Congestion in the bathtub

Mogens Fosgerau
Technical University of Denmark, Denmark, and
Royal Institute of Technology, Sweden
mf@transport.dtu.dk
May 21, 2014

Abstract

This paper presents a bathtub model of downtown traffic congestion that accounts for hypercongestion and for the fact that people drive different distances. Holding travel demand constant, the model shows, in a leading case, that a toll rate of zero is welfare optimal among a wide class of time-varying tolls. In general, the potential welfare gain from tolling limited. These findings stand in sharp contrast to the prevailing wisdom that builds on the Vickrey bottleneck model.

Keywords: dynamic; congestion; urban; traffic; bottleneck; bathtub
JEL codes: D0; R4
1 Introduction

Anybody living in a major city will appreciate that congestion is a significant issue for economic policy. For the US, for example, it is estimated that urban road congestion in 2011 caused a total of 5.5 billion hours of delay (Schrank et al., 2012). The basic Pigouvian analysis of traffic congestion leads to the suggestion that car drivers should be charged their marginal external cost. In reality it is, however, not so clear what this cost is. The Pigouvian analysis is static, using just that the marginal cost of traveling diverges from the average cost as the traffic volume increases. This perspective ignores an essential fact about traffic congestion, namely that there are queues that build and dissipate over the course of a day. Thus traffic congestion is a dynamic phenomenon and the timing of trips must be taken into account.

The seminal Vickrey (1969) bottleneck model has shaped our intuition about urban congestion dynamics. That model describes queueing before a bottleneck, for example located at the entrance to the central business district of a city where drivers enter during the morning commute. The bottleneck allows cars to enter only at a certain maximal rate. Drivers have similar preferences regarding when they would like to arrive at work and therefore a queue first builds up before the bottleneck and then dissipates every morning. In equilibrium, drivers trade off the inconvenience of deviating from their preferred schedule against the time lost queueing. This means that the inconvenience of the timing of trips as well as the dynamics of congestion should be included in the calculation of the costs of congestion.

The defining property of the bottleneck congestion technology is the fixed rate at which cars can exit from the bottleneck. In this case, if drivers in the middle of the peak could be induced to delay their departures, then it would be possible to reduce the queue without drivers arriving later at work. This underlies the finding that there are large potential efficiency gains available from the retiming of trips. Arnott et al. (1993) exhibit a stylized case in which an efficiency gain can be harvested through the imposition of a time-varying toll that eliminates queueing and the efficiency gain is equal to half the congestion cost that the bottleneck imposes on drivers in unregulated equilibrium. Drivers will be indifferent between the tolled and the untolled equilibrium, all of the efficiency gain will be captured as toll revenue. The bottleneck model is briefly reviewed in section 3.

The bottleneck model is behind the now common intuition that there are large potential efficiency gains available from retiming traffic and that these gains can be harvested without making drivers worse off. This paper will challenge this

---

1Vickrey’s paper is extensively cited and has spawned a lively literature on regulating congestion dynamics.
intuition by employing a more realistic description of congestion.

It is well established that the instantaneous speed at a single point on a road is a decreasing function of the instantaneous density of cars at that point (Green-shields, 1935). The fundamental identity of traffic flow holds that flow, i.e. the number of cars passing a point per time unit, equals speed times density, where density is the number of cars per distance unit. Flow is then (with appropriate shape restrictions on the speed-density relationship) an inverse u-shaped function of density. On the upward sloping part we talk about congestion, as higher density leads to higher flow but reduced speed. On the downward sloping part we talk about hypercongestion. Here higher density is associated with both lower flow and reduced speed. The combined relationships between speed, density and flow are called the fundamental diagram of traffic flow. This is illustrated in Figure 1.

A recent range of contributions have shown that such a fundamental diagram of traffic flow also applies at the level of an urban neighborhood (Geroliminis and Daganzo, 2008; Daganzo et al., 2011). The underlying mechanism is that drivers continuously adapt their route choices to avoid more congested places in the network. This adaptation process equalizes congestion such that it tends to be constant across space. A stable relationship emerges between the density of cars in the network and the space-averaged speed. Vickrey actually preceded Daganzo in a way, working on what he called a bathtub model of congestion, which is based on the same intuition.\footnote{This work was never completed but a handwritten note has been preserved (Vickrey, 1991).} Just like the water level is the same everywhere

![Figure 1: Fundamental diagram of traffic flow](image)
in a bathtub, the bathtub model holds that the level of congestion is the same everywhere in an urban neighborhood.

The inverse-u relationship between flow and density does not occur in the bottleneck model. Flow out of the bottleneck does increase with density before the bottleneck until the point where the capacity flow is reached. At higher densities, however, the flow stays constant and does not decrease as in the fundamental diagram. Thus the bottleneck does not generate hypercongestion and this means that bottleneck congestion is a poor description of downtown urban congestion.

The bathtub model uses an aggregate speed-density relationship to describe downtown urban congestion and thus incorporates hypercongestion. It considers a fixed mass of homogeneous drivers who care about the timing of their trips, traveling in an urban area where the speed at any instant is governed by a macroscopic fundamental diagram of traffic flow, depending on the instantaneous number of cars in transit. Section 4 considers first the case where all have the same distance to drive and shows that Nash equilibrium exists uniquely in this model. Social optimum in this setting is investigated through simulation. A finding is that some drivers may be worse off in social optimum than in Nash equilibrium, which would not occur under bottleneck congestion. The bottleneck model has the bottleneck as the only source of congestion and there it is possible to remove congestion completely. In the bathtub there is still congestion in the social optimum.

Section 5 allows trip lengths to be heterogeneous. Under a regularity assumption that puts a bound on the rate at which instantaneous speed changes, drivers will sort such that drivers with long trips depart earlier and arrive later than drivers with short trips; the paper calls this regular sorting. Regular sorting implies that the equilibrium travel time for a driver does not depend on the particular form of trip timing preferences. Any toll that does not break the regular sorting of drivers can then only be welfare decreasing as it can not improve travel times. This result is in sharp contrast to the bottleneck model. In the bathtub model with regular sorting, the Nash equilibrium is also the social optimum.

The paper goes on to consider time-varying tolls that do break the regular sorting and finds that such tolls will have to be rather extreme. They must be large relative to utility rates and must have an inverse u-shape, such that drivers are charged at the lowest rate when there is most traffic.

Section 6 presents some comparative statics results for the cases with homogeneous and heterogeneous trip lengths.

Section 7 extends the analysis to the case where there is heterogeneity with respect to trip lengths as well as the temporal location of trip timing preferences. Again under a regularity condition, drivers will sort regularly according to trip length within each temporal location group. A simulation example shows efficiency gains from tolling that are small relative to the toll revenue.
In summary, the Vickrey bottleneck model implies large potential efficiency gains from retiming and these efficiency gains can be realized without making drivers worse off while queueing is eliminated. In the bathtub with homogeneous trip lengths there is scope for an efficiency gain, but some drivers may be worse off in social optimum and some congestion remains. When drivers are allowed to be heterogeneous with respect to trip length, then no efficiency gain from retiming is possible under regular sorting. Some efficiency gain is possible when heterogeneity with respect to trip timing preferences is allowed for.

Thus we should not expect that the large efficiency gains from trip retiming available in the bottleneck are generally available in the bathtub - they derive specifically from the bottleneck congestion technology and are not there under more realistic descriptions of downtown traffic congestion.

An important issue is the role played by distance. Distance is essentially ignored in the basic Vickrey bottleneck model. Vickrey (1969) employed so-called $\alpha-\beta-\gamma$ trip timing preferences, represented by a utility function that is linear in travel time and separable in travel time and arrival time. Then driver heterogeneity with respect to distance to the bottleneck can be ignored. In the bathtub, each driver has some distance to cover in the urban area and there is an aggregate distribution of trip lengths. Given the trip length and the departure time, the arrival time is determined from the evolution of speed over the day. This is a realistic aspect of the bathtub model but it also leads to some mathematical challenges. The main issue is that all drivers simultaneously in transit interact with each other. In the bottleneck, an additional trip initiated at time $t$ affects only later departures. In the bathtub it also affects trips initiated earlier that are still in transit at time $t$.

