Solving for equilibrium in the basic bathtub model

Richard Arnott \textsuperscript{a,*}, Joshua Buli \textsuperscript{b}

\textsuperscript{a}Department of Economics, University of California, Riverside, USA
\textsuperscript{b}Department of Mathematics, University of California, Riverside, CA 92521, USA

A R T I C L E   I N F O

Article history:
Received 16 June 2017
Revised 9 December 2017
Accepted 10 December 2017

Keywords:
Traffic congestion
Bathtub model
Computational solution

A B S T R A C T

The basic (identical individuals) bathtub model has an unfamiliar mathematical structure, with a delay differential equation with an endogenous delay at its core. The early papers on the model circumvented this complication by making approximating assumptions, but without solution of the proper model it is unclear how accurate the results are. More recent work has either considered special cases that can be solved analytically using familiar methods, or has turned to generic computational solution. This paper develops a customized method for computational solution of equilibrium in the basic bathtub model with smooth preferences that exploits the mathematical structure of the problem. An inner loop solves numerically for the entry rate, conditional on the equilibrium utility level, by verifying a trip distance condition. An outer loop uses the computed start time from the inner loop to solve for the population that commutes over the rush hour, then lowers the equilibrium utility level to repeat the inner loop for a new level of utility. One result in that, even though tastes and the congestion technology are smooth, the entry rate and exit rate functions exhibit discontinuities at breakpoints. Another result is that, depending on the form of tastes and the congestion technology, the user cost curve as a function of population and may be backward bending.

© 2017 Published by Elsevier Ltd.

1. Solving for equilibrium in the basic bathtub model

Prior to a decade ago, essentially no data had been collected on downtown traffic congestion at the level of an entire downtown neighborhood or over an entire downtown area. That changed with the publication of a seminal paper by Geroliminis and Daganzo (2008 – GD hereafter) that analyzed data on traffic flow and traffic density for a neighborhood of Yokohama, Japan. The paper contained two central findings. The first was that there was a stable relationship between traffic flow and traffic density at the level of the neighborhood over the course of the day, and from day to day, which the authors referred to as the neighborhood’s Macroscopic Fundamental Diagram (MFD). The second was that this relationship has an inverted-U shape, with traffic flow rising with traffic density up to a critical density and then falling with density. Economists refer to the phenomenon in which traffic flow falls as traffic density rises under heavily congested conditions as hypercongestion. Thus, for the first time GD documented hypercongested traffic flow at the level of a downtown neighborhood. Subsequent studies have confirmed the empirical regularities identified by GD, though the form of the MFD, and its degree of stability, vary over cities.

* Corresponding author.
E-mail address: richard.arnott@ucr.edu (R. Arnott).

https://doi.org/10.1016/j.trb.2017.12.003
0191-2615/© 2017 Published by Elsevier Ltd.
### Notational Glossary

\( A_0, B_0, a_1, b_1 \)  
exponential utility function parameters

\( A_i(t; \mathbf{u}) \)  
k\( (t_i(t; \mathbf{u}); \mathbf{u}) \)

\( B_i(t; \mathbf{u}) \)  
1 + \( T(t_i(t; \mathbf{u}); \mathbf{u}) \)

\( C_i(t_i(t; \mathbf{u}); \mathbf{u}) \)  
\(-\( T(t_i(t; \mathbf{u}); \mathbf{u}) \)

\( D \)  
departure interval

\( E(t) \)  
cumulative number of entries at time \( t \)

\( F_i(t, \mathbf{u}) \)  
\( \prod_{j=1}^{n} B_j(t, \mathbf{u}) \), for \( i = 1, 2, \ldots, I \)

\( H(t, \mathbf{u}, L) \)  
\( \int_{L}^{t+T(t, \mathbf{u})} \nu(\hat{k}(t; \mathbf{u})) \) \( dt - L \)

\( I \)  
number of cycles where entries occur

\( L \)  
exogenous trip length

\( L \)  
computed trip length

\( N \)  
exogenous number of commuters

\( N \)  
computed population

\( M \)  
time of latest possible arrival

\( T(\cdot) \)  
trip duration

\( U(t, T(t)) \)  
utility function

\( V(t, T(t), y) \)  
total utility function

\( X(t) \)  
cumulative number of exits at time \( t \)

\( c \)  
generalized commuter cost

\( e(t) \)  
numerically computed entry rate

\( e(t) \)  
entry rate

\( \hat{k}(\cdot) \)  
numerically computed density

\( k_i \)  
capacity density

\( k_j \)  
jam density

\( k(\cdot) \)  
density function

\( q \)  
capacity flow

\( q_0, r_1, s_0, s_1 \)  
logarithmic utility function parameters

\( r \)  
time of the first departure

\( r' \)  
time of the last exit

\( t' \)  
time in the first cycle corresponding to \( t^* \) in cycle \( i \)

\( t^* \)  
time of the last departure

\( t_i(t; \mathbf{u}) \)  
time in the \( i \)th cycle corresponding to time \( t \) in the first cycle

\( t \)  
clock time

\( u \)  
equilibrium utility

\( v_f \)  
free flow velocity

\( v(\cdot) \)  
velocity function

\( x(t) \)  
exit rate

\( y \)  
income

\( \Delta \)  
time increments

\( \alpha \)  
unnormalized unit value of travel time

\( \beta \)  
unnormalized unit value of time early

\( \gamma \)  
unnormalized unit value of time late

\( v \)  
y + \( \max_{i} U(t, 0) \)

In both transportation science and transportation economics, Vickrey’s bottleneck model (1969), as adapted by Arnott et al. (1993), has been the workhorse model of equilibrium rush-hour traffic dynamics for a quarter century. Traffic congestion takes the form of queues behind bottlenecks of fixed flow capacity, which rules out hypercongestion. Urban transportation economists have long recognized the potential importance of hypercongestion, particularly in downtown areas, and have been exploring alternative ways of adapting the bottleneck model to accommodate hypercongestion. Two approaches have been taken. The first is to make the capacity of a bottleneck a function of the length of the queue behind it. The second is to develop isotropic models of downtown rush-hour traffic dynamics that incorporate MFD congestion (which assumes a stable relationship between traffic flow at a point in time and traffic density at that point in time), which have come to be referred to generically as bathtub models. The publication of GD has catalyzed their development. In the basic bathtub...
model, which is the focus of this paper, over the morning rush hour a fixed number of identical commuters travel an equal distance from home to work over an isotropic downtown area, with the velocity of traffic at a point in time depending on the density of traffic at that time. Thus, a particular commuter’s speed varies over the course of his journey, as traffic density changes.

Bathtub models have a serious weakness – solution of their equilibria and optima is generally analytically intractable, giving rise to a class of delay differential equations (DDES) whose analysis is at the research frontier in applied mathematics. Furthermore, there is little experience to draw on in their solution by numerical methods. The initial bathtub models, which were all basic bathtub models (assuming identical individuals), dealt with this intractability by making approximating assumptions that transform the delay differential equation into an ordinary differential equation (ODE) which is solvable using standard methods from the theory of ODEs. None of these models, which shall be reviewed individually in the next section, has gained widespread acceptance, since without solution of the “proper” model that they approximate, it is not possible to judge the accuracy of their approximated solutions. Subsequently, there have been two other approaches to dealing with the intractability of bathtub models, both of which shall be reviewed later. The first focuses on special cases for which analytical, and in some cases, closed-form solutions can be obtained. The second entails working with a continuum of individuals who differ in trip length, and perhaps some other characteristics as well, which alters the mathematical structure of the problem, and may surprisingly simplify solution.

This paper takes a different approach to dealing with the analytical intractability. Rather than attempt to obtain an exact or approximate analytical solution, it moves directly to numerical solution. By drawing on the mathematical and economic structure of the problem, the solution algorithm we develop avoids most of the problems that would arise from application of generic numerical solution methods. The algorithm is tailored to a solution of equilibrium in the basic bathtub model under the assumption that the utility or trip cost function is a smooth function of its arguments. We illustrate application of the algorithm using two simple utility functions.

In the process of developing the solution algorithm, we have derived some mathematical properties of the bathtub model. Some hold for all smooth utility functions; others hold under more restrictive conditions. We present these formally, in proposition-proof format, in a companion paper (Buli and Arnott, 2017). In this paper, we shall note these results as they arise.

Since knowledge of the structure of the basic bathtub model is necessary background to the literature review, the next section, Section 2, will present the basic bathtub model before Section 3’s literature review. Section 4 presents the solution algorithm, and Section 5 illustrates its application via an extended numerical example. Section 6 presents directions for future research and concluding remarks.

2. The bathtub model

• A thumbnail sketch A fixed number of commuters per unit area travel from home to work in the morning rush hour in an isotropic downtown area. Trip origins and destinations are uniformly distributed over the area and each commuter travels the same exogenous trip distance, \( L \). Commuters also have identical tastes, which are described by a utility function having as its arguments departure time and trip duration.\(^1\) The congestion technology is described by a function exhibiting a negative relationship between velocity and traffic density per unit area. Traffic density at a point in time equals the cumulative entries to the road system (which equals cumulative departures from home) minus the cumulative exits from the road system (which equals cumulative arrivals at work) to that point in time. Furthermore, for any commuter, cumulative entries at the time she enters the street network equal cumulative exits at the time she exits it (a FIFO condition). Equilibrium is obtained when no commuter can increase her utility by departing at a different time, and all commuters travel. Since the model can be completely solved for from the entry rate function, the objective of the solution algorithm is to solve for the entry rate function (or functions) consistent with the equilibrium conditions.

• The demand side of the model A fixed number \( N \) of commuters per unit area travel from home to work in the morning rush hour over an isotropic downtown area. They have the same tastes that are described by the smooth utility function, \( U(t, T) \), where \( t \) is the departure time from home and \( T \) is trip duration. The dependence of the utility function on \( t \) captures scheduling preferences. The utility function has the properties that: i) there is increasing marginal disutility to travel time, \( U_T < 0, U_{TT} < 0 \); ii) \( U_T < 0 \), with \( U_t \) being positive for early departure times, zero at the most preferred departure time (conditional on \( T \)), and negative for late departure times; iii) it is smooth. The trip duration function is obtained by inverting the equal utility condition, \( U(t, T(t)) = u \), to give \( T(t; u) \). We assume as well that the utility function is quasi-concave, which implies that the trip duration function is strictly concave; i.e. \( T(t; u) < 0 \).\(^2\) Commuters differ in their home and work locations. Work and home locations are uniformly distributed over the downtown area, and each commuter travels the same distance, \( L \), between home and work. Each commuter chooses when to leave home so as to maximize her utility, taking as given the equilibrium relationship between trip duration and departure time, \( T(t) \).

---

\(^1\) It is assumed that the commuter drives directly from home to work. Thus, one could just as well take utility to be a function of departure time and arrival time, or arrival time and trip duration.

