behaving optimally in their short-run scheduling problem. Since they both do not face queuing costs, the only costs they face besides the short-run scheduling costs are long-run scheduling costs, which thus have to be equal across these two drivers. Given the assumption that $Z^E(t)$ is linear as well as our assumption that $g := b = c$, this is only the case if $t^* = 0$ and $Z(0) = \theta N$. We can thus substitute $-\theta N/s_{min}$ for $t_q(s_{min})$ in Eq. 30 and take the derivative with respect to t. Setting the derivative equal to 0, the equilibrium density of SRPATs (Eq. 14) can be derived:

$$\begin{aligned}\frac{\mathrm{d}EC^E(t)}{\mathrm{d}t} =& p\beta\frac{\dot{Z}^E(t)}{s_{min}} + p\beta\left(1 - \frac{\dot{Z}^E(t)}{s_{min}}\right) + ap\beta\frac{\dot{Z}^E(t)}{s_{min}} + -g\beta \\ =& \beta\left(p - g + ap\frac{\dot{Z}^E(t)}{s_{min}}\right) = 0 \Rightarrow \dot{Z}^E(t) = \frac{g-p}{ap}s_{min}\end{aligned} \tag{31}$$

Indeed it shows that $\dot{Z}^E(t)$ is a constant, and $Z^E(t)$ therefore linear. The same $Z^E(t)$ as given in Eq. 31 is obtained if the cost function is defined such that it implies lateness both in the short as well as in the long run. The SRPATs are therefore uniformly distributed in equilibrium.

A.2 Derivation of ETC^E

The total expected costs in the unpriced equilibrium, ETC^E, consist of the sum of (expected) short-run and long-run queuing costs, ETC_T^E, as well as l(expected) short-run and long-run scheduling costs, denoted by $ETC_{SD}^{E,SR}$ and $TC_{SD}^{E,LR}$, respectively. Each of these cost elements can be derived by integrating the corresponding driver- (or more precisely, SRPAT-) specific costs across all drivers, starting from the driver with the earliest SRPAT, $t_{'}^E$, to the driver with the latest SRPAT, $t_{l'}^E$ (Eq. 15). The density of SRPATs, $\dot{Z}^E(t)$ has been derived in Eq. 14.

Based on the definitions of the expected short-run and long-run queuing costs in equilibrium (see Eqs. 5 and 10, respectively) and the finding that no queuing occurs if the s_{max} state is realized, total expected queuing costs can be derived in the following way:

$$\begin{aligned}ETC_T^E =& (1+a)\int_{t_l^E}^{t_{l'}^E} EC_T^{SR}(t)\dot{Z}^E(t)\mathrm{d}t = \\ & (1+a)p\left(\int_{t_l^E}^{t_{l'}^E} C_T^{SR}(t, s_{min})\mathrm{d}t\right) = \delta\frac{N^2}{2}p\frac{1+a}{s_{min}}\end{aligned} \tag{32}$$

Similarly, expected short-run schedule delay costs (Eq. 3) can be computed, again taking into account only delay costs in the s_{min} state:

$$ETC_{SD}^{E,SR} = \int_{t_l^E}^{t_{l'}^E} EC_{SD}^{SR}(t)\dot{Z}^E(t)\mathrm{d}t = \tag{33}$$
$$p\int_{t_l^E}^{t_{l'}^E} C_{SD}^{SR}(t, s_{min})\dot{Z}^E(t)\mathrm{d}t = \delta\frac{N^2}{2}\frac{p}{s_{min}}\left(1+a-\frac{ap}{(g-p)s_{min}}\right)$$

Finally, it follows from Eq. 12 that long-run schedule delay costs are equal to:

$$TC_{SD}^{E,LR} = \int_{t_l^E}^{t_{l'}^E} C_{SD}^{LR}(t)\dot{Z}^E(t)\mathrm{d}t = \delta\frac{N^2}{2}\frac{p}{s_{min}}\frac{ap}{(g-p)s_{min}} \tag{34}$$

A.3 Derivation $\tau^{F,LR}(t)$

If only first-best short-run tolls were implemented, the price function, $EP^F(t)$, would be given by the following equation, using the earlier derived results for the costs of short-run and long-run schedule delays (Eq. 3 and 12) and first-best short-run tolls (Eq. 21) (for the case that $t \leq 0$ (hence, SRPAT≤LRPAT)).

$$EP^F(t) = \underbrace{p\beta\left(t - t_q(s_{min}) - \frac{Z(t)}{s_{min}}\right)}_{EC_{SD}^{SR}(t)} + \underbrace{g\beta(-t)}_{C_{SD}^{LR}(t)} + \underbrace{p\beta\frac{Z(t)}{s_{min}}}_{E\tau^{SR}} \tag{35}$$