Arnott (2013) develops a version of Vickrey’s bathtub model, which is similar to the present model. In comparison to Arnott (2013), this paper uses a more general congestion technology but of the same type as Arnott/Vickrey. The present paper also uses more general trip timing preferences that comprise the $\alpha-\beta-\gamma$ trip timing preferences of Arnott/Vickrey as a limiting case. The main difference concerns the treatment of trip distance. Arnott (2013) makes an assumption that can be interpreted as saying that each driver’s trip length is random with a certain distribution and that drivers do not know their trip lengths at the time they make their departure decisions. This assumption is the price for obtaining analytical tractability in his model. Arnott (2013) finds in his model that the efficiency gain from congestion tolling might be smaller than the toll revenue when congestion is light and larger, even much larger, when congestion is severe.

Fosgerau and Small (2013) also analyze a bathtub type model. In their model

---

3Fosgerau and de Palma (2012) present a model with more general scheduling preferences where distance from the home to the bottleneck does matter and use this model to analyze commuting in a city where workers live at various distances from the CBD.
there is a bottleneck with a capacity that depends on the number of cars queueing; tractability is achieved by simplifying the bottleneck capacity function to a step function and there is no congestion outside the bottleneck. The present model does not resort to such ad hoc devices and allows for a general distribution of trip lengths where travel takes place.

2 Driver preferences

We begin by formulating the drivers’ preferences for the timing of a trip. Let \( h, w \) be real functions satisfying \( h, w > 0, h' < 0 < w' \) and \( h(0) = w(0) \). We furthermore require that the rates of change \( h'(\cdot) = \partial \ln h(t) / \partial t \) and \( w'(\cdot) \) exist everywhere and are bounded. The function \( h \) describes utility accumulated at the origin of the trip from some initial time, set to zero at no loss of generality, until departure time \( a \), this amounts to \( \int_0^a h(s) \, ds \). Similarly, \( \int_0^b w(s) \, ds \) is the utility accumulated at the destination from the arrival time until some arbitrary time, also set to zero at no loss of generality. Let \( \tau \) be a toll payment and define utility as

\[
 u(a, b, \tau) = \Lambda \left( \int_0^a h(s) \, ds + \int_b^0 w(s) \, ds \right) - \tau
\]

where \( \Lambda \) is a strictly increasing transformation. The marginal utilities of later departure and earlier arrival are positive and decreasing. The transformation \( \Lambda \) makes a difference when toll payments occur and when a social welfare function is considered.

A special case is when \( h \) and \( w \) are exponential functions

\[
 h(s) = \exp(\alpha_0 - \alpha_1 s), \quad w(s) = \exp(\alpha_0 + \beta_1 s),
\]

where \( \alpha_1, \beta_1 > 0 \). In this case \( h'(s) = h'(s) / h(s) = -\alpha_1 \) and \( w'(s) = w'(s) / w(s) = \beta_1 \), which will be a useful simplification at some places.

There is a continuum of \( N \) drivers and they have identical trip timing preferences represented by \( u \). We are concerned with the interaction of congestion dynamics with the timing of departures and therefore we regard \( N \) as fixed. The model could conceivably be extended to allow the total demand \( N \) to depend on utility \( u \).

The marginal utility of travel time is computed in the appendix for the case of exponential utility rates. This is useful for comparing changes in utility to changes in travel time. Computing the marginal utility of travel time requires a little consideration since it depends on when the trip takes place and on how departure and arrival times are supposed to adjust. Appendix C computes the
marginal utility change for a trip with specific beginning and end times with the change corresponding to an increase in travel time that maintains the ratio between the marginal utility of distance at the beginning and the end of the trip. When the trip is optimally timed by the driver then this corresponds to the optimal change in the timing of the trip.

3 The bottleneck model

This section reviews the Vickrey (1969) bottleneck model in order to have a basis for comparison. Drivers have the trip timing preferences just described. In the bottleneck model, all drivers must drive a certain distance before they reach a bottleneck and this part of their trips is uncongested. The distance before the bottleneck is constant and is normalized to zero at no loss of generality such that drivers arrive at the bottleneck immediately after departure. The bottleneck allows drivers to pass at a maximal rate of $\psi$, and a queue forms before the bottleneck if drivers arrive at the bottleneck at a higher rate. The queue discipline is first-in-first-out.\(^4\) This form of congestion may be thought to describe the morning commute when congestion occurs mainly at the entrance to the central business district.

Denoting by $R(t)$ the cumulative mass of departures at time $t$ and by $t_0$ the last time prior to $t$ when there was no queue, then the queue at the bottleneck at time $t$ is

$$Q(t) = R(t) - R(t_0) - \psi \cdot (t - t_0).$$

A driver arriving at the bottleneck at this time will exit the bottleneck and arrive at his destination at time $t + Q(t)/\psi = t_0 + (R(t) - R(t_0))/\psi$.

Consider Nash equilibrium where no driver has incentive to change his departure time. This boils down to the condition that utility is constant across active departure times and lower at inactive departure times. It is easy to find that equilibrium requires that departures, as well as arrivals, take place during an interval, that there is queue during the interior of this interval, and that the first and last drivers do not experience a queue. Let $t_0$ be the time of the first departure where $R(t_0) = 0$ and let $t_1$ be the time of the last departure where $R(t_1) = N$. Then $\psi \cdot (t_1 - t_0) = N$. The common utility for active departure times is

$$u_0 = \Lambda \left( \int_0^t h(s) \, ds + \int_{t_0 + R(t)/\psi}^0 w(s) \, ds \right)$$

and from the constancy of this expression it is possible to back out the departure schedule $R$ and to verify that the queue is first increasing and then decreasing.

\(^4\)de Palma and Fosgerau (2013) relax this assumption.
during the interval $[t_0, t_1]$. A toll that is small enough and does not change too steeply will leave unchanged the properties that departures take place during $[t_0, t_1]$ and that all drivers can pass the bottleneck during this interval. If the toll is non-zero only in the interior of the departure interval, then equilibrium utility $u_0$ is unaffected by the toll. Hence the toll revenue represents a pure efficiency gain. Equilibrium utility is maintained in the tolled equilibrium since the toll payment is balanced by reduced queueing resulting from drivers departing later than in the no-toll equilibrium.

4 The bathtub model with homogeneous trip lengths

The bathtub model considers trips that take place in an urban area where a uniform but time-varying speed $S(t) > 0$ prevails. For a given departure time $a$ and trip length $l$, the arrival time $b(a, l)$ is given implicitly by

$$l = \int_a^{b(a, l)} S(t) \, dt. \quad (3)$$

Denote by $D(t)$ the number of drivers on the road at time $t$, in a real city this quantity is proportional to the density of cars on the streets when the road network is held constant. The speed-density relationship is $S(t) = \psi(D(t))$, where $\psi' > 0, \psi'' < 0$, and it relates the instantaneous speed to instantaneous density by

$$S(t) = \psi(D(t)).$$

This section analyzes the bathtub model when all drivers travel the same distance $l$ within the bathtub area. Drivers are indexed by $q \in [0, N]$ and the departure time for driver $q$ is denoted $a(q)$. Drivers have trip timing preferences as before. Assume at no loss of generality that drivers depart in sequence given by $q$ such that $a(q)$ is increasing. All drive the same distance and therefore arrivals take place in the same sequence as departures; the arrival time is denoted by $b(q)$. Define for later convenience the function

$$\Gamma(q) = \frac{\psi(q)}{\psi(N - q)}.$$ 

We consider differentiable departure and arrival schedules $a(q), b(q)$. Given any departure schedule $a(q)$, differentiation of (3) with respect to $q$ leads to a differential equation that relates $b(q)$ to $a(q)$. The number of drivers on the road at time $t$
This differential equation is intractable in general since it involves departure and
arrival schedules as well as their derivatives and their inverses. Fortunately, there
is a simplification available when departures and arrivals do not overlap in time.
In that case the number of drivers on the road reduces to

\[ D(t) = \begin{cases} 
0, & t < a(0) \\
q, & t = a(q) \\
N, & t \in [a(N), b(0)] \\
N - q, & t = b(q) \\
0, & t > b(N). 
\end{cases} \]

When departures and arrivals do not overlap in time we shall say that they are
separated and we shall talk about separated Nash equilibrium and later separated
social optimum. We first consider separated Nash equilibrium where no driver has
incentive to change his departure time.