\(^2\) Differentiate \( U(t, T(t)) = u \) twice. Differentiating once gives \( U_t + U_{TT} T = 0 \), so that \( \dot{T} = -U_t/U_{TT} \) (\( T(t) \) therefore equals the marginal rate of substitution between trip duration and departure time). Differentiating a second times yields \( U_{tt} + 2U_{TT} \dot{T} + U_{TTT} \dot{T}^2 + U_t \ddot{T} = 0 \). Thus, \( \dot{T} = -[U_t + 2U_{TT} \dot{T} + U_{TTT} \dot{T}^2]/U_T = -[U_t - 2U_{TT} U_t/U_T + U_{TT} (U_t/U_T)^2]/U_T \), which is strictly negative by quasi-concavity of the utility function.
The supply side of the model. Traffic congestion is described by a function relating velocity to traffic density: \( v = v(k) \). The fundamental identity of traffic flow is that the flow, \( q \), equals density times velocity. Combining the identity and the velocity function yields \( q(k) = kv(k) \), which is the model’s macroscopic fundamental diagram and provides an alternative description of the form of traffic congestion in the model. It is assumed that the MFD has the following properties: (i) \( q(0) = 0 \); (ii) \( q(k_f) = 0 \), where \( k_f \) is a finite jam density; (iii) it is smooth; and (iv) flow has a unique maximum, capacity flow, \( q_c \), which occurs at capacity density, \( k_c \).

The evolution of the traffic density over the rush hour is described using the following boundary conditions, functions, and differential equation: (i) \( k(t) = 0 \) where \( t \) is the time of the first departure from home; (ii) \( k(\bar{t}) = 0 \), where \( \bar{t} \) is the time of the last arrival at work; (iii) \( E(t) \) denotes cumulative entries to the road at time \( t \), \( X(t) \) denotes cumulative exits, and density is given by \( k(t) = E(t) - X(t) \); (iv) \( X(t + T(t)) = E(t) \), which states that cumulative arrivals by time \( t + T(t) \) equal cumulative departures by time \( t \); (v) \( E(t) = e(t) \) is the rate at which cars enter the traffic stream, which is also the departure rate from home, and \( X(t) = x(t) \) is the rate at which cars exit the traffic stream, which is also the arrival rate at work; (vi) \( e(t) \geq 0 \), which states simply that the departure rate from home is non-negative; and (vii) the differential equation \( k(t) = e(t) - x(t) \).

Equilibrium. A trip-timing equilibrium is an entry/departure rate function, \( e(t) \), an exit/arrival rate function, \( x(t) \), a density function, \( k(t) \), and a trip duration function, \( T(t) \), such that:

(i) No commuter can increase her utility by altering her departure time. This implies that utility is equalized over the departure set, \( D \), and is no higher outside the departure set

\[
U(t, T(t)) = \begin{cases} 
  u & \text{for } t \in D \\
  \leq u & \text{for } t \notin D.
\end{cases}
\]

(ii) Since the downtown area is isotropic, and since all cars on the road at a point in time travel at the same speed, all cars entering at the same time have the same trip duration and therefore exit at the same time. Thus, the cumulative number of cars that have exited the downtown traffic stream by time \( t + T(t) \) equals the cumulative number of cars that have entered the downtown traffic stream at time \( t \). This FIFO condition is

\[
X(t + T(t)) = E(t).
\]

Differentiation of (2.2) gives

\[
x(t + T(t))(1 + T(t)) = e(t).
\]

(iii) Since by assumption traffic density equals zero immediately prior to the first entry \( t \), \( k(t^-) = 0 \), the FIFO condition implies that

\[
k(t) = E(t) - X(t),
\]

\[
k(t) = e(t) - x(t).
\]

and traffic density equals zero immediately after the last exit \( t \), \( k(t^+) = 0 \).

(iv) Since all commuters travel an exogenous distance \( L \) and since at each point in time all cars travel at the same speed, trip duration as a function of departure time is given implicitly by the function

\[
\int_t^{t + T(t)} v(k(x)) \, dx = L.
\]

which states that trip distance equals the integral of velocity over the duration of the trip.

(v) Let \( t^* \) be the time of the last departure. Since by assumption all \( N \) commuters travel from home to work over the morning rush hour,

\[
\int_{\frac{t^*}{2}}^{t^*} e(t) \, dt = N.
\]

3. Literature review

This section will review the literature from the perspective of transportation economics.

Transportation economists have struggled for over fifty years to incorporate hypercongestion into mainstream transportation economics. Though Vickrey published his bottleneck model in 1969, it was not until the 1980’s that transportation economists began to develop an economic theory of rush-hour traffic dynamics based on it. Until then, transportation economic analysis was static, being interpreted as describing a steady state. Beckmann et al. (1956) treated steady-state travel time, user cost, and marginal social cost as increasing in steady-state flow. That specification continues to dominate practical transportation economic analysis.
Walters (1961) developed the first economic model of steady-state traffic that admits hypercongestion. Walters considered a steady-state flow of identical drivers traveling on a road of unit length with a common origin and destination, combining a technological relationship describing traffic congestion in which velocity is negatively related to density ($v = v(k)$ with $v'(k) < 0$) and the fundamental identity of traffic flow theory, that flow equals density times velocity gives an equation relating flow and density: $q(k) = kv(k)$. Simplifying by ignoring the money cost of travel, Walters treated the steady-state user cost, $uc$, as being steady-state travel time, $T(k) = 1/v(k)$, times the common value of time, $ho : uc(k) = \rho/v(k)$. He then plotted steady-state user cost against flow, as shown in Fig. 3.1. The steady-state user cost curve has two portions, one upward sloping, in which steady-state user cost is positively related to steady-state flow (congestion), the other backward bending in which steady-state user cost is negatively related to steady-state flow (hypercongestion). For his equilibrium analysis, he added a steady-state demand curve relating the steady-state flow demand for trips to the steady-state trip price, and argued that steady-state user equilibrium occurs at the point (or points) of intersection of the steady-state user cost curve and the steady-state demand curve. According to this analysis, there may be multiple steady-state equilibria. Furthermore, defining the steady-state marginal social cost of a trip to be the derivative of steady-state total cost with respect to steady-state flow, and the steady-state social optimum to occur at the point of intersection of the steady-state marginal social cost and steady-state demand curves, he obtained the result that there always exists a unique steady-state social optimum, which is congested. He went on to argue that the steady-state social optimum can be attained by imposing a congestion toll equal to the difference between the steady-state marginal social cost and the steady-state user cost of travel, evaluated at the socially optimal level of steady-state travel.

As the first paper to derive the optimal steady-state congestion toll, Walters (1961) remains the cornerstone of the economic theory of congestion pricing and a landmark in the history of transportation economic thought. Lacking a well-articulated model of traffic dynamics out of steady state, transportation economists dismissed hypercongestion as being an unstable, transient phenomenon that cannot occur in steady-state equilibrium. However, they did so with some discomfort, recognizing that applied to transient dynamics, Walters’ result of negative marginal social travel under hypercongestion jars with intuition.

Such was the conventional wisdom until Vickrey’s analysis of the bottleneck model (1969) started to gain currency in the 1990’s. Vickrey’s paper made two distinct contributions. The first was to provide a simple but sound formulation of traffic dynamics out of steady state. The second was to introduce a new equilibrium concept applicable to rush-hour traffic dynamics. The dynamics of rush-hour traffic is in (Vickrey or trip-timing) equilibrium when no user (commuter) has an incentive to alter her travel time. This naturally gave rise to the question: Can hypercongestion occur in trip-timing equilibrium? This question cannot be answered in the context of the bottleneck model, in which the flow rate through the bottleneck equals maximum or capacity flow.

As noted in the introduction, two classes of models were developed to address this question. The first directly extended the bottleneck model, allowing the exit capacity of the bottleneck to vary with the length of the queue behind it; for example, for this generalization, using optimal control theory Yang and Huang (1997) solve for the optimal time-varying pricing of a bottleneck with elastic demand, allowing for queuing. We do not review that branch of literature here. The second were bathtub models that treat MPF flow congestion in an isotropic downtown area.

We shall refer to the bathtub model described in the previous section as either the bathtub model (with identical commuters) or the “proper” bathtub model. It is proper in the sense that it properly reflects the physics of traffic congestion, in which a commuter’s trip duration is measured by summing up the travel time over each distance increment of his trip, taking into account that velocity varies over the course of the trip. In the proper bathtub model, there is a delay from the time the first commuter enters the traffic stream to the time he exits it, equal to his trip duration. Other, “improper” bathtub models make approximating assumptions in order to simplify the mathematics.

In incomplete notes that remain unpublished, Vickrey (1991) was the first to develop a bathtub model. According to the terminology of the previous paragraph, it was an improper bathtub model since it assumed that the exit rate from downtown traffic (i.e., the arrival rate at work) depends only on the density of downtown traffic, and hence that the first
exit occurs as soon as the first entry. Small and Chu (2003) presented the first fully developed bathtub model of equilibrium and optimum rush-hour traffic dynamics. While they specified the proper bathtub model, they solved an approximation to it in which a commuter’s velocity throughout his journey to work depends on the density of traffic when he enters the street system. Subsequently, Geroliminis and Levinson (2009) presented and analyzed a bathtub model with an especially simple form of the congestion function, and solved it on the approximating assumption that a commuter’s velocity throughout his journey depends on the density of traffic when he exits the street system. Arnott (2013) solved for equilibrium and optimum rush-hour traffic dynamics in the Vickrey (1991) model. According to the terminology employed here, all the models mentioned in this paragraph are improper bathtub models. None of them has gained general acceptance, likely because, without solution of the proper bathtub model, it is not possible to gauge the accuracy of their results.

Three more recent papers in the literature specify and solve proper bathtub models. The first, Arnott et al. (2016), is a special case of the proper basic bathtub model in which the commuter’s trip-timing decision is modeled in terms of a trip cost function rather than a utility function, with the \( \alpha - \beta - \gamma \) form of the cost function being assumed, and in which the congestion technology is described by Greenshields’ Relation (in which velocity is a negative linear function of density). They derived a closed-form solution of an equilibrium when no late arrivals are permitted, and analytical solution (viz. described by a system of algebraic equations) of both an equilibrium and a social optimum with both early and late arrivals. All these solutions entail departure masses; a subset of commuters depart, travel, and arrive together, with no other commuters entering the traffic stream during their travel, followed immediately by another subset of commuters who do the same.

The other two recent papers with proper bathtub models, Fosgerau (2015) and Lamotte and Geroliminis (2017), treat a continuum of heterogeneous commuters rather than identical commuters. One would expect solution of a proper bathtub with a continuum of heterogeneous commuters to be more complex than that with homogeneous commuters. However, with homogeneous commuters (or with a finite number of groups of commuters) the solution to the delay differential equation entails discontinuities in the departure rate at “breakpoints”, as will be explained later. Having a continuum of heterogeneous commuters smooths such discontinuities. If both the departure order and the arrival order of commuters with respect to trip length are monotonic, the dynamics can then be expressed in terms of ODEs rather than DDEs. Fosgerau (2015) and Lamotte and Geroliminis (2017) solve for equilibria and optima with heterogeneous commuters that satisfy this property.