As argued in Section A.1, $t_q(s_{min})$ must be equal to $-\theta N/s_{min}$ if the distribution of SRPATs is uniform, which is true also in the first-best optimum, where $Z^F(t)$ is equal to s_{max} (Eq. 19). We can then substitute this expression for $t_q(s_{min})$ in Eq. 35. Moreover, it is easy to see that the $p\beta Z(t)s_{min}$ terms cancel out. The derivative of the price function with respect to t, $\mathrm{d}EP^F(t)/\mathrm{d}t$, is thus equal to $(p-g)\beta$. In order to provide for price equality among travelers in the optimum, the long-run toll, $\dot{\tau}^{F,LR}(t)$, must therefore be set such that its derivative is equal to $(g-p)\beta$. The toll itself, $\tau^{F,LR}(t)$, must then equal $(g-p)\beta Z^F(t)/s_{max}$, since it starts from 0 for the driver with the earliest SRPAT. Similarly, it can be shown that for $t > 0$, $\dot{\tau}^{F,LR}(t)$ must be equal to $(p-g)\gamma$, and $\tau^{F,LR}(t)$ to $(g-p)\gamma Z^F(t)/s_{max}$

References

Arnott, R. (2007). Congestion tolling with agglomeration externalities. *Journal of Urban Economics*, *62*(2), 187–203.

Arnott, R., De Palma, A., and Lindsey, R. (1990). Economics of a bottleneck. *Journal of Urban Economics*, *27*(1), 111–130.

Arnott, R., De Palma, A., and Lindsey, R. (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 161–179.

Arnott, R., De Palma, A., and Lindsey, R. (1996). Information and usage of free-access congestible facilities with stochastic capacity and demand. *International Economic Review*, 181–203.

Arnott, R., De Palma, A., and Lindsey, R. (1999). Information and time-of-usage decisions in the bottleneck model with stochastic capacity and demand. *European Economic Review*, *43*(3), 525–548.

Börjesson, M. (2009). Modelling the preference for scheduled and unexpected delays. *Journal of Choice Modelling*, *2*(1), 29–50.

Börjesson, M., Eliasson, J., and Franklin, J. (2012). Valuations of travel time variability in scheduling versus mean-variance models. *Transportation Research Part B: Methodological*. (forthcoming)

Chorus, C., Molin, E., and Wee, B. van. (2006). Travel information as an instrument to change car drivers travel choices. *EJTIR*, *6*(4), 335–364.

Daganzo, C. (1985). The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation Science*, *19*(1), 29–37.

De Vany, A., and Saving, T. (1982). Life-cycle job choice and the demand and supply of entry level jobs: some evidence from the air force. *The Review of Economics and Statistics*, *64*(3), 457–465.

Fosgerau, M., and Small, K. (2010). Endogenous scheduling preferences and congestion. In *Kuhmo nectar conference on transport economics, Valencia, Spain.*

Henderson, J. (1981). The economics of staggered work hours. *Journal of Urban Economics*, *9*(3), 349–364.

Hendrickson, C., and Kocur, G. (1981). Schedule delay and departure time decisions in a deterministic model. *Transportation Science*, *15*(1), 62–77.

Jenelius, E., Mattsson, L., and Levinson, D. (2011). Traveler delay costs and value of time with trip chains, flexible activity scheduling and information. *Transportation Research Part B: Methodological*.

Lindsey, R. (1995). Optimal departure scheduling for the morning rush hour when capacity is uncertain. In *7th world conference on transport research, Sydney, Australia* (pp. 16–21).

Lindsey, R. (2004). Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes. *Transportation Science*, *38*(3), 293–314.

Newell, G. (1987). The morning commute for nonidentical travelers. *Transportation Science*, *21*(2), 74–88.

Peer, S., Verhoef, E., Knockaert, J., Koster, P., and Tseng, Y. (2011). Long-run vs. short-run perspectives on consumer scheduling: Evidence from a revealed-preference experiment among peak-hour road commuters.

Pigou, A. (1920). *The economics of welfare.* London, UK.

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

Small, K., and Verhoef, E. (2007). *The economics of urban transportation.* London, UK: Routledge.

Smith, M. (1979). The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, *13*(4), 295–304.

Tseng, Y., Knockaert, J., and Verhoef, E. (2011). A revealed-preference study of behavioural impacts of real-time traffic information. *Transportation Research Part C: Emerging Technologies*.

Van Ommeren, J., and Fosgerau, M. (2009). Workers' marginal costs of commuting. *Journal of Urban Economics*, *65*(1), 38–47.

Van Ommeren, J., Van Den Berg, G., and Gorter, C. (2000). Estimating the marginal willingness to pay for commuting. *Journal of Regional Science*, *40*(3), 541–563.

Vickrey, W. (1969). Congestion theory and transport investment. *The American Economic Review*, *59*(2), 251–260.

Weiss, Y. (1996). Synchronization of work schedules. *International Economic Review*, 157–179.

Zhang, X., Yang, H., Huang, H., and Zhang, H. (2005). Integrated scheduling of daily work activities and morning-evening commutes with bottleneck congestion. *Transportation Research Part A: Policy and Practice*, *39*(1), 41–60.