**Theorem 1** For \( l \) sufficiently large, a separated Nash equilibrium with differentiable departure and arrival schedules exists in the bathtub model with homogeneous trip lengths and it is unique (at least) when utility rates are exponential. The travel time \( b(q) - a(q) \) first increases, then decreases, and it is maximal for the median driver with \( q = N/2 \).

All proofs are in Appendix B. The proof of Theorem 1 shows that constant utility implies

\[ h(a(q)) = w(b(q)) \Gamma(q), \tag{4} \]

which fixes the arrival schedule as a function of the departure schedule. This
expression leads to

\[ a'(q) = \frac{\dot{\Gamma}(q)}{h(a(q)) - \dot{w}(b(q)) \Gamma(q)}. \]

The specific exponential form for the utility rates is convenient here since \( \dot{h}(t) = -\alpha_1, \dot{w}(t) = \beta_1 \) such that the derivative of the departure schedule is completely
determined by

\[ a'(q) = -\frac{\dot{\Gamma}(q)}{\alpha_1 + \beta_1 \Gamma(q)}. \]

\[ ^6 \text{Define } a^{-1}(t) = 0 \text{ if } t < a(0) \text{ and } a^{-1}(t) = N \text{ if } t > a(N) \text{ and similarly for } b(\cdot). \]
The time $a(0)$ of the first departure is determined by the condition that the distance driven is $l$.\footnote{The proof that equilibrium is unique under exponential utility rates uses that $a'$ does not depend on $a(0)$. It would be sufficient for existence that $a'$ did not vary too strongly with $a(0)$.} Social optimum is defined as the departure schedule that maximizes the average utility of drivers.

$$SW(a) = \int_0^N u(a(q), b(q), 0) dq.$$ \hspace{1cm} (5)

The simulation example in the following section shows that social optimum may not be separated. This means that an explicit solution to the social optimum is not available\footnote{Unless one can guess and verify a solution, which seems unlikely.} and numerical methods have to be used.

### 4.1 Simulation example

This section presents an example of Nash equilibrium and social optimum in the bathtub model with homogeneous trip lengths. There is a mass $N = 1$ of drivers and the speed-density relationship is linear, $\psi(D) = 1 - D/2$, which indicates a free-flow speed of 1 that is reduced to 1/2 when all drivers are in transit. This is a realistic speed reduction for very congested urban areas. Exponential utility rates $h, w$ are defined by (2) with $\alpha_1 = \beta_1 = 1$ and the utility transformation $\Lambda(x) = x$ is applied. Note that the simulation may be arbitrarily scaled by redefining the units for time and distance with no qualitative effect on results. Model time is reversible and the symmetry implied by $\alpha_1 = \beta_1$ provides an opportunity to check the quality of the simulation. The trip length is $l = 1$ and this is sufficiently large that Nash equilibrium is separated. Experimentation with the simulation revealed no cases where departures and arrivals were separated in social optimum.

The Nash equilibrium was computed analytically as $a(q) = -\ln(2 - q) - \frac{1}{2}$, $b(q) = \ln(1 + q) + \frac{1}{2}$ and is shown in Figure 2, where the lower curve is departures and the upper is arrivals.\footnote{Per Olsson pointed out that an analytical solution was available for the Nash equilibrium. The derivation is available from the author on request.} Since no explicit solution is available for the social optimum, a flexible approximation was computed instead.\footnote{The simulation is programmed in Ox (Doornik, 2001) and is available from the author on request. $a'(q)$ was defined as the square of a sixth-order polynomial, which ensures that $a(q)$ is increasing and very flexible. The arrival schedule $b(q)$ was computed by loading cars into the bathtub in time steps, computing current speed, incrementing distances and recording when trips are completed. The coefficients of the polynomial in $a'(q)$ together with the first departure time were chosen numerically using an optimization algorithm.} This is shown in Figure 3.

The simulation results are summarized in table 1.
Figure 2: Departures and arrivals in Nash equilibrium with homogeneous trip lengths.

Figure 3: Departures and arrivals in social optimum with homogeneous trip lengths.
Table 1: Simulation results with homogeneous distances

<table>
<thead>
<tr>
<th></th>
<th>Nash equilibrium</th>
<th>Social optimum</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average utility</td>
<td>-4.9462</td>
<td>-4.6087</td>
<td>0.3375</td>
</tr>
<tr>
<td>Minimum utility</td>
<td>-4.9467</td>
<td>-5.3688</td>
<td>-0.4221</td>
</tr>
<tr>
<td>Maximum utility</td>
<td>-4.9456</td>
<td>-4.2627</td>
<td>0.6829</td>
</tr>
<tr>
<td>Avg. duration</td>
<td>1.7726</td>
<td>1.4237</td>
<td>-0.3489</td>
</tr>
<tr>
<td>First departure</td>
<td>-1.1928</td>
<td>-1.5277</td>
<td>-0.3349</td>
</tr>
<tr>
<td>Last arrival</td>
<td>1.1930</td>
<td>1.5250</td>
<td>0.3320</td>
</tr>
<tr>
<td>Marginal utility of travel time</td>
<td>-4.5033</td>
<td>-2.4513</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Speed in social optimum with homogeneous trip lengths.
Nash equilibrium has drivers departing at a high rate. It is separated so there is an interval where everybody is on the road and the speed is minimal at $1/2$. The average duration of trips is 1.8, which may be compared to the uncongested travel time of 1.

Social optimum yields higher average utility but lower minimum utility than Nash equilibrium. This shows that if social optimum is implemented via a non-negative toll then the toll will make at least some drivers worse off when the use of toll revenues is not accounted for. The average duration of trips is reduced compared to Nash equilibrium. The departure and arrivals are stretched such that departures begin earlier than in Nash equilibrium, arrivals end later and departures and arrivals overlap. Figure 4 shows the speed profile for social optimum; the speed reaches a minimum of 0.6 which is again better than Nash equilibrium.\footnote{The simulation should yield exactly the same results with time and the sequence of drivers reversed. Thus the time of the first departure should be equal to minus the time of the last arrival and this is approximately the case. Figures 2 and 3 should be invariant with respect to a 180 degree rotation, which also seems about right. The speed profile in Figure 4 seems also to be about symmetric in time. Thus the quality of the simulation seems to be adequate.}

The marginal utility of travel time is computed as the average over drivers and may be used to translate utility changes into time units. The average utility increases by 0.34, which corresponds to about 0.1 time units. This may be compared to the change in the average trip duration of 0.35. Thus most of the travel time decrease is offset by worse timing of trips.

5 The bathtub model with heterogeneous trip lengths

Instead of all drivers having the same trip length we now allow for a distribution of trip lengths $l$. This step towards realism dramatically changes the properties of equilibrium. Let demand $\Phi(l)$ be the number of drivers with trip length of at least $l$ and assume that this distribution is absolutely continuous with density $\phi$ and supported on an interval that includes zero.