Fosgerau (2015) assumes a general class of tastes in which utility is additive in a subutility function of the period spent at home prior to the commute, a subutility function of the period spent at work after the commute, and money (so that the utility function is money-metric) and a congestion function in which speed is negatively related to traffic density. He presents both analytical and simulation results. In Section 2, he focuses on continuous distributions of commuters differing only in trip length that satisfy the property that in equilibrium, shorter trips are carried out entirely within the duration of longer trips, which he terms regular sorting but could also be termed FILO (first-in, last-out) or LIFO sorting. Whether regular sorting occurs in equilibrium depends in a complicated way on tastes, the congestion technology, and the distribution of trip lengths. Under an endogenous condition that guarantees regular sorting, he shows that equilibrium is unique and coincides with the social optimum if the social optimum too entails regular sorting. Using simulation, he constructs numerical examples for which regular sorting occurs in equilibrium, and examines the properties of the equilibria. He also constructs numerical examples in which regular sorting does not occur in the social optimum. Section 3 considers elastic demand and Section 4 heterogeneous preferences.

Lamotte and Geroliminis (2017) assume the same class of tastes and the same form of the congestion function as Fosgerau (2015). Sections 3 through 5 present analytical results, while Section 6 reports on simulation results. With a continuous distribution of commuters by trip length, and perhaps other characteristics too, in Section 3 they show that, in equilibrium, density is a continuous function of time. In Section 4, with identical \( \alpha - \beta - \gamma \) preferences and a common desired arrival time at work, they show that, in the early morning rush hour, there is FIFO sorting with longer trips departing and arriving earlier, while in the late morning rush hour there is also FIFO sorting, but with shorter trip departing and arriving less late. They also investigate the situation where commuters have the same trip length but differ in \( \beta/\alpha \) or \( \gamma/\beta \). Section 5 investigates the social optimum and its decentralization via tolling. Section 6 first presents simulation results that confirm the analytical results, and then simulation results with more general preferences.

In both papers hypercongestion may occur in equilibrium, but it remains an open question as to whether it can occur in the social optimum.

The simulation method that Fosgerau employs to solve for equilibrium when commuters differ only in terms of trip length takes account of the analytical results derived in Section 2 of his paper. Lamotte and Geroliminis apply a more generic method to numerically solve for equilibrium, iterating towards a “fixed function” (the generalization of a fixed point) – a velocity function with the property that the utility-maximizing departure function, conditional on the velocity function, generates that velocity function. To solve numerically for the social optima (departure patterns that minimize total costs) both papers use hill-climbing methods. The analytical properties of all these methods (existence, uniqueness, and stability) have not yet been fully investigated.

---

3 As will be explained in remark 2 of Section 4.3, this difference is inconsequential.
4 “An” equilibrium rather than “the” equilibrium since the paper does not establish uniqueness.
There is also a body of literature that solves MFD-models of rush-hour traffic dynamics with exogenous entry functions. The most recent paper in this vein, Mariotte et al. (2017) contrasts the numerical solution properties of models which assume that a car’s velocity throughout a trip equals traffic speed at the time it enters the “reservoir”, which the authors term “accumulation-based”, to those which assume that a car’s velocity at a point in time is related to traffic density at that point in time, which the authors term “trip-based”, with a focus on the inaccuracy of the accumulation-based approach.

As perhaps the simplest possible model of downtown rush-hour traffic dynamics that is completely sound in the sense of fully respecting individual rationality and the physics of traffic flow, the proper bathtub model is very appealing. Unfortunately, however, it is analytically intractable. Previous papers have dealt with this intractability by sidestepping it, the earlier literature by making approximating assumptions, the more recent literature by considering special cases. This paper and its companion, Buli and Arnott (2017), take the bull by the horns, confronting the intractability directly. This paper develops a reliable algorithm to compute equilibrium in the basic proper bathtub model that is applicable when both the utility function and the congestion function are smooth. Buli and Arnott (2017) analytically derives equilibrium properties of the basic proper bathtub model under the same smoothness conditions.

4. The solution algorithm

The central feature of our solution strategy is to solve for equilibrium as a function of \( y \). One advantage of this procedure is that the equal utility condition can be inverted to give equilibrium trip duration as a function of departure time and the exogenous utility level, \( T(t; y) \). Another advantage is that, as we shall see, for every value of \( y \), there is a unique solution, which is not generally the case for \( t^* \), or \( N \). Intuitively, if the equilibrium trip duration function is known, it should be possible to solve for the time of the first departure, the time of the last departure, and the equilibrium entry rate function over \( D \). We define the refinement procedure in which the time of the first and last departure, and the equilibrium entry rate function are computed, such that the computed trip length \( L \) is within some tolerance to the exogenous trip length \( L_e \) as the inner loop of the solution algorithm. Once this refinement is completed, the population \( N \) is computed, then the inner loop is repeated for different values of exogenous utility. The function \( N(y) \) can then be constructed, from which the equilibrium (or equilibria) as a function of the exogenous population can be determined. We define the repeated inner loop computation for different exogenous utility levels as the outer loop of the solution algorithm. This intuition is sound, but it turns out to be more difficult than we had initially expected to solve completely for the full equilibrium given the equilibrium trip duration function.

Before proceeding to the structure of the inner and outer loops, some terminology is in order. Recall that \( t \) is the first departure, \( y \) is the exogenous utility level, and \( T(t; y) \) is trip duration as a function of departure time, conditional on \( y \). Consider the first commuter to depart. He departs at \( t \) and arrives at \( t + T(t; y) \). We define the interval \( (t, t + T(t; y)) \) to be the first entry cycle. And we define the second entry cycle to be \( (t + T(t; y), t + T(t; y) + T(t + T(t; y))) \); it is the travel time interval of the commuter who departs when the first commuter to depart arrives; and so on. We say that an entry cycle is a full entry cycle when commuters depart over the entire cycle, and is a partial entry cycle when commuters depart over the first part of the cycle but not over the entire cycle. In equilibrium, the departure set is connected. Thus, equilibrium may entail only a partial entry cycle, or an integer number of full entry cycles, followed by a partial entry cycle.

Since the notation becomes cumbersome for large numbers of full entry cycles, with some abuse of notation, for the rest of this section we shall let \( t \) denote some time in the first entry cycle. We then define the following condensed notation:

\[
t_1(t; y) = t, \quad t_2(t; y) = t + T(t; y), \quad t_3(t; y) = t + T(t; y) + T(t + T(t; y)), \quad \text{and so on. Thus, } t_j(t; y) = t \text{ in the first entry cycle corresponding to the time } t \text{ in the first entry cycle, which is just } t; \quad t_j(t; y) = t \text{ in the second entry cycle corresponding to time } t \text{ in the first entry cycle, and so on. Note that a first order condition can be obtained by taking a time derivative of } (2.6) \text{ (or equivalently, } (4.3)). \text{ This condition naturally divides the departure set into entry cycles. Using the condensed notation, the first full entry cycle runs from } t_1(t; y) \text{ to } t_2(t; y), \text{ the period over which the first commuter to depart travels; the second full entry cycle runs from } t_2(t; y) \text{ to } t_3(t; y), \text{ the travel period of the commuter who departs when the first commuter arrives; and so on. The final entry cycle is typically only a partial entry cycle. This situation is discussed in the proceeding subsection.}}
\]

The above definitions can be generalized as

\[
t_{j+1}(t; y) = t_j(t; y) + T(t_j(t; y); y).
\]

so that

\[
t_j'(t; y) = t_j'(t; y)(1 + T(t_j(t; y); y)), \quad \text{where } t_j'(t; y) = \frac{dt_j(t; y)}{dt}, \text{ etc.}
\]

(4.2)

Throughout the remaining sections of the paper, we will use the condensed entry cycle notation to simplify the mathematical expressions.

To summarize the description of the solution algorithm, we proceed in three steps:

(a) Determine the number of entry cycles for a given equilibrium utility level \( y \).
(b) In the inner loop, solve for the equilibrium entry rate function, conditional on starting values for \(y\) and \(t\). The loop continues updating the value of \(t\) and \(t^*\), with \(y\) fixed, until the computed trip length, \(L\), is within a provided tolerance of the exogenous trip distance \(L\) for the commuter who departs at \(t\). At this point, \(t\) and \(t^*\) are known.

(c) In the outer loop, using the computed values for \(t\) and \(t^*\) as functions of \(y\), the complete solution as a function of \(y\) is calculated, including \(N(y)\) for the fixed \(y\). The level of utility is then decreased by a provided \(\Delta u\), and the inner loop is repeated.

Now we introduce the detailed cycle structure that is inherently present due to the delay differential equation, and explain how to determine the number of cycles.

### 4.1. Cycle structure and cycle counting

Consider two commuters, the second of whom departs a period of time \(dt\) after the first. For most of their trips, the two commuters experience the same travel conditions. However, the first commuter, but not the second, travels on the road from \(t\) to \(t + dt\). While the second commuter, but not the first, travels on the road from \(t + T(t; y)\) to \(t + T(t; y) + d(t + T(t + dt; y))\). Recall that the entry rate function, hence each commuter, must satisfy

\[
\int_{t}^{t+T(t; y)} v(k(x; t; y)) \, dx = L \quad \text{for} \quad t \in \mathcal{D},
\]

which states that over the departure interval, \(\mathcal{D}\), the integral of velocity from the start time of a trip to the end time of the trip consistent with the start time and the equilibrium utility level must equal the exogenous trip length. We term this the distance condition.

Differentiating (4.3) with respect to \(t\) yields

\[
v(k(t + T(t; y); t; y))(1 + \dot{T}(t; y)) - v(k(t; t; y)) = 0,
\]

where going from (4.3) to (4.4), a constant of integration, exogenous trip distance \(L\) was dropped, and will used later in the refinement for the inner loop. Since both commuters travel the same distance, the distance traveled by the first commuter between \(t\) and \(t + dt\), \(v(k(t; t; y))dt\), is the distance traveled by the second commuter between \(t + T(t; y)\) and \(t + T(t; y) + (1 + \dot{T}(t; y))dt\). Thus, (4.4) indicates how the velocity at \(t\) and \(t + T(t; y)\) must be related in order for the trip distance of the two commuters to be the same. Since (4.4) provides a condition for equilibrium in terms of velocities, we refer to it as the velocity condition.

Our solution algorithm is formulated in terms of entry, but not exit, cycles. For the remainder of the body of the paper, we shall use the term "cycle" to refer to an "entry cycle". We distinguish between full and partial cycles. In a full cycle, entry occurs throughout the cycle, as discussed at the beginning of this section; in a partial cycle, which is always the last (entry) cycle, entry occurs over only an initial part of the cycle. Suppose, for the point of illustration, that the departure set comprises two full cycles, followed by a partial cycle. We say that there are three cycles. Let \(I\) denote the number of cycles, which is normally comprised of \(I - 1\) full cycles, followed by a single partial cycle, and \(i\) index the entry cycle.

A precise terminology for the start and end points of the cycles can be given in terms of the language of delay differential equations given in Bellen and Zennaro (2013), which defines the points \(t_i(t; y)\) as primary breakpoints, which will prove to be central in the numerical solution procedure. The use of the generic term, "breakpoint," will refer to a primary breakpoint, as there exists a second type of breakpoint. The breakpoints occur at the beginning of the first entry cycle, second entry cycle, etc.