Now the arrival time $b(a,l)$ for a driver depends both on the departure time $a$ and the trip length $l$. It follows immediately from (3) that the partial derivatives of $b = b(a, l)$ are

$$\frac{\partial b(a,l)}{\partial a} = \frac{S(a)}{S(b)} \frac{\partial b(a,l)}{\partial l} = \frac{1}{S(b)}.$$ 

Drivers are indexed by their trip length. If $a(l)$ is the departure time for drivers with trip length $l$ then, denoting $b(l) = b(a(l), l)$,

$$b'(l) = \frac{1}{S(b(l))} + \frac{S(a(l))}{S(b(l))}a'(l).$$
Define for convenience the functions
\[ H(a) = \frac{h(a)}{S(a)}, \quad W(b) = \frac{w(b)}{S(b)}, \]
expressing the utility rates in terms of utility per distance unit rather than utility per time unit, and note that \( \dot{H}(a) = \dot{h}(a) - \dot{S}(a), \quad \dot{W}(b) = \dot{w}(b) - \dot{S}(b) \). We shall use the following regularity assumption for the analytical results.

**Assumption 1** For all times \( a, b : \dot{H}(a) < 0, \dot{W}(b) > 0 \).

This assumption is satisfied when the speed is constant. It remains satisfied if speed does not drop too quickly at times of departure or rise too quickly at times of arrival. It is also satisfied in Nash equilibrium in the numerical example below.

Theorem 2 states that drivers sort such that drivers with large \( l \) depart earlier and arrive later than drivers with small \( l \). We label this situation *regular sorting*. It is illustrated in Figure 5.

Theorem 2 Under Assumption 1, Nash equilibrium exists uniquely in the bathtub model with heterogeneous trip lengths. Drivers sort regularly such that \( a' (l) < 0 < b' (l) \) for all \( l \). Specifically,
\[
\begin{align*}
a' (l) &= \frac{1}{\psi'(l)} \frac{w'(b(l))}{h'(a(l)) - w'(b(l))}, \\
b' (l) &= \frac{1}{\psi'(l)} \frac{w'(b(l))}{h'(a(l)) - w'(b(l))}.
\end{align*}
\]
(6)

The travel time achieved by each driver depends only on the distribution of trip lengths and the speed-density relationship. The Nash equilibrium is socially optimal among departure schedules that maintain regular sorting.

Appendix B establishes the statements of the theorem regarding unique existence of Nash equilibrium and sorting, while the other statements are established here. The driver at \( l = 0 \) will depart at time 0 such that \( a(l) < 0 < b(l) \) for all \( l > 0 \). The drivers in transit at time \( a(l) \) are those with trips longer than \( l \) and the same is true at time \( b(l) \), which implies that \( S(a(l)) = S(b(l)) = \psi(\Phi(l)) \), see Figure 5. Then the first-order condition \( 0 = S(a) (H(a) - W(b(a,l))) \) reduces to
\[ h(a(l)) = w(b(l)). \]
(7)

Differentiating with respect to \( l \) leads to (6). Next, note that the first-order condition for the choice of departure time for a driver facing a constant travel time \( b(l) - a(l) \) is also (7); hence all drivers depart at the time that would be optimal...
if their travel time was fixed at the equilibrium value. Moreover, the travel time
does not depend on the specific form of preferences since
\[ b'(l) - a'(l) = \frac{1}{\psi(\Phi(l))}. \]  
This implies immediately that the Nash equilibrium is already the social optimum
among the cases with regular sorting. Then any time-varying toll that preserves
regular sorting can not increase welfare, when the welfare gain from the toll rev-
eneue is equal to the drivers’ welfare loss from the toll payment.

5.1 Simulation example

The social optimum is quite intractable. The example in this section uses simula-
tion to compute the equilibrium and the social optimum for the special case when
the utility rates are exponential given by (2). Appendix B establishes that
\[ a(0) = b(0) = 0, \]
\[ a'(l) = -\frac{1}{\psi(\Phi(l))} \frac{\beta_1}{\alpha_1 + \beta_1}, b'(l) = \frac{1}{\psi(\Phi(l))} \frac{\alpha_1}{\alpha_1 + \beta_1}, \]
constitutes the Nash equilibrium. The speed-density relationship \( \psi \) and the dis-
tribution of trip lengths \( \Phi \) must be chosen to conform to Assumption 1, which
according to the proof in the Appendix is equivalent to
\[ \frac{\alpha_1 \beta_1}{\alpha_1 + \beta_1} \geq \psi'(\Phi(l)) \phi(l). \]  

Figure 5 shows departure and arrival schedules in Nash equilibrium using
again \( N = 1, \alpha_1 = \beta_1 = 1 \) in the exponential specification of utility rates (2), trip
lengths that are uniformly distributed on \([0, 1]\) and the speed density relationship
\( \psi(D) = 1 - D/2 \). This is consistent with (9) and hence with regular sorting. The
utility transformation \( \Lambda(x) = x \) is applied. Figure 6 shows the corresponding
speed profile.

Social optimum is approximated by the same method as before. The resulting
departure and arrival schedules and the corresponding speed profile are shown in
Figures 7 and 8. The following table summarizes some figures from the simulation
of Nash equilibrium and social optimum.

The average utility increases as it should from Nash equilibrium to social op-
timum, but both the minimum and the maximum utilities are smaller in social
optimum than in equilibrium. Trips are much faster on average in social optimum
but are more spread out in time. The average duration of trips decreases by 0.11
time units from 0.77, but the average utility increases only 0.02 utility units cor-
responding to about 0.02 time units. Thus most of the travel time gain is offset by
worse timing of trips.
Figure 5: Sorting of departures and arrivals in Nash equilibrium with heterogeneous trip lengths

Figure 6: Speed profile in Nash equilibrium with heterogeneous trip lengths
Table 2: Simulation results with heterogeneous distances

<table>
<thead>
<tr>
<th></th>
<th>Nash equilibrium</th>
<th>Social optimum</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average utility</td>
<td>-2.9985</td>
<td>-2.9822</td>
<td>0.0163</td>
</tr>
<tr>
<td>Minimum utility</td>
<td>-3.9970</td>
<td>-4.1898</td>
<td>-0.1928</td>
</tr>
<tr>
<td>Maximum utility</td>
<td>-2.0000</td>
<td>-2.2983</td>
<td>-0.2983</td>
</tr>
<tr>
<td>Avg. duration</td>
<td>0.77159</td>
<td>0.66616</td>
<td>-0.1054</td>
</tr>
<tr>
<td>First departure</td>
<td>-0.69240</td>
<td>-0.53909</td>
<td>0.1533</td>
</tr>
<tr>
<td>Last arrival</td>
<td>0.69240</td>
<td>1.1261</td>
<td>0.4337</td>
</tr>
<tr>
<td>Marginal utility of travel time</td>
<td>-1.4983</td>
<td>-0.79921</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7: Sorting of departures and arrivals in social optimum with heterogeneous trip lengths
5.2 Tolling and resorting

A question that emerges is how social optimum could be implemented by a time-varying toll. This section discusses tolls that reverse the regular sorting of either departures or arrivals. Tolls may be defined up to a constant as 

$$\tau_1 (a) = \int_a^0 \pi_1 (s) \, ds,$$

which is a toll charged at the origin of a trip, as 

$$\tau_2 (b) = \int_0^b \pi_2 (s) \, ds,$$

which is a toll charged at the destination, or as 

$$\tau_3 (a, b) = \int_a^b \pi_3 (s) \, ds,$$

which is a toll charged while the trip is ongoing. Subtraction of these tolls from utility (1) may be cast as a modification of the utility rates as follows, using here $\Lambda (x) = x$ for simplicity.

$$u (a, b, \tau_1 (a)) = \int_0^a (h (s) + \pi_1 (s)) \, ds + \int_0^b w (s) \, ds,$$

$$u (a, b, \tau_2 (a)) = \int_0^a h (s) \, ds + \int_0^b (w (s) + \pi_2 (s)) \, ds,$$

$$u (a, b, \tau_3 (a, b)) = \int_0^a (h (s) + \pi_3 (s)) \, ds + \int_b^0 (w (s) + \pi_3 (s)) \, ds.$$

Consider again Assumption 1, but with $h, w$ modified by time-varying tolls to $h + \pi_1, w + \pi_2$. The assumption translates into

$$(h' (a) + \pi_1' (a)) < (h (a) + \pi_1 (a)) \hat{S} (a),$$

$$(w' (b) + \pi_2' (b)) > (w (b) + \pi_2 (b)) \hat{S} (b).$$
If both inequalities hold then regular sorting is maintained in equilibrium. However, if the first inequality is reversed then $b' < 0$, and if the second inequality is reversed then $a' > 0$. It is not possible to reverse both inequalities at the same time since then the second-order condition for utility maximization would necessarily be contradicted.