Let \(t^*\) denote the time of the last entry, which always occurs in cycle \(I\), as stated previously. Then we define \(t^* = t_1(t^*; y)\) as the point in the first cycle corresponding to \(t^*\) in the \(I^{th}\) cycle. Generalizing this notation, \(t_i(t^*; y)\) is the point in the second cycle, \(t_2(t^*; y)\) is the point in the third cycle, etc., corresponding to \(t^* = t_i(t^*; y)\) in the \(I^{th}\) cycle. The \(t_i(t^*; y)\) are referred to as the secondary breakpoints in the delay differential equation literature. Both primary and secondary breakpoints play a role in the solution, as we shall see, and are the cause of the discontinuities in the entry rate function.

Using the notion of primary and secondary breakpoints, we can further decompose the cycles into subcycles. Each cycle (except when \(t^*\) occurs at a breakpoint) contains two subcycles, the first extending from \(t_i(t^*; y)\) to \(t_i(t^*; y)\), which we term the first subcycle, and the second extending from \(t_i(t^*; y)\) to \(t_{i+1}(t^*; y)\) which is referred to as the second subcycle. Fig. 4.1 illustrates the notation and terminology with respect to cycles.

With the description of the cycles and their notation complete, we now turn to how the number of cycles is determined. For the remaining portion of this subsection, we will assume that there are exactly \(I\) full entry cycles. Rewriting (4.4) using the cycle notation, we have

\[
v(k(t_{i+1}(t; y); t; y))(1 + \dot{T}(t_{i}(t; y); y)) - v(k(t_{i}(t; y); t; y)) = 0,
\]

for \(i = 1, \ldots, I + 1\). The index \(i\) must range from \(i = 1\) to \(i = I + 1\) due to the notation \(t_i = t_{i+1}(t; y)\) and \(\dot{T} = \dot{T}(t_i(t; y); y)\), since \(t^*\) occurs at the \((I + 1)^{st}\) breakpoint which is at the end of the \(I^{th}\) cycle, and the time of the last exit, \(\ddot{T} = \ddot{T}(t_{I+2}(t; y); y)\), occurs at the \((I + 2)^{nd}\) breakpoint which is at the end of the \((I + 1)^{st}\) cycle. These notational conventions will influence the boundary conditions, as we shall see shortly.
The cycle notation clearly demonstrates that \( (4.5) \) relates corresponding points in adjacent entry cycles. If there are exactly \( l \) full entry cycles and we fix an equilibrium utility level and first entry time, \( u \) and \( t \), \( (4.5) \) can be used to give a system of equations for each cycle \( i = 1, \ldots, l + 1 \), that yields

\[
\begin{align*}
v(k(t_1; t; u); t; u) &= v(k(t_2; t; u); t; u)(1 + \hat{T}(t_1; t; u); u) \\
v(k(t_2; t; u); t; u) &= v(k(t_3; t; u); t; u)(1 + \hat{T}(t_2; t; u); u) \\
&
\vdots \\
v(k(t_{i-1}; t; u); t; u) &= v(k(t_i; t; u); t; u)(1 + \hat{T}(t_{i-1}; t; u); u)
\end{align*}
\]

which gives the equations that relate the velocity at each breakpoint. Note that \( (4.6) \) has \( l + 1 \) equations; comprised \( l \) of equations for the entry cycles, and one equation for the last cycle of only exits, thus providing a total of \( l + 1 \) equations. We shall require \( l + 1 \) equations since \( (4.6) \) can be simplified using boundary conditions. Substituting out consecutive velocity terms line by line, \( (4.6) \) becomes

\[
\begin{align*}
v(k(t_1; t; u); t; u) &= v(k(t_{i-2}; t; u); t; u)(1 + \hat{T}(t_{i-1}; t; u); u) (1 + \hat{T}(t_i; t; u); u) \\
&\quad \cdots (1 + \hat{T}(t_2; t; u); u) (1 + \hat{T}(t_1; t; u); u) \\
&= v(k(t_{i-2}; t; u); t; u) \prod_{i=1}^{l+1} (1 + \hat{T}(t_i; t; u); u).
\end{align*}
\]

The first boundary condition is that traffic density is equal to zero at the time of the first entry, which is equivalent to the condition that velocity equal free-flow velocity at the time of the first entry. Using the above notation, this is written as

\[
v(k(t_1; t; u); t; u) = v_f.
\]

The second boundary condition is that traffic density equals zero at the time of the last arrival, which is equivalent to the condition that velocity equals free-flow velocity at the time of the last arrival. In terms of the above notation, this is written as

\[
v(k(t_{i+2}; t; u); t; u) = v_f.
\]

We now apply these conditions and \( (4.7) \) to divide \((t; u)\)-space into mutually exclusive and collectively exhaustive regions in which there are \( 1, 2, \ldots \) entry cycles. This will prove useful in the algorithm since, for any \((t; u)\), it will allow us to identify the number of entry cycles. Consider the boundary between the region where there is one entry cycle and the region where there are two entry cycles. This is the locus of \((t; u)\) for which there is exactly one full entry cycle with no partial entry cycle following it (so, \( i = 1 \), hence when the last entry occurs at the second breakpoint, and hence when the last exit occurs at
the third breakpoint, so that \( t_{i+1}(t'; u) = t_2(t'; u) = t_3(t'; u) \). Thus, in terms of \((4.9)\), the locus of \((t, u)\) for which there is one full cycle and no partial cycle is

\[
v(k(t_3(t'; u); t; u)) = v_f. \tag{4.10}
\]

From \((4.7)\),

\[
v(k(t_1(t'; u); t; u)) = v(k(t_3(t; u); u)) \prod_{i=1}^{2} \left( 1 + \tilde{T}(t_i(t; u); u) \right). \tag{4.11}
\]

Substituting \((4.8)\) and \((4.10)\) into \((4.11)\) gives

\[
1 = \prod_{i=1}^{2} \left( 1 + \tilde{T}(t_i(t; u); u) \right). \tag{4.12}
\]

Now we define the function

\[
G_1(t, u) = \left[ \prod_{i=1}^{2} \left( 1 + \tilde{T}(t_i(t; u); u) \right) \right] - 1. \tag{4.13}
\]

Then the locus of \((t, u)\) for which there is exactly one full entry cycle is given by \(G_1(t, u) = 0\). The argument can be extended to an arbitrary number of full entry cycles, so that the locus of \((t, u)\) for which there are exactly \(l\) full entry cycles is given by

\[
G_l(t, u) = \prod_{i=1}^{l} \left( 1 + \tilde{T}(t_i(t; u); u) \right) - 1 = 0. \tag{4.14}
\]

In the region between \(G_1(t, u) = 0\) and \(G_{l+1}(t, u) = 0\), there are \(i + 1\) cycles comprising \(i\) full cycles and a partial entry cycle.

The algorithm has an outer loop and an inner loop, as discussed previously. The outer loop incrementally lowers \(u\). The inner loop solves for equilibrium conditional on \(u\). The obvious starting point for the algorithm is the maximum feasible level of utility, which is the utility achieved when the road is completely uncongested, so that the trip duration is \(L/v_f\), and when the single commuter departs at his most preferred departure time. This utility is \(U_{\text{max}}(t, L/v_f)\) which we denote by \(u_{\text{max}}\). The inner loops start with an initial guess of the equilibrium \(\bar{t}\) for the specified level of utility. For \(u_{\text{max}}\), the equilibrium \(\bar{t}\) is obviously \(U_{\text{max}}(t, L/v_f)\).

Once a utility function that satisfies the conditions laid out in Section 2 is specified, the \(G\)-functions can be determined exclusively without any other computation. By definition, the \(G\)-functions are only the product of the derivative of the trip duration function, which is just an inversion of the utility function. At this point, for any fixed value of \(\bar{t}\) and \(u\), we can determine exactly how many cycles of entries will be present.

One could start with an arbitrary fixed value of \(\bar{t}\) and \(u\), but there would be no guarantee that \((4.3)\) would be satisfied for the exogenous value of \(L\). At the beginning of the algorithm, we compute the maximum level of utility corresponding to one commuter who travels at free flow speed for distance \(L\). This maximum level of equilibrium utility, \(u_{\text{max}}\) provides a natural starting point for the inner loop to begin, as it provides an upper bound on the feasible equilibrium utility levels. Since only one commuter entered the system, we also know there exists a partial cycle of entries for the equilibrium utility \(u_{\text{max}}\). Now how to find the endogenous time of the first departure, \(\bar{t}\)? The inner loop provides us that solution.

### 4.2. The inner loop

The inner loop solves for the equilibrium entry rate function conditional on \(t\) and \(u\), and updates the value of \(\bar{t}\) until the trip length condition \((4.3)\) is satisfied within a specified tolerance. Thus to obtain the full equilibrium solution conditional on \(u\), we need to determine, for each \(u\), the \(\bar{t}\) or \(\bar{t}'\) that is consistent with the exogenous trip length, \(L\). For clarity, the functional dependence on \(t\) and \(u\) will be dropped until further notice.\(^6\)

Define the function

\[
H(t, u; L) = \int_{L}^{L+T(t; u)} v(\tilde{k}(t_1(t; u))) \, dt - L. \tag{4.15}
\]

where \(\int_{L}^{L+T(t; u)} v(\tilde{k}(t_1(t; u))) \, dt\) is the algorithm’s computed trip distance, \(L\), which gives the distance traveled by the first commuter to depart, conditional on \(t\) and \(u\). The function \(\tilde{k}(t_1(t; u))\) is defined to be the density at time \(t_1(t; u)\) conditional on the exogenous utility level and the exogenous time of the first entry, computed via the algorithm. By construction, \(L\) is the same as the distance traveled by all other commuters. Therefore the function \(H(t, u; L) = 0\) implicitly defines the set

\(^6\) In full notation, we would write for example, the velocity function as \(v(k(t_1(t; u); t; u))\). In the simplified notation, we drop the second two arguments, so the functions will be written as \(v(k(t_1(t; u)))\). This change occurs in the density, entry, and exit rate functions.
of \((t, u)\) consistent with trip distance equaling the exogenous trip distance \(L\). We shall refer to \(H(t, u; \cdot) = 0\) simply as the \(H\)-function. The inner loop computes the \(H\)-function\(^7\) and continues until the value of (4.15) is within the chosen tolerance, for the specified exogenous trip length \(L\).

With the inner loop refinement mechanism defined, we now develop a method to construct the endogenous velocity, density, and entry rate functions to compute (4.15).

Differentiate (4.5) with respect to \(t\):

\[
v'(k(t_{i+1}(t; u))) = \frac{k(t_{i+1}(t; u))}{k_{ij}}.
\]

Using (4.2), this simplifies to

\[
v'(k(t_{i+1}(t; u))) = \frac{k(t_{i+1}(t; u))}{k_{ij}} \left(1 + \tilde{T}(t_{i+1}(t; u))\right) + v(k(t_{i+1}(t; u))) \tilde{T}(t_{i+1}(t; u)) t'(t; u) = 0.
\]

(4.16)

Eq. (4.5) relates equilibrium velocity at \(t_{i+1}(t; u)\) to that at \(t_{i}(t; u)\). Eq. (4.17) therefore relates equilibrium acceleration at \(t_{i+1}(t; u)\) to that at \(t_{i}(t; u)\). For this reason, we refer to (4.17) as the acceleration condition. Eq. (4.17) is a delay differential equation relating \(k(t_{i}(t; u))\), \(k(t_{i+1}(t; u))\), \(k(t_{i+1}(t; u))\), and \(k(t_{i+1}(t; u))\). Since given \(u\), the values of \(T(t; u)\), \(T(t; u)\), and \(\tilde{T}(t; u)\) are known for any value of \(t\). In moving from (4.3) to (4.4), a constant of integration was dropped. In moving from (4.5) to (4.16), an additional constant of integration is dropped. The dropped constant of integration is added back in later, in (4.28).

For the rest of the paper, in order to simplify the algebra we shall assume that the congestion technology takes the form of Greenshields’ Relation, which specifies a negative linear relationship between velocity and density:

\[
v(k) = v_f \left(1 - \frac{k}{k_j}\right),
\]

where \(v_f\) is free-flow velocity and \(k_j\) is jam density. This simplifies the algebra since the \(v'(k(\cdot))\) terms in (4.17) equal \(-v_f/k_j\), a constant. We also make two normalizations, setting \(v_f = 1\) and \(k_j = 1\). Hence measuring velocity as a ratio of free-flow velocity and density as a ratio of jam density. With these simplifications and substituting out the \(k\)'s using (2.5), (4.17) becomes

\[
-(e(t_{i+1}(t; u)) - x(t_{i+1}(t; u))) (1 + \tilde{T}(t_{i+1}(t; u))) + v(k(t_{i+1}(t; u))) \tilde{T}(t_{i+1}(t; u)) + e(t_{i+1}(t; u)) - x(t_{i+1}(t; u)) = 0.
\]

(4.19)

Substitute (2.3) into (4.19):

\[
-e(t_{i+1}(t; u)) (1 + \tilde{T}(t_{i+1}(t; u))) + e(t_{i+1}(t; u)) (1 + \tilde{T}(t_{i+1}(t; u))) + v(k(t_{i+1}(t; u))) \tilde{T}(t_{i+1}(t; u)) + e(t_{i+1}(t; u)) - \frac{e(t_{i+1}(t; u))}{1 + \tilde{T}(t_{i+1}(t; u))} = 0.
\]

(4.20)

Define

\[
B_i(t; u) \equiv (1 + \tilde{T}(t_{i+1}(t; u))) > 0 \quad \text{and} \quad C_i(t; u) \equiv -\tilde{T}(t_{i+1}(t; u); u) > 0,
\]

(4.21)

which are exogenous functions that can be derived from the utility function. Under our assumptions, both are strictly positive. Suppressing the explicit dependence of the \(B_i\)’s and \(C_i\)’s on \(u\), and collecting terms in \(e(t_{i+1}(t; u))\), rewrite (4.20) as

\[
-e(t_{i+1}(t; u)) B_i(t; u)^2 + e(t_{i+1}(t; u)) (1 + B_i(t; u)) - v(k(t_{i+1}(t; u))) C_i(t; u) - \frac{e(t_{i+1}(t; u))}{B_i(t; u)} = 0.
\]

(4.22)

Since the \(B_i\) and \(C_i\) functions are exogenous, and since the \(v(\cdot)\) function is the equilibrium velocity function, (4.22) indicates how the entry rates at corresponding points in time in three contiguous cycles must be related in order to satisfy the acceleration condition. When we are solving for entries in a particular cycle \(i\), we shall use the short-hand notation \(e(t_{i+1}(t; u))\), letting \(e(t_{i+1}(t; u))\) denote the entry rate at the corresponding time in the previous cycle, and \(e(t_{i+1}(t; u))\) denote the entry rate at the corresponding time in the next cycle.

Eq. (4.22) defines a system of equations. There are two cases to consider. In case A, there is only one partial cycle. In case B, there is at least one full cycle of entries, corresponding to the situation where there exist two subcycles for each full cycle, as discussed earlier. It should also be mentioned that there are analytical results that can be obtained for Case A, which will be discussed after Cases A and B.

- **Case A: One partial cycle** In this case, since \(i = 1\), with \(i = 1\), there are entries in the first subcycle of cycle 1, but there are no entries with \(i = 0\) and \(i = 2\), so that (4.22) reduces to

\[
\frac{e(t_{i+1}(t; u)) (1 + B_i(t; u)) - v(k(t_{i+1}(t; u))) C_i(t; u)}{B_i(t; u)} = 0.
\]

(4.23)

\(^7\) The algorithm runs based on the estimated value of \(t\) as stated earlier. Therefore, the computed trip distance \(L\) is almost never the exogenous trip length \(L\). The construction of the \(G\)-functions described in Section 4.1 allow for a good starting value of \(t\) for the algorithm, and allowing us to know when to switch from one partial cycle to one full and one partial entry cycle, etc. The loop runs until the computed value is within some tolerance of \(L\).
Since density is zero at $t$, we have that $e(t_1(t; u)) = v_f \frac{C_1(t; u)}{(1 + B_1(t; u))B_1(t; u)} > 0$. We now time step forward in increments of $\Delta$. At time $t + \Delta$, the number of entries that have occurred is approximately $e(t_1(t; y))\Delta$. This is the density of traffic on the road at $t + \Delta$, $k(t_1(t + \Delta; y))$, which is strictly positive. We then use (4.23) with $t = t + \Delta$ to calculate $e(t_1(t; u) + \Delta)$, which is strictly positive, and the process is repeated until $t' = t_i(t'; u)$ is reached. The value of $t'$ is determined by how the dynamics affect the last cycle. Recall that $\tilde{\tau}$ is the time of the last exit, which is defined by the right-hand boundary condition described in Section 4.1. The value of $\tilde{\tau}$ is the $t$ that satisfies $\nu(k(\tilde{\tau} + \tilde{\tau}(t; y))) = \nu(k(t_{i+1}(t'; y))) = v_f$, which relates the time $\tilde{\tau}$ in the $(I + 1)^{st}$ cycle to the value of $t'$ in the first cycle. This also provides the time of the last entry, $t^* = t_i(t'; u)$ in the $I^{th}$ cycle. Therefore, the algorithm time-steps forward until the right-hand boundary condition is satisfied in the $(I + 1)^{st}$ cycle, thus the values of $t', \tilde{\tau}$, and $t^*$ can be computed.

Now define $\hat{e}(t_i(t; y))$ to be the entry rate at time $t_i(t; y)$ conditional on the exogenous utility level and the exogenous time of the first entry, computed per this procedure. We can then write

$$\hat{e}(t_i(t; y))(1 + B_1(t; y)) - \nu(k(t_i(t; y)))\frac{C_1(t; u)}{B_1(t; y)} = 0,$$

from which it follows that the entry rate is strictly positive over $\mathcal{D}$. The entry rate also increases discontinuously from zero at $t$, and then decreases discontinuously to zero at $t = t_i(t'; u)$.

**Case B: Greater than one full cycle** In this case, since $I > 1$, there are entries in cycles 1 through $I$. Using the cycle structure presented in Fig. 4.1, there exists a natural way to separate the problem into two parts. We will restructure (4.22) to be used for the set of first subcycles, then for the set of second subcycles.

Recall that the first subcycles occur in the subcycles between the points $t_i(t; y)$ and $t_i(t'; y)$, so there are a total of $I$ subcycles, in which there are entries over all of the first subcycles for $i = 1, 2, \ldots, I$. In the $I^{th}$ first subcycle, $I = 1$, there are no entries before $t = t_i(t; y)$, therefore (4.22) reduces to

$$-e(t_1(t; y))B_1(t; y)^2 + e(t_1(t; y))(1 + B_1(t; y)) - \nu(k(t_2(t; y)))C_1(t; y) = 0.$$

In the last of the first subcycles where $i = 1$, the opposite situation is true, i.e. there are no entries after $t^* = t_i(t'; y)$. Thus (4.22) reduces to

$$e(t_i(t; y))(1 + B_1(t; y)) - \nu(k(t_i(t; y)))C_1(t; y) - \frac{e(t_{i-1}(t; y))}{B_{i-1}(t; y)} = 0.$$

For the intermediate first subcycles $i = 3, \ldots, I - 1$, the full form of (4.22) is applied. Defining $A_i(t; y) = k(t_i(t; y))$, the system of entry rate equations for the first subcycles, (4.22), (4.25), and (4.26) can be written in matrix form as:

$$
\begin{bmatrix}
I_1 + B_1 & -B_2^0 & 0 & 0 & 0 & \cdots & 0 \\
-B_1^0 & I_1 + B_2 & -B_2^0 & 0 & 0 & \cdots & 0 \\
0 & -B_1^0 & I_1 + B_3 & -B_3^0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \frac{1}{B_{i-2}} & (1 + B_{i-1}) & -B_{i-1}^0 \\
0 & 0 & \cdots & 0 & 0 & \frac{1}{B_{i-1}} & (1 + B_i)
\end{bmatrix}
\begin{bmatrix}
\nu(A_1)C_1 \\
\nu(A_1)C_2 \\
\nu(A_1)C_3 \\
\vdots \\
\nu(A_{i-2})C_{i-2} \\
\nu(A_{i-1})C_{i-1} \\
\nu(A_i)C_i
\end{bmatrix} =
\begin{bmatrix}
e(t_1(t; y)) \\
e(t_2(t; y)) \\
e(t_3(t; y)) \\
\vdots \\
e(t_{i-1}(t; y)) \\
e(t_i(t; y))
\end{bmatrix}$$

(4.27)

Note that (4.4) may be rewritten as

$$\nu(A_{i+1}(t; y))B_i(t, y) = \nu(A_i(t; y)) \quad \text{for} \quad i = 1, \ldots, I.$$

Applying this equation recursively gives the compact version of (4.6)

$$\nu(A_{i+1}(t; y)) = \frac{\nu(A_i(t; y))}{f_i(t; y)} \quad \text{for} \quad i = 1, \ldots, I,$$

(4.28)

where

$$f_i(t; y) = \prod_{j=1}^i B_j(t; y).$$

(4.29)