The simulation of social optimum above shows that this has the driver with trip length $l = 0$ departing not at his preferred time given by the first-order condition for a trip with zero duration $h(a) = w(a)$. This driver would have zero payment of a toll $\tau_3$ charged during trips. Thus, such a toll cannot implement social optimum.

There are essentially two ways of escaping the maximal congestion when homogenous drivers travel different distances, namely to design a toll to violate one of the inequalities of Assumption 1. This requires only a toll that is charged either at the time of departure or at the time of arrival. Let us consider just a toll $\tau_1$ charged at the time of departure, which means we can design it to achieve $b'(l) < 0$ for all $l$. This requires

$$\frac{h'(a) + \pi'(a)}{h(a) + \pi_1(a)} > \hat{S}(a).$$

(10)

The math involved in solving for equilibrium in this case is beyond this author.\textsuperscript{12} So it seems not possible to solve directly for the toll that leads to social optimum. However, given the departure and arrival schedules of social optimum, it is possible to back out the toll that implements social optimum from the first-order condition for the choice of departure time. Figure 9 shows the result of this exercise, a U-shaped toll. The following below suggests that this is a general result for a toll that induces a departure pattern like in Figure 7. We shall reason using Figure 10.

The figure shows first the utility rates $h$ and $w$, crossing at time 0. The unregulated equilibrium involves travel during an interval centered at this point and we shall consider a toll that leads to travel within the same interval.

With $a', b' < 0$ we expect the speed in the tolled equilibrium to be increasing and to reach maximum for the last driver with trip length $l = 0$. By (10), increasing speed requires that the tolled utility rate $h + \pi$ is strictly increasing. Drivers with $l = 0$ depart last at time $t_1$, when $h(t_1) + \pi(t_1) = w(t_1)$.

We may then infer that the toll rate at the time of travel for the last driver, $\pi(t_1)$, is large and positive. There is a time $t_2$ where $\pi(t_2) = 0$. At earlier times like $t_3$, $\pi(t_3) < 0$, since $h + \pi$ is increasing and $h$ is decreasing.

Thus we conclude as likely that a toll that would induce the sorting pattern with $b' < 0$ must have a minimum in the middle of the demand peak and be

\textsuperscript{12}For the same reasons as discussed in Section 4. It might be beyond current mathematics, according to Arnott (2013).
Figure 9: Toll at departure to implement social optimum with heterogeneous trip lengths

Figure 10: Graphical tolling example
largest on the shoulders of the peak. This is the complete opposite of what one
would expect of a congestion toll, namely that it should be highest when there is
most traffic. The second case with the sorting pattern \( 0 < a' < b' \) is the mirror
image of the first and the same conclusion regarding the shape of the toll applies
again. So the toll patterns that are required to induce the two sorting patterns that
lead to minimal congestion are quite peculiar. We have seen before that any toll
that does not break regular sorting will have a (weakly) welfare decreasing effect
on trip timing. We see now that tolls that do break the regular sorting probably
will have to be quite awkward. This suggests the general conclusion that it is
hard, and perhaps even impossible, to achieve efficiency gains in this model from
changes in trip timing induced by tolling. This may be the strongest conclusion
available since the problem of determining social optimum without the constraint
of regular sorting seems quite intractable.

6 Some comparative statics

Modifying the speed-density relationship into \( S(t) = \gamma \psi(D(t)) \) allows us to
derive the effect on social welfare in Nash equilibrium of varying \( \gamma \) around a
value of 1.

**Theorem 3** In the bathtub model with homogeneous trip length, the effect in sep-
arated equilibrium of a proportional change in the speed-density relationship is

\[
\frac{\partial SW}{\partial \gamma} = l \int_0^N \Lambda' \left( \int_0^{a(q)} h(s) \, ds + \int_{b(q)}^0 w(s) \, ds \right) W(b(q)) \, dq.
\]

The theorem has an analogue in the case of heterogeneous trip lengths and it
is stated without proof.

**Theorem 4** The effect of a proportional change in the speed-density relationship
is

\[
\frac{\partial SW}{\partial \gamma} = l \int_0^\infty \phi(l) \Lambda' \left( \int_0^{a(l)} h(s) \, ds + \int_{b(l)}^0 w(s) \, ds \right) W(a(l)) \, dl
\]

under Assumption 1 in the bathtub model with heterogeneous trip lengths.

Holding \( \Lambda \) constant, both theorems reveal that the welfare effect of a propor-
tional speed change increases more than proportionally to the trip length as well
as the number of drivers.
Theorem 5 In the bathtub model with homogeneous trip length and separated Nash equilibrium, when $\Lambda$ is the identity and $\psi$ is linear with slope $\psi' < 0$, then the marginal effect on the social welfare function of a change number of drivers simplifies to

$$\frac{\partial SW}{\partial N} = u(a(0), a(0), 0) + \psi' \int_0^N W(b(q)) (b(q) - a(N)) dq.$$

The marginal external loss of utility due to a marginal driver is bounded by

$$- \left( \frac{\partial SW}{\partial N} - u(a(q), b(q), 0) \right) \geq \frac{\psi(0) - \psi(N)}{\psi(0)} (b(N) - a(N)) w(b(N)).$$

Thus, under the simplifications assumed in the theorem, the marginal external loss of utility due to a marginal driver is at least the relative speed decrease, times the travel time for the last driver, times the utility rate at the destination at the arrival time for the last driver.

Define for convenience the reciprocal speed-density relationship as $\sigma(D) = 1/D$.

Theorem 6 Given Assumption 1 in the bathtub model with heterogeneous trip lengths, the marginal external loss of utility associated with a unit mass of additional drivers with trip lengths distributed according to survivor function $\eta(l)$ is

$$\int \phi(l) h(a(l)) \left( \int_0^l \sigma' (l') \eta(l') dl' \right) dl.$$

When $\eta$ collapses into a point mass at distance $l_0$ and $\sigma'$ is constant, then the marginal external loss of utility associated with an additional driver at trip length $l_0$ simplifies to

$$\sigma' \int \phi(l) h(a(l)) (l \wedge l_0) dl.$$

In the simplified expression for the marginal external utility loss in Theorem 6, the term $\sigma' \cdot (l \wedge l_0)$ is the additional travel time for a driver with trip length $l$, and $h(a(l))$ is the marginal utility of time.

7 The bathtub model with heterogeneous trip lengths and preferences

In this section we allow for heterogeneity with respect to the temporal location of preferences, while maintaining heterogeneity with respect to trip length. Recall
that the utility rates are defined with $h(0) = w(0)$ and that this implies that a
driver with zero trip length will depart and arrive at time 0. We now introduce a
constant $c$ to shift the temporal location of preferences and define

$$u(a, b, \tau|c) = u(a - c, b - c, \tau),$$

such that a driver with zero trip length and location $c$ will depart and arrive at time $c$. The following theorem generalizes Assumption 1 to allow for heterogeneity in
the location of preferences and shows that, in Nash equilibrium, drivers sort on $c$
while regular sorting holds for each $c$.

**Theorem 7** Assume that Nash equilibrium exists and satisfies $\dot{h}(a - c) - \dot{S}(a) < 0 < \dot{w}(b - c) - \dot{S}(b)$ for all $a, b, c$. Then, for fixed $c$, drivers sort regularly in
equilibrium and, for fixed $l$, the departure time is increasing as a function of $c$.

With heterogeneous location $c$, there is scope for achieving a welfare gain
from tolling, without breaking regular sorting. We shall investigate this using the
simulation model, now extended with a uniform distribution of $c$ on the interval
$[-1, 1]$. If we interpret the time unit as hours then the simulation could describe a
commuting peak that in the absence of congestion would last from 6.30 am to 9.30
am with short trips taking place during 7-9 am. Congestion increases the duration
of the peak. The speed-density relationship is taken again to be linear, which is in
accordance with the empirical evidence in Geroliminis and Daganzo (2008) and
the slope is set such that the speed-drop during the peak is about 20%, which is
well within a plausible range. Thus the simulation resembles a peak hour in a real
city.

The simulation is carried out by first computing the departure and arrival times
for each $c$ and $l$ using the first-order condition conditional on a speed profile $S$.
Then the speed profile is updated using the resulting departure and arrival sched-
ules. These steps are repeated until the speed profile is constant to a high degree
of precision. The simulation allows for modification of the utility rates by tolling
as discussed in Section 5.2 and does not rely on regular sorting to hold.

A series of simulations are carried out to illustrate the effect of a toll $\tau_3$ charged
during trips. The toll rate $\tau_3$ is zero until time $-1.5$ (or 6.30 am), increases linearly
until time $-1$ (7 am), is constant until time $1$ (9 am), decreases linearly until time
$1.5$ (9.30 am) and is zero thereafter. The maximum toll rate $\tau_3(0)$ is ranges from
0 to 0.5. The resulting toll rates are shown in Figure 11.

Increasing the maximum toll rate induces drivers with $c < 0$ to depart earlier
and drivers with $c > 0$ to depart later. Figure 12 shows the departure and arrival
schedules for drivers with $c = -1$ and varying trip length. Figure 13 shows the
resulting speed profiles. As trips are shifted apart in time, speed increases during
Figure 11: Toll rates in simulation

Figure 12: Departure and arrival schedules for $c = -1/2$. 
Figure 13: Speed profiles under various toll regimes

Table 3: Simulation results with varying maximum toll rates

<table>
<thead>
<tr>
<th>$\pi_3(0)$</th>
<th>Avg. gross utility</th>
<th>Avg. toll revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$-2.7609$</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>$-2.7586$</td>
<td>0.057764</td>
</tr>
<tr>
<td>0.2</td>
<td>$-2.7573$</td>
<td>0.11421</td>
</tr>
<tr>
<td>0.3</td>
<td>$-2.7561$</td>
<td>0.16903</td>
</tr>
<tr>
<td>0.4</td>
<td>$-2.7556$</td>
<td>0.22213</td>
</tr>
<tr>
<td>0.5</td>
<td>$-2.7560$</td>
<td>0.27342</td>
</tr>
</tbody>
</table>

an interval around time 0 and there is a smaller speed decrease on the shoulders of the peak.

Table 3 shows the average utility gross of toll and the toll revenue per driver. Utility is money-metric such that these figures are comparable. The welfare measure is the average gross utility. As the maximum toll rate $\pi_3(0)$ increases, the average gross utility first increases and then decreases indicating that the maximum welfare for this type of toll is attained with $\pi_3(0)$ around 0.4. The average toll revenue is much larger than the increase in gross utility and increases faster as $\pi_3(0)$ increases. At $\pi_3(0) = 0.4$, the toll revenue is 42 times the average welfare gain.
8 Conclusion

This paper has contributed by formulating and analyzing a tractable bathtub model of downtown congestion. Combining analytical results and simulation, the paper has examined Nash equilibrium, social optimum and tolling, allowing for driver heterogeneity both with respect to trip length and the temporal location of trip timing preferences. The overall conclusion is that the scope for efficiency gains from retiming of trips seems limited under bathtub congestion. This finding stands in sharp contrast to the case of bottleneck congestion, which has dominated the economic literature on congestion dynamics. It is also a very important finding since bathtub congestion is most likely the norm in real cities, while it is less easy to find examples where bottleneck congestion is a good description of congestion at the city level.

It is a positive conclusion that it is less important to seek to regulate the timing of trips, since it means that attention can return to regulating the number of trips. This is the obvious focus in static congestion models. However, as should be clear by now, congestion dynamics are very important. The economic literature on congestion dynamics has so far relied almost exclusively on the bottleneck model of congestion and combined this with assumptions that allow trip timing and the demand for trips to be treated separately. Attention has then focused on trip timing.

From the perspective of the current model a new kind of question seems natural, namely what happens when the demand for trips of each length and time depends on the cost of such trips. We would ideally like to be able to analyze policies that suppress demand at the height of the peak. So we would like to endogenize the decisions about whether and how far to drive. Such extension is not straightforward and the issue has been avoided by the extant literature.

References


A Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s, t$</td>
<td>Points in time</td>
</tr>
<tr>
<td>$S(t)$</td>
<td>Speed at time $t$</td>
</tr>
<tr>
<td>$h(t)$</td>
<td>Utility rate at origin</td>
</tr>
<tr>
<td>$\alpha_0, \alpha_1$</td>
<td>Parameters in specific form, $h(s) = \exp(\alpha_0 - \alpha_1 s)$</td>
</tr>
<tr>
<td>$H(t) = \frac{h(t)}{S(t)}$</td>
<td>Utility rate at origin in distance terms</td>
</tr>
<tr>
<td>$w(t)$</td>
<td>Utility rate at destination</td>
</tr>
<tr>
<td>$\beta_0, \beta_1$</td>
<td>Parameters in specific form, $w(s) = \exp(\beta_0 + \beta_1 s)$</td>
</tr>
<tr>
<td>$W(t) = \frac{w(t)}{S(t)}$</td>
<td>Utility rate at destination in distance terms</td>
</tr>
<tr>
<td>$u$</td>
<td>Utility</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>Strictly increasing transformation used in definition of utility</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Toll</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Toll rate</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of drivers</td>
</tr>
<tr>
<td>$l$</td>
<td>Trip length</td>
</tr>
<tr>
<td>$q, r$</td>
<td>Index of drivers, $q, r \in [0, N]$</td>
</tr>
<tr>
<td>$\Phi, (\phi = -\Phi'), F, \eta$</td>
<td>Survivor functions for trip length in the model with heterogeneous trip lengths</td>
</tr>
<tr>
<td>$D(t)$</td>
<td>Density of cars at time $t$</td>
</tr>
<tr>
<td>$\psi(D)$</td>
<td>Speed-density relationship</td>
</tr>
<tr>
<td>$\Gamma(q) = \frac{\psi(q)}{\psi(N-q)}$</td>
<td></td>
</tr>
<tr>
<td>$\dot{x}(t) = \frac{x'(t)}{x(t)}$</td>
<td>Dot above a function of time denotes rate of change</td>
</tr>
<tr>
<td>$SW$</td>
<td>Social welfare function</td>
</tr>
<tr>
<td>$a(q), b(q)$</td>
<td>Departure and arrival time for driver $q$</td>
</tr>
<tr>
<td>$b(a,l)$</td>
<td>Arrival time as a function of departure time and distance</td>
</tr>
</tbody>
</table>

B Proofs

**Proof of Theorem 1.** We shall establish that the necessary conditions for equilibrium identify a single solution with differentiable departure rates, which hence exists and must be unique. Equilibrium implies that utility is constant which in turn implies that

$$h(a(q))a'(q) = w(b(q))b'(q).$$

Distance traveled is constant $l$ for all drivers

$$l = \int_{a(q)}^{b(q)} S(t) \, dt$$
implying
\[ \psi(q) a'(q) = \psi(N-q) b'(q) \]
or equivalently
\[ h(a(q)) a'(q) = w(b(q)) \Gamma(q) a'(q). \]
This determines \( b(q) \) as a function of \( a(q) \) whenever \( a'(q) > 0 \) (\( a'(q) \geq 0 \) by construction).