---

8 Note that $i = 2$ is not included in the intermediate subcycles case. For this special case, $I = 2$, and only (4.25) and (4.26) are required, so then $i = I = 2$ corresponds to (4.26) being used in the second cycle. Thus the $2 \times 2$ matrix of (4.27) is applied for the first subcycles.
Substituting (4.29) into (4.27), and substituting the numerical approximation \( \hat{e}(t_i(t; y)) \) for the true entry rate \( e(t_i(t; y)) \) gives
\[
\begin{bmatrix}
(1 + B_1) & -B_1^2 & 0 & 0 & 0 & \ldots & 0 \\
- \frac{1}{B_1} & (1 + B_2) & -B_2^2 & 0 & 0 & \ldots & 0 \\
0 & - \frac{1}{B_2} & (1 + B_3) & -B_3^2 & 0 & \ldots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & - \frac{1}{B_{i-2}} & (1 + B_{i-1}) & -B_{i-1}^2 \\
0 & 0 & \ldots & 0 & 0 & - \frac{1}{B_{i-1}} & (1 + B_i)
\end{bmatrix}
\begin{bmatrix}
\hat{e}(t_1(t; y)) \\
\hat{e}(t_2(t; y)) \\
\hat{e}(t_3(t; y)) \\
\vdots \\
\hat{e}(t_{i-1}(t; y)) \\
\hat{e}(t_i(t; y))
\end{bmatrix}
= \begin{bmatrix}
\frac{\nu(A_1)C_1}{\hat{R}_1} \\
\frac{\nu(A_1)C_2}{\hat{R}_2} \\
\frac{\nu(A_1)C_3}{\hat{R}_3} \\
\vdots \\
\frac{\nu(A_1)C_{i-1}}{\hat{R}_{i-1}} \\
\frac{\nu(A_1)C_i}{\hat{R}_i}
\end{bmatrix}
\] (4.30)

Conditional on \( \hat{t} \), the entry rate function over the first subcycle of the first cycle, \( \hat{e}(t_1(t; y)) \), the corresponding density function, \( \tilde{k}(t_1(t; y)) \), and the corresponding velocity function, \( \nu(\tilde{k}(t_1(t; y))) \), may be computed by time-stepping forward in \( t \) until the appropriate boundary condition is satisfied. The boundary condition is \( \nu(t_{i+1}(t'; y)) = \nu_f \), i.e. the velocity reaches free-flow velocity in the (\( i+1 \))st cycle. We can then locate the value of \( t_1(t'; y) \) knowing \( t_{i+1}(t'; y) \), via (4.1).

Note that the computation for the entry rates requires the discretization of time over only the first and second subcycles of the first cycle. The points in time for the subsequent cycles \( i = 2, \ldots, l+1 \) are automatically determined by relation (4.1). So then, the functions \( \hat{e}(t_i(t; y)) \), \( \tilde{k}(t_i(t; y)) \), etc. are also determined at these points. The computations for the set of first subcycles is complete.

Now we turn to the case for the set of second subcycles. The second subcycles occur in the subcycles between the points \( t_i(t'; y) \) and \( t_{i+1}(t; y) \), so there are a total of \( l-1 \) second subcycles, in which there are entries over all of the second subcycles for \( i = 1, 2, \ldots, l-1 \). In the 1st second subcycle, \( i = 1 \) for \( l > 2 \), there are no entries before \( \hat{t} \), therefore (4.22) reduces to (4.25). In the special case when \( l = 2 \), then there is only one second subcycle, which corresponds to (4.23), which is equivalent to the 1 \( \times \) 1 case of (4.32). For the case when \( l = 3 \), the same reasoning for the equations used in footnote 9 for the \( l = 2 \) first subcycles is used, hence the \( 2 \times 2 \) case of (4.32) is applied. In the last of the first subcycles where \( i > 3 \) and \( i = l-1 \), there are no entries after \( \hat{t} = t_i(t'; y) \), then
\[
\hat{e}(t_{i-1}(t; y))(1 + B_{i-1}(t; y)) - \nu(\tilde{k}(t_{i-1}(t; y)))C_{i-1}(t; y) - \hat{e}(t_i(t; y)) \frac{1}{B_{i-2}(t; y)} = 0
\] (4.31)
is used. For the intermediate second subcycles \( i = 4, \ldots, l-2 \), the full form of (4.22) is used. In terms of the matrix equation (4.30), the \( i^{th} \) row and \( i^{th} \) column are deleted, yielding
\[
\begin{bmatrix}
(1 + B_1) & -B_1^2 & 0 & 0 & \ldots & 0 \\
- \frac{1}{B_1} & (1 + B_2) & -B_2^2 & 0 & \ldots & 0 \\
0 & - \frac{1}{B_2} & (1 + B_3) & -B_3^2 & \ldots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & - \frac{1}{B_{i-2}} & (1 + B_{i-1})
\end{bmatrix}
\begin{bmatrix}
\hat{e}(t_1(t; y)) \\
\hat{e}(t_2(t; y)) \\
\hat{e}(t_3(t; y)) \\
\vdots \\
\hat{e}(t_{i-1}(t; y))
\end{bmatrix}
= \begin{bmatrix}
\frac{\nu(A_1)C_1}{\hat{R}_1} \\
\frac{\nu(A_1)C_2}{\hat{R}_2} \\
\frac{\nu(A_1)C_3}{\hat{R}_3} \\
\vdots \\
\frac{\nu(A_1)C_{i-1}}{\hat{R}_{i-1}}
\end{bmatrix}
\] (4.32)

Conditional on \( \hat{t} \), the entry rate function over the second subcycle of the first cycle, \( \hat{e}(t_1(t; y)) \), \( \tilde{k}(t_1(t; y)) \), and \( \nu(\tilde{k}(t_1(t; y))) \), may be computed by time-stepping forward for all \( t \) values in the interval \( (t_1(t'; y), t_2(t'; y)) \).

Thus, we have derived the entry rate function conditional on \( \hat{t} \) and \( y \). We have done this without imposing the restriction that the entry rate be non-negative over \( D \). It is proved in Buli and Arnott (2017) that this restriction is implied by the structure of the equation system (4.30) and (4.32).

Consider the case of one full cycle followed by a partial cycle. At \( t_1(t'; y) \) the entry rate increases discontinuously from zero; at \( t_2(t'; y) \) the entry rate changes discontinuously as the equation system determining the entry rate switches from having one rather than two equations; at \( t_2(t'; y) \) the exit rate increases discontinuously from zero, and to maintain

\[ A \text{ special case which occurs when there are only full cycles, only requires the computation of (4.30), as } t' \text{ will be a breakpoint. This situation is discussed in Section 4.4.} \]
the continuity of the velocity function required by the equal-trip-cost condition, the entry rate changes discontinuously too. And so on. Thus, discontinuities in the entry and exit rates at the primary breakpoints between entry cycles, and between subcycles within each cycle (secondary breakpoints) are properties of the solution, despite the smoothness of the congestion and utility functions.

- **Analytical Solution for Case A** Recall Case A, where there are only entries in first cycle (including the case of one full cycle). The entry rate is governed by (4.23), which can be rewritten in terms of cumulative entries, $E(t)$, to give the ODE

$$
\dot{E}(t) = \left(1 + B_1(t; u)\right)C(t; u)\left(1 - E(t)ight)(1 + B_1(t; u)) - \left(1 - E(t)ight)C(t; u)\left(1 + B_1(t; u)\right) = 0,
$$

where the normalized Greenshields’ Relation and $k(t) = E(t)$ are used. The latter holds, since there are no exits in the first cycle. The ODE (4.33) has the exact solution in terms of the original exogenous functions:

$$
E(t) = \begin{cases} 
1 - \frac{B_1(t; u)}{E(t^*)} & \text{for } t \in [t_1(t; u), t_1(t' = t^*; u)] \\
\frac{C_1(t; u)}{1 + B_1(t; u)^2} & \text{for } t \in [t_1(t' = t^*; u), t_2(t' = u)],
\end{cases}
$$

$$
e(t) = \begin{cases} 
\frac{C_1(t; u)}{1 + B_1(t; u)^2} & \text{for } t \in [t_1(t; u), t_1(t' = t^*; u)] \\
0 & \text{for } t \in [t_1(t' = t^*; u), t_2(t' = u)],
\end{cases}
$$

where the constant $\tilde{C}$ is found using the initial condition $E(\xi) = 0$ is given by

$$
\tilde{C} = \frac{1 + B_1(t; u)}{B_1(t; u)}.
$$

Recall that in the case of one partial cycle, $t_1(t' = u) = t^*$. Using (2.4), $k(t_1(t; u)) = E(t_1(t; u); y)$, and $k(t_2(t; u)) = E(t_2(t; u); y) - X(t_2(t; y))$. Since $E(t; y)$ is cumulative entries, $E(t_2(t; u); y) = E(t^*; y)$. Therefore, density in the second cycle is given by $k(t_2(t; y)) = E(t^*; y) - X(t_2(t; u))$, where $X(t_2(t; u)) = E(t_1(t; u))$ by (2.2). The velocity function is easily determined through the normalized version of (4.18): $v(t; y) = 1 - k(t; y)$. Finally, we can compute the normalized population as a function of the equilibrium utility level, $N(y)$, by using (2.7), which yields:

$$
N(y) = \tilde{C}\left[\frac{B_1(t; u)}{1 + B_1(t; u)} - \frac{B_1(t'; u)}{1 + B_1(t'; u)}\right].
$$

Note that the solution presented here is general, as a specific utility function was not used; it need only satisfy the conditions presented in Section 2. When the number of cycles is greater than one, the endogenous delay is present during the first cycle via (4.25), and thus the delay differential equation is recovered, as in Case B. Therefore, the same kind of computation is no longer possible.

Before moving to the outer loop, we make a number of remarks.

4.3. Remarks

1. We have employed the boundary condition that traffic density is zero immediately before the first commuter departs $\{v(k(t^*)) = \nu_f\}$ in numerically solving the function $\nu(k(t_1(t)))$. We have also employed the boundary condition that traffic density is zero immediately after the last commuter exits $\{v(k(t^* + t; t^*; y)) = \nu_f\}$ in solving for $t^*(t; u) = t_2(t'; u)$. The derivations of the previous section drew heavily on (4.4), the velocity condition, and its derivative, (4.17), the acceleration condition. They did not, however, employ (4.3), the distance condition, from which both (4.4) and (4.17) are derived. Since (4.4) is derived through time differentiation of (4.3), a constant of integration is lost from in moving from (4.3) to (4.4). That constant of integration is the exogenous trip distance, $L$, which is used to obtain a complete solution of equilibrium, conditional on $u$. The previous subsection solved for the unique equilibrium, conditional on $t$ and $y$, which corresponds to some trip distance, but not generally the exogenous trip distance. Bull and Arnott (2017) prove that, conditional on $\xi$, there is a one-to-one mapping from $t$ to $L$, a result that is used in the next subsection. Thus, there is a unique equilibrium condition on $y$ and $L$. Calculating the population corresponding to this equilibrium generates the function $N(y; \xi)$. The equilibrium or equilibria with an exogenous trip distance and an exogenous population then correspond to utility levels that solve $N = N(y; L)$. We shall see in the next section that there may be two utility levels corresponding to an exogenous population and trip distance, one corresponding to aggregate congestion, the other to aggregate hypercongestion.