Next take logs and differentiate to obtain (I am now omitting function arguments to make expressions clearer, no ambiguities arise)
\[ \dot{h}a' = \dot{w}b' + \hat{\Gamma} = \dot{w}\Gamma a' + \Gamma \]
\[ a'(q) = \frac{\dot{\Gamma}(q)}{h(a(q)) - \dot{w}(b(q))\Gamma(q)} \]
The rates of change are bounded and thus \( a' \) is also bounded by say \( k \), such that \( a(N) - a(0) < Nk \). The speed is smaller than \( \psi(0) \) and the distance covered during \([a(0), a(N)]\) is then at most \( Nk\psi(0) \). A common trip length \( l \) larger than this ensures that equilibrium is separated.

We have already determined \( b(\cdot) \) as a function of \( a(\cdot) \). All that remains is to determine \( a(0) \). The distance covered during \([a(0), b(0)]\) is
\[
\int_{a(0)}^{a(N)} S(t) \, dt + \left[ w^{-1} \left( \frac{h(a(0))}{\Gamma(0)} \right) - a(N) \right] \psi(N)
\]
\[ = \int_{0}^{N} \psi(q) a'(q) \, dq + \left[ w^{-1} \left( \frac{h(a(0))}{\Gamma(0)} \right) - a(0) - \int_{0}^{N} a'(q) \, dq \right] \psi(N)
\]
\[ = \int_{0}^{N} (\psi(q) - \psi(N)) a'(q) \, dq + \left[ w^{-1} \left( \frac{h(a(0))}{\Gamma(0)} \right) - a(0) \right] \psi(N). \]

When utility rates are exponential, then \( a' \) does not depend on \( a(0) \). Hence it follows by straight differentiation that the distance covered during \([a(0), b(0)]\) strictly decreases as \( a(0) \) increases. Hence for large enough \( l \) there is a unique \( a(0) \) that lets all drivers cover that distance.

Existence follows with general utility rates since \( 0 < a'(q) < k \) shows that the distance covered can be made smaller or larger than \( l \) by changing \( a(0) \). A value of \( a(0) \) such that the distance covered is \( l \) follows by the mean value theorem.

It remains to check that no driver has incentive to depart earlier or later than given by \( a(\cdot) \). We shall look at infinitesimal changes. Consider first a driver who departs an infinitesimal \( \Delta \) time units earlier. In the time up to \( a(0) \) he will cover a distance of \( \Delta \psi(0) \). He arrives \( \nabla \) time units earlier than \( b(0) \) and
\( \Delta \psi (0) = \nabla \psi (N) \). This driver would gain utility \(- \Delta h (a (0)) + \nabla w (b (0)) = \Delta \left( - h (a (0)) + \Gamma (0) w (b (0)) \right) = 0 \) and so he would not gain by departing earlier. Even earlier departure times would be more costly since \( h \) is decreasing and \( w \) is increasing, while speeds are lower at the time this driver would arrive than when he departs. A similar argument can be made for a driver considering departure later than time \( a (N) \).

Thus equilibrium is completely determined by necessary conditions and is then unique. We have exhibited a solution which establishes existence.

Recall that

\[
a'(q) = \frac{\hat{\Gamma}(q)}{\hat{h}(a(q)) - \hat{w}(b(q)) \Gamma(q)}.
\]

The assertion that travel time is maximal for the median driver follows since the first-order condition leads to

\[
b'(q) = \frac{\hat{h}(a(q)) a'(q) - \hat{\Gamma}(q)}{\hat{w}(b(q))} = \Gamma(q) \frac{\hat{\Gamma}(q)}{\hat{h}(a(q)) - \hat{w}(b(q)) \Gamma(q)} = \Gamma(q) b'(q),
\]

such that

\[
b'(q) - a'(q) = a'(q) (\Gamma(q) - 1).
\]

This has the same sign as \( \Gamma(q) - 1 \), which is strictly decreasing and it is zero when \( q = N/2 \). Thus \( b(q) - a(q) \) first increases and then decreases with maximum at the median driver. \( \blacksquare \)

**Proof of Theorem 2.** The first-order condition for utility maximization for a driver with trip length \( l \) taking the speed profile \( S \) as given is

\[
0 = S(a) (H(a) - W(b(a, l))).
\] (11)

The corresponding second-order condition is\(^{13}\)

\[
0 \geq S(a) \left( H'(a) - W'(b) \frac{S(a)}{S(b)} \right) = S(a) H(a) \left( \hat{H}(a) - \hat{W}(b) \frac{S(a)}{S(b)} \right),
\]

and by Assumption 1 it is satisfied with strict inequality.

The first-order condition holds for all \( l \) and is equivalent to \( H(a(l)) = W(b(a(l), l)) \).

\(^{13}\)Some function arguments omitted for the sake of clarity.
Differentiating this with respect to \( l \) shows that

\[
0 = H' (a (l)) a' (l) - W' (b (a, l)) \left( \frac{1}{S (b (l))} + \frac{S (a (l))}{S (b (l))} a' (l) \right),
\]

which can be solved using the first-order condition to yield

\[
a' (l) = \frac{W (b (a, l))}{S (b (l)) H (a (l)) - W (b (a, l)) S (a (l))},
\]

\[
b' (l) = \frac{1}{S (b (l))} \frac{H (a (l))}{H (a (l)) - W (b (a, l)) S (a (l))},
\]

(13)

Under Assumption 1, we find that \( a' (l) < 0 < b' (l) \) for all \( l \).

The unique existence of Nash equilibrium follows since (13) is a pair of differential solutions that has a unique solution given initial conditions \( a (0) = b (0) = 0 \).

**Proof of Example in section 5.1.** This appendix establishes that the departure and arrival rates given in Section 5.1 constitutes Nash equilibrium. There will be sorting by construction and hence the speed is

\[
S (a (l)) = S (b (l)) = \psi (\Phi (l)) = 1 - \Phi (l) \cdot \frac{1}{3} = 1 - (1 - l) \cdot \frac{1}{3}.
\]

It is straightforward that \( l = \int_{a (l)}^{b (l)} S (t) \, dt \). The first-order condition for utility maximization, \( h (a (l)) = w (b (l)) \) is equivalent to \( 0 = \alpha_1 a (l) + \beta_1 b (l) \), which holds by the definition of \( a' (l) \) and \( b' (l) \).

Note that (omitting some function arguments)

\[
S (a (l)) = S (b (l)) = \psi (\Phi (l)),
\]

\[
S' (a (l)) = -\frac{\psi' (\Phi (l)) \phi (l)}{a' (l)},
\]

\[
S' (b (l)) = -\frac{\psi' (\Phi (l)) \phi (l)}{b' (l)}.
\]

Then Assumption 1 is satisfied since

\[
0 > \dot{H} (a (l))
\]

\[
\Downarrow
\]

\[
\frac{\alpha_1 \beta_1}{\alpha_1 + \beta_1} > -\psi' (\Phi (l)) \phi (l)
\]

\[
\Downarrow
\]

\[
0 < \dot{W} (b)
\]

and these inequalities hold by construction of \( \psi \) and \( \Phi \). The second-order condition is implied by Assumption 1. \( \blacksquare \)
Proof of Theorem 3. Consider Nash equilibrium with $\gamma = 1$. Then $a(q), b(q), D(t)$ and $S(t)$ depend on $\gamma$. We suppress this in the notation. Note that $q = D(a(q))$ implies that $0 = \frac{\partial a}{\partial q}(a(q))$. The derivative of the density is similarly zero at times of arrival and during the interval between the last departure and the first arrival. From $S(t) = \gamma \psi(D(t))$, we then find that $\frac{\partial S(t)}{\partial \gamma}|_{\gamma = 1} = S(t)$. Differentiating (3) then leads to

$$0 = S(b(q)) \frac{\partial b(q)}{\partial \gamma} - S(a(q)) \frac{\partial a(q)}{\partial \gamma} + l.$$  

Differentiating equilibrium utility and using the first-order condition for utility maximization (4) leads to

$$\frac{\partial u(a(q), b(q), 0)}{\partial \gamma} = h(a(q)) \frac{\partial a(q)}{\partial \gamma} - w(b(q)) \frac{\partial b(q)}{\partial \gamma}$$

$$= \frac{l w(b(q))}{S(b(q))}.$$  

Applying the definition of the social welfare function (5) yields the desired result.

Proof of Theorem 5. The change in density at time $t$ following a change in $N$ is $D(t) = 1$ when $t \geq a(N)$ and zero otherwise. Trip length is constant and differentiation and some rearrangement leads to

$$0 = S(b(q)) \frac{\partial b(q)}{\partial N} - S(a(q)) \frac{\partial a(q)}{\partial N} + \int_{a(N)}^{b(q)} \psi'(D(t)) \, dt$$

$$\Downarrow$$

$$\frac{\partial b(q)}{\partial N} = \frac{S(a(q)) \frac{\partial a(q)}{\partial N}}{S(b(q))} - \frac{1}{S(b(q))} \int_{a(N)}^{b(q)} \psi'(D(t)) \, dt.$$  

The derivative of equilibrium utility with respect to $N$ is

$$\frac{\partial u(a(q), b(q), 0)}{\partial N} = N \left( \int_{0}^{a(q)} h(s) \, ds + \int_{b(q)}^{0} w(s) \, ds \right)$$

$$\left( h(a(q)) \frac{\partial a(q)}{\partial N} - w(b(q)) \frac{\partial b(q)}{\partial N} \right).$$  

Applying the first-order condition for utility maximization leads to the following
simplification.

\[
\frac{\partial u (a(q), b(q), 0)}{\partial N} = \Lambda \left( \int_0^{a(q)} h(s) \, ds + \int_0^{b(q)} w(s) \, ds \right) \\
- W(b(q)) \int_{a(N)}^{b(q)} \psi'(D(t)) \, dt.
\]

Hence the derivative of the social welfare function is

\[
\frac{\partial SW}{\partial N} = u(a(q), b(q), 0) \\
+ \int_0^N \Lambda' \left( \int_0^{a(q)} h(s) \, ds + \int_0^{b(q)} w(s) \, ds \right) W(b(q)) \int_{a(N)}^{b(q)} \psi'(D(t)) \, dt \, dq.
\]

When \( \Lambda \) is the identity and \( \psi \) is linear with slope \( \psi' < 0 \), this simplifies to

\[
\frac{\partial SW}{\partial N} = u(a(q), b(q), 0) + \psi' \int_0^N W(b(q)) (b(q) - a(N)) \, dq
\]

and by Assumption 1, \( W(b(q)) \) is increasing as a function of \( q \), such that the marginal external loss of utility due to a marginal driver is bounded by

\[
u(a(q), b(q), 0) - \frac{\partial SW}{\partial N} = -\psi' \int_0^N W(b(q)) (b(q) - a(N)) \, dq \\
\geq -\psi' NW(b(N)) (b(N) - a(N)) \\
= \psi(0) - \psi(N) (b(N) - a(N)) w(b(N)).
\]

**Proof of Theorem 6.** Let \( \eta(l) \) be the survivor function for the trip length for a unit mass of drivers. This will be used to represent the trip length of marginal drivers. Let \( \delta \geq 0 \) and consider the survivor function of trip lengths to be \( F(l) = \Phi(l) + \delta \eta(l) \), such that \( \frac{\partial F(l)}{\partial \delta} |_{\delta=0} = \Phi_l \). Define for convenience the reciprocal of the speed-density relationship as \( \sigma(\Phi) = 1/\psi(\Phi) \), such that (8) becomes \( b'(l) - a'(l) = \sigma(F(l)) \). Note that the dependency of \( a, b \) on \( \delta \) is suppressed in the notation. Now,

\[
\frac{\partial (b'(l) - a'(l))}{\partial \delta} |_{\delta=0} = \sigma'((\Phi(l)) \eta(l)
\]
and $b(0) - a(0) = 0$ leads to

$$\frac{\partial}{\partial \delta} (b(l) - a(l)) = \int_0^l \sigma'(\Phi(l')) \eta(l') \, dl'.$$

The marginal external utility loss associated with additional drivers with trip lengths distributed according to $\eta(l)$ is

$$- \int_0^\infty \phi(l) \frac{\partial u(a(l), b(l), 0)}{\partial \delta} \bigg|_{\delta=0} \, dl = \int \phi(l) \, h(a(l)) \frac{\partial (b(l) - a(l))}{\partial \delta} \, dl = \int \phi(l) \, h(a(l)) \left( \int_0^l \sigma'(\Phi(l')) \eta(l') \, dl' \right) \, dl.$$

A point mass at trip length $l_0$ has survivor function $\eta(l) = 1_{\{l \leq l_0\}}$. When $\eta$ collapses into this and $\sigma'$ is constant, then the marginal external utility loss associated with an additional driver at trip length $l_0$ simplifies to

$$\sigma' \int \phi(l) \, h(a(l)) \, (l \land l_0) \, dl.$$

Proof of Theorem 7. The proof that there is regular sorting for each $c$ is the same as the proof of Theorem 6. Consider then some fixed trip length $l$. The first-order condition for the choice of departure time $a(l|c)$ is

$$0 = \frac{h(a(l|c) - c)}{S(a(l|c))} - \frac{w(b(a(l|c), l) - c)}{S(b(a(l|c), l))}.$$

For visual clarity, write $a$ for $a(l|c)$ and $b$ for $b(a(l|c), l)$. Differentiate the first-order condition with respect to $c$ to find

$$0 = \frac{h'(a - c)}{S(a)} \left( \frac{\partial a}{\partial c} - 1 \right) - \frac{h(a - c)}{S(a)^2} S'(a) \frac{\partial a}{\partial c} - \frac{w'(b - c)}{S(b)} \left( \frac{S(a)}{S(b)} \frac{\partial a}{\partial c} - 1 \right) + \frac{w(b - c)}{S(b)^2} S'(b) \frac{S(a)}{S(b)} \frac{\partial a}{\partial c},$$

which can be solved, using the first-order condition to simplify, to yield

$$\frac{\partial a}{\partial c} = \frac{\hat{h}(a - c) - \hat{w}(b - c)}{\hat{h}(a - c) - \hat{S}(a) - \frac{S(a)}{S(b)} \left( \hat{w}(b - c) + \hat{S}(b) \right)},$$

34
which is strictly positive as required. ■

\section{The marginal utility of travel time}

We consider the marginal utility associated with an increase in travel time for a trip beginning at time \( a \) and ending at time \( b \). This will provide a way to translate utility into travel time terms. Both the departure time and the arrival time are flexible, so we need to decide how they both change while \( b' - a' = 1 \). The first-order condition for utility maximization, \( h (a) / S (a) = w (b) / S (b) \), states that the marginal utility associated with increasing distance is the same at the beginning and the end of the trip. Differentiation of this equation and solving for \( a' \) leads to

\[
a' = \frac{\left( \dot{w} - \dot{S} (b) \right)}{\left( \dot{h} - \dot{S} (a) \right) - \left( \dot{w} - \dot{S} (b) \right)}.
\]

We use this expression also for trips that are non-optimally timed, which maintains the ratio between the marginal utilities associated with increasing distance. Differentiating gross utility when \( \Lambda \) is the identity, \( u (a, b, 0) = \int_a^b h (s) \, ds + \int_b^0 w (s) \, ds \) leads to

\[
u' (a, b, 0) = (h (a) - w (b)) a' - w (b)
\]

\[
= \frac{h (a) \left( \dot{w} - \dot{S} (b) \right) - w (b) \left( \dot{h} - \dot{S} (a) \right)}{\left( \dot{h} - \dot{S} (a) \right) - \left( \dot{w} - \dot{S} (b) \right)}
\]

Specialize this to exponential utility rates (2) to find that

\[
u' (a, b, 0) = -e^{\alpha a} \frac{e^{-\alpha_1 a} \left( \beta_1 - \dot{S} (b) \right) + e^{\beta_1 b} \left( \alpha_1 + \dot{S} (a) \right)}{\left( \beta_1 - \dot{S} (b) \right) + \left( \alpha_1 + \dot{S} (a) \right)}.
\]