2. This is a convenient point for a digression on the relationship between the utility function employed in this paper and the user cost function. The easiest way to relate them is to assume, first, that the utility function $U(t, T(t))$ is fact a subutility function of a mixed direct-indirect total utility function having $t, T(t)$, and $y$, income, as its arguments: $V(t, T(t), y)$, and, second, that the total utility function is additively separable between $U(t, T(t))$ and $y$. Then $V(t, T(t), y) = U(t, T(t)) + y$. Since the marginal utility of income is then one, $U(t, T(t))$ is measured is money units. Now define $v$ to be the utility that would be achieved if trip duration were zero and if utility were maximized with respect to $v$: 

$$
\dot{E}(t_1(t; u))\left(1 + B_1(t_1; u)\right) - \left(1 - E(t_1(t; u))\right)\dot{C}_1(t_1(t; u))B_1(t_1(t; u)) = 0,
$$

(4.33)
\[ v = \max_t U(t, 0). \] Then \( v - U(t, T(t)) \) equals the money-metric loss in utility from a commuting trip with travel duration \( T \) and departure time \( t \) compared to a commuting trip of zero duration and departure at the optimal time. Define this to be the user cost. Since all commuters experience the same utility in equilibrium, \( u \), one may define \( \bar{c}(u) = v - y \) to be the equilibrium user cost. Then this function and the function \( N(u) \) together give a relationship between user cost and population, which is the user cost “function” \( c = c(N) \).

3. The solution of the last subsection did not take into account several inequality conditions that equilibrium must satisfy: (i) neither the entry rate nor the exit rate can be negative; (ii) traffic density cannot be negative; and (iii) utility be below the equilibrium utility level everywhere outside the departure interval. Buli and Arnott (2017) have shown that the mathematical structure of the problem implies that the entry rate is strictly positive over the departure interval, which implies via (2.3) and (4.4) that the exit rate is strictly positive from the first exit to the time of the last exit. Since cumulative exits lag cumulative entries, since cumulative entries are strictly positive over the interior of the departure interval, and since the number of entries over the rush hour equals the number of exits, density cannot be negative. By construction, the entry rate function solved for in the previous subsection, which depends on \( f \) and \( y \), has the property that utility is constant over the departure interval. Furthermore, it is shown in Buli and Arnott (2017) that the utility is everywhere lower outside the departure interval, so that (2.1) is satisfied.

4. The dynamic structure of the problem is unfamiliar. If traffic density were instead determined as the solution to an ordinary differential equation, the equation of motion would have the form \( k(t) = f(e(t), k(t)) \), and the solution would be obtained by combining this equation of motion with appropriate boundary conditions on functions of the state variable. Here instead, the equation of motion takes the form \( k(t + T; y) = e(t + T; y) - x(t + T; y) - e(t + T; y) - e(t)/(1 + T(c; y)). \) The boundary conditions are unfamiliar too. In the simplest form of a delay differential equation, the boundary condition is an exogenous initial history, the analog of which here would be the entry function over the first cycle. But in the problem at hand, the initial history has to be solved for by the conditions that the first entry cycle generates a final exit cycle with the property that all the commuters exiting during the final exit cycle satisfy the equilibrium trip-timing conditions. To see this, consider starting with an arbitrary non-negative entry function over the first cycle. The entry function over the first cycle determines the exit function over the second cycle, and the entry function over the second cycle is chosen so that all commuters in the first cycle satisfy the trip timing condition. This process continues until the last entry, but then there are no subsequent entries to choose to satisfy the equilibrium trip-timing condition for commuters still traveling at this time. Considering the complexity of the dynamic structure of the problem, it is perhaps surprising that the solution procedure is as “simple” as it is.

4.4. The outer loop

Recall Walters’ analysis of steady-state traffic flow with MFD congestion. It solves for the user cost and flow as a function of traffic density, and then plots user cost against flow as density increases, as shown in Fig. 3.1, obtaining the user cost curve. Equilibrium is then characterized as points of intersection of the demand curve and the user cost curve.

Our procedure here is analogous, except that we solve for user cost and the commuter population as a function of utility, and then plot user cost against commuter population to obtain the user cost curve, as shown in Fig. 5.4. In Walters’ construction, there is always a backward-sloping portion of the user cost curve, which corresponds to hypercongestion. In our construction, depending on the form of the utility function, there may or may not be a backward-sloping portion of the user cost curve, which corresponds to aggregate hypercongestion.

The possibility of aggregate hypercongestion has not, to our knowledge, been noted before, but is not difficult to understand intuitively. In our central example, the utility function is such that the earliest feasible departure from home occurs at 5 am, while the latest feasible arrival at work occurs at 11 am. These two constraints, along with the exogenous trip length, imply a maximum feasible commuter population that can be accommodated over the rush hour. Now suppose that demand is high and price-sensitive. The only way that equilibrium can be achieved is with a user cost that is higher than that consistent with congested traffic flow.

When aggregate hypercongestion is a possibility, the number of equilibria associated with an exogenous population is a priori unknown, which creates numerical problems. However, even with the possibility of aggregate hypercongestion, there is at most one equilibrium consistent with an exogenous user cost or utility level. For this reason, we generate the user cost curve by starting out with the maximum level of utility and then incrementally lowering utility for each utility level, determining the population associated with the corresponding unique equilibrium.

We proceed as stated above, by incrementally lowering \( y \) from \( u_{\max} \), and for each \( u \) solving for the \( f \) satisfying (4.15) using the procedure described in Section 4.2. At each equilibrium utility level, \( u \), the population is computed by using (2.7). The outer loop repeats as long as the number of cycles remains constant. The number of cycles is computed at the beginning of the loops, as discussed in Section 4.1. Once the number of cycles changes, the matrices in Section 4.2 either increase or decrease by one dimension in the number of equations and the number of variables, i.e. a row and column is either added or deleted. The matrix size increases (decreases) when the number of cycles increases (decreases). Details of the G-functions and H-function are outlined in the following section.
4.5. Details of the G-functions and H-function

The procedure developed in the inner loop to compute the integral term in (4.15) requires knowing the number of entry cycles over the departure set. At the end of Section 4.1, it was shown how to solve for the number of entry cycles for a particular \((t, y)\). But to perform this exercise for each guess of the value of \(t\) that solves (4.15) for a particular \(y\), and to repeat this for each value of \(y\) investigated is inefficient. This inefficiency led to the development of the G-functions and H-function.

Recall the shorthand notation from the previous subsections:

\[
F_i(t) = \prod_{i=1}^{t} B_i(t, u) \quad \text{and} \quad B_i = 1 + \bar{T}(t_i(t; y)).
\]

We can rewrite the G-functions (4.14) in terms of the shorthand notation, specifically

\[
G_i(t, u) = F_{i+1}(t, u) - 1 = \left[ \prod_{i=1}^{t} B_i(t, u) \right] - 1 = \left[ \prod_{i=1}^{t} (1 + \bar{T}(t_i(t, u); y)) \right] - 1, \tag{4.35}
\]

so that \(G_i(t, y) = 0\) is the set of \((t, y)\) for which there are exactly \(l\) full entry cycles in equilibrium, as stated before.

Two general properties of the G-functions bear note.

1. Recall that the G-functions are defined without reference to trip length. Suppose that \(U = \max_{t} U(t, 0) \equiv v\). This can only be achieved with zero trip length. But with zero trip length, everyone can travel at the utility-maximizing time however many entry cycles there are. Thus, the G-functions share the point \((v, \arg \max_{t} U(t, 0))\) in common.

2. As \(u\) decreases from \(v\), in equilibrium the number of full entry cycles increases to a maximum number of cycles and then decreases. Despite the increasing then decreasing nature of the number of full entry cycles, as \(u\) decreases, the earlier is the time of the first departure.

Fig. 4.2 displays this construction for the central numerical example. Suppose, for example that we have solved for the H-function from \(u = u_{\text{max}}\) to \(u = 10\). and now wish to solve for \(t\) satisfying (4.15) with \(u = 9\). From inspection of the Figure, the solution entails three cycles, two full cycles, followed by a partial cycle. We could then determine our initial guess of the equilibrium value of \(t\) for \(u = 10\). Applying (4.15), taking into account that the solution entails two full cycles followed by a partial cycle, it would be found that the trip distance so computed exceeds the exogenous trip length, which would indicate that the equilibrium value of \(t\) for \(u = 9\) is lower than that for \(u = 10\).

In the central numerical example, both the H-function and the G-functions are positively sloped over the entire economically relevant region of \(\xi - u\) space. These properties are specific to the example. However, Buli and Arnott (2017) shows for both the H-functions and the G-functions that, for every value of \(u\) there is a unique value of \(\xi\) satisfying each function, which is all that is needed for our algorithm to work smoothly.

Since the solution algorithm solves for the entry rate function conditional on \(y\) from the time of the first departure to the time of the last departure, it is straightforward to calculate and record the equilibrium commuter population as a function of \(y\).

Fig. 4.3 gives a flowchart that summarizes the full algorithm employed, including both the inner and the outer loop. The computer program that generates the examples is given at the hyperlink: http://math.ucr.edu/~buli/Bathtub_Code/Bathtub_Code.html. The program follows exactly the algorithm as described in this section. The program can be readily adapted to
treat other utility functions that satisfy the properties assumed in Section 2, as long as they yield closed-form formulas of the $\tilde{T}$ and $\tilde{T}$ function. The program is written to be easy to understand and to avoid numerical difficulties, rather than to be efficient in terms of run time.

5. An extended numerical example

This section demonstrates the implementation of the algorithm in a test case. We generated two extended numerical examples, one for an exponential utility function, and another for an logarithmic utility function. In most of the section, we focus on the extended numerical example with logarithmic utility since it was computationally more straightforward and its results easier to present and explain, but in Section 5.2 we shall briefly contrast the results from the logarithmic utility function with those obtained from the exponential utility function.

Section 5.1 presents the logarithmic utility function and derives and explains its properties, and at the end gives the parameters used in the example. The only parameter that varies between runs is the exogenous population. Section 5.2 reports on the results with the logarithmic utility function for three populations. Section 5.3 briefly contrasts the results obtained with the exponential utility function to those obtained with the logarithmic utility function. Finally, Section 5.4 comments on computational aspects of the extended numerical example.

5.1. The logarithmic utility function

The logarithmic utility function is

$$U(t, T(t)) = r_0 \log(r_1 t) + s_0 \log(s_1(M - (t + T(t)))).$$  

(5.1)
Table 5.1
Logarithmic utility and travel duration functions.

<table>
<thead>
<tr>
<th>Functions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(U(t, T(t)) = r_0 \log(r_1 t) + s_0 \log(s_1 (M - (t + T(t)))))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T(t) = M - t - \frac{1}{s_0} \exp \left( \frac{1}{s_0} \left( u - r_0 \log(r_1 t) \right) \right) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\bar{T}(t) = -1 + \frac{r_0}{s_0} \exp \left( \frac{2}{s_0} (r_1 t) \right) t^{-2} e^{-1} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T(t) = \frac{r_0}{s_0} \left( -\frac{u}{s_0} - 1 \right) \exp \left( \frac{1}{s_0} (r_1 t) \right) t^{-2} e^{-2} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is the sum of two sub-utility functions. The sub-utility function on the left gives the “at home” utility function. A commuter derives infinitely negative utility if he leaves home at \(t = 0\) and thereafter derives diminishing marginal utility from staying at home longer. In particular, his utility from being at home is the log of the time he leaves home. The sub-utility function on the right gives the “at work” utility function. A commuter derives infinitely negative utility if he arrives at work at \(t = M\), and derives diminishing marginal utility from arriving at work earlier. In particular, his utility from being at work is the log of the time he arrives at work earlier than \(M\). The simplicity of the utility function makes its properties easy to derive, and its intuitiveness makes its properties easy to understand.

The most obvious implication of the utility function is that the rush hour cannot start before \(t = 0\) or end after \(t = M\). Since the congestion technology is such that there is a maximum or capacity flow, this puts an upper bound on the population the system can accommodate over the rush hour. Recall that the way the algorithm proceeds is to successively lower the utility level, for each utility level computing the corresponding equilibrium, including the equilibrium population. The process generates a user cost curve. Since there is a strict lower bound on departure time and a strict upper bound on arrival time, the only way a very low equilibrium utility can be achieved is for the median commuter to have a very high trip duration, requiring severe hypercongestion, and hence flow considerably below capacity flow, for most of his trip. Thus, the logarithmic form of the utility function results in a user cost curve with a backward-sloping portion corresponding to aggregate hypercongestion.

Table 5.1 displays the logarithmic utility and travel duration functions. It has all the properties that we have assumed for the general utility function employed in the analysis of the previous section; in particular \(1 + \bar{T}(t; y) > 0\) and \(\bar{T}(t; y) < 0\) for all \(t \in (0, M)\). We also note that, with the logarithmic utility function, \(\bar{T}_i(t; y) > 0\) (which guarantees that the \(G_i(\cdot)\) functions are all positively sloped for all \(i\), \(\bar{T}_g < 0\), and \(\bar{T}(t; y) > 0\). Given the simple form of the utility function, it is not surprising that its properties are so simple.

There is no empirical work on the basis of which to choose the parameters of the utility function. We obtained the parameters instead through experimentation, with the aim of getting reasonable results. They were

\[
r_0 = 15 \quad r_1 = \frac{1}{2} \quad s_0 = 18 \quad s_1 = 1 \quad M = 6.
\]

The units are hours; thus, one may view 5 am as the earliest feasible departure, and 11 am as the latest feasible arrival. In order to simplify the algebra, in the previous subsection results were derived using normalized units. But in presenting the results, we work with unnormalized units:

\[
v_f = 15 \text{ mph} \quad k_j = 10^6 \text{ vehicles} \quad L = 4 \text{ miles}
\]

For interpretation of the results, it will be useful to work out some of the implications of these parameters. As a simple point of reference, consider the situation (which is not consistent with equilibrium) where the entry rate is such that density is at capacity density over the rush hour. With Greenshield’s Relation, capacity density is \(0.50 \times 10^6\), capacity flow is \(3.75 \times 10^6\), and capacity velocity is 7.5 mph. In steady state, the inflow = outflow of the street system equals flow divided by trip length or 0.94 \times 10^6. Since the duration of a trip at the velocity corresponding to capacity flow is 0.53, the maximum feasible duration of the departure interval is 5.47, and the corresponding maximum feasible number of commuters over the rush hour is 5.13 \times 10^6. The maximum equilibrium number of commuters is significantly smaller than this, since in order to achieve equilibrium, velocity must be higher than capacity velocity (and flow correspondingly lower) in the shoulders of the rush hour and lower than capacity velocity (and flow correspondingly lower) during the peak.

Recall that, where \(v = \max \bar{T}(t; 0)\) and \(\bar{v}_\text{max} = \max \bar{T}(t; L/v_f)\), \(v - \bar{v}_\text{max}\) is trip cost at free-flow travel speed. With the assumed utility function, it is simple to calculate\(^{10}\) that \(v = 25.99\) and \(\bar{v}_\text{max} = 24.49\), implying that trip cost at free-flow speed is 1.50 monetary units. Since free-flow travel time equals 4/15 hrs, and since the cost of travel time is around $20/hr, the trip cost in dollars is $3.33, so that the money unit may be taken to be around $3.55.

\(^{10}\) The utility-maximizing departure time with \(T = 0\) solves \(r_0 t - s_0 (M - t) = 0\) yielding \(t = r_0 M / (r_2 + s_2)\), so that \(v = r_0 \log(r_1 r_0 M / (r_2 + s_2)) + s_0 \log(M (1 - s_2 r_0 / (r_2 + s_2))) = 25.99\). The utility-maximizing departure time with \(T = 0.2667\) solves \(r_0 t - s_0 (M - 0.2667 - t) = 0\), yielding \(t = r_0 (M - 0.2667) / (r_2 + s_2)\), so that \(\bar{v}_\text{max} = r_0 \log(r_1 r_0 (M - 0.2667) / (r_2 + s_2)) + s_0 \log((M - 0.2667) (1 - s_2 / M)) = 3.97027 + 20.5226 = 24.493\).
Fig. 5.1. All the plots in Fig. 5.1 are for the case of one partial entry cycle. The outputs for equilibrium are \( u = 24.43473, t = 2.54243, N = 4.11892 \times 10^4 \)

\( t^* = 2.66181 \), and \( f = 2.93739 \). Fig. 5.1a plots the functions \( T(t), \dot{T}(t), \) and \( \ddot{T}(t) \). Fig. 5.1b plots the velocity function, \( v(t) \), over the departure interval. Fig. 5.1c displays the density function, \( k(t) \), over the departure interval. Fig. 5.1d plots the trip length of the individual who leaves at \( t = t^* \). Fig. 5.1e displays the entry and exit rates, \( e(t) \) and \( x(t) \), respectively.

### 5.2. Numerical results with the logarithmic utility function

We present numerical results for three levels of utility, high, medium, and low. The example with high utility has only one partial entry cycle and exhibits only very modest congestion. The example with medium utility has three full entry cycles and one partial entry cycle, and, though corresponding to the congested portion of the aggregate user cost curve, still exhibits quite severe hypercongestion at the peak of the rush hour, with a minimum speed somewhat under 4 mph. The example with low utility corresponds to the hypercongested portion of the aggregate user cost curve. It exhibits very severe hypercongestion over almost the entire rush hour, with a minimum speed of about 1 mph.

- **high utility** (\( u = 24.43, N = 4.12 \times 10^4 \)) Fig. 5.1 displays five panels. The top panel displays \( T(t), \dot{T}(t), \) and \( \ddot{T}(t) \) against clock time. The panel in the middle on the left plots velocity against time, and that on the right density over time; since under Greenshields’ Relation velocity is a negative linear function of density, the two graphs are closely related. The panel at the bottom left gives the distance traveled within successive entry cycles as a function of time; in this example,
there is just one partial cycle.\textsuperscript{11} The panel at the bottom right gives the entry and exit rates over time. Clock time is normalized so that $t = 0$ corresponds to 5 am. Breakpoints are indicated by circles. The vertical axes are scaled to make the results easy to read.

The five panels together tell a simple story of equilibrium rush-hour dynamics in a situation where there is generous capacity relative to the population of commuters, which is $N = 4.11892 \times 10^4$. The rush hour is short, lasting only about 0.4 h or 24 min, and travel exhibits little congestion, with a minimum speed of around 14.4 mph, the free-flow speed being 15 mph.

There is only one partial entry cycle, which starts at $t = 2.542$ and ends at $t^* = 2.662$. Since trip length is 4 miles and speed is around 14.4 mph, trip duration averages around 0.274 hrs; first rising for early commuters to offset the increasingly desirable departure time and then falling for late commuters to offset the increasingly undesirable departure time. Since the trip duration of the first commuter to depart exceeds the period over which commuters depart, there is a period in the middle of the rush hour when all commuters are on the roads, after the last entry but before the first exit, which explains the flat portions of the velocity and density curves in the middle of the rush hour. The entry rate, calculated from (4.24), adjusts so as to ensure that utility is constant over the departure interval; it declines slowly over the departure interval. In contrast, the exit rate increases slowly over the exit interval, as traffic speed increases.

- **medium utility ($u = 20.000, N = 2.246 \times 10^6$)** Fig. 5.2 displays the equilibrium traffic dynamics with the considerably higher population of commuters of $N = 2.246 \times 10^6$, using the same five panels. The vertical axes of some of the panels are scaled differently from the corresponding panels in Fig. 5.1 to improve readability. The equilibrium rush-hour traffic dynamics are markedly different from those of the previous example. The rush hour lasts 2.859 h, extending from $t = 1.353$ to $t = 4.212$, and there are three full entry cycles and one partial entry cycle. The separate cycles are displayed in different colors. Traffic is hypercongested at the peak.

The most striking feature of this case is the discontinuities in the entry and exit rates, particularly since the utility and congestion functions are smooth. The source of these discontinuities was explained in the previous section. Apart from the discontinuities in the entry and exit rates at the breakpoints, the rush-hour traffic dynamics accord with intuition. Velocity falls in the early morning rush hour and increases in the late morning rush hour. As the population of commuters increases, one expects both the rush hour to lengthen and average velocity to fall, and that is what is observed. Also, since average velocity is lower, the duration of the entry cycles is longer.

- **low utility ($u = -25, N = 1.656 \times 10^6$)** This case is displayed in Fig. 5.3. Apart from the extremely low velocity at the peak of the rush hour, indicating severe hypercongestion, a cursory look at the rush-hour traffic dynamics suggests that they are well behaved. But on closer examination, the equilibrium in this case is anomalous.

The rush hour is longer than in the previous example, despite the considerably lower population, and there is one full and one partial entry cycle, whereas in the previous example there were three full and one partial entry cycle. At the start of the rush hour, the entry rate is very high, causing hypercongestion to set in almost immediately. Speed is so low that the full entry cycle is over two hours long, so that even the first commuter to travel has an average speed of less than 2 mph. Because traffic is so slow at the end of the first entry cycle, the exit rate at the start of the second entry cycle is low, requiring only a lower entry rate to satisfy the equilibrium trip-timing condition. The second, partial entry cycle has an even longer duration. Together these results indicate severe hypercongestion at the aggregate level.

Fig. 5.4a plots the aggregate user cost curves, $c(N)$, for the logarithmic utility function. The points plotted in the figure are not the primary and secondary breakpoints described earlier, but instead indicate levels of population at which there is a switch in the number of entry cycles, which we define as switch-points. Letting $(m, n)$ denote a switch from $m$ to $n$ full cycles with movement up the user cost curve, the indicated switch-points are $(0,1)$, $(1,2)$, $(2,3)$, $(3,2)$, and $(2,1)$. The switch-point at the top of the Figure, beyond the plotted portion of the user cost curve is $(1,0)$.

One might think that aggregate hypercongestion is an artifact of the logarithmic utility function, which effectively restricts the length of the rush hour, but Fig. 5.4b shows that aggregate hypercongestion also arises with the exponential utility function that Fosgerau has employed, which places no effective restrictions on the length of the rush hour.

Now add an exogenous population level to the user cost curve. By definition, any intersection of the exogenous population line and the user cost curve, of which there are two, is an equilibrium. The obvious question is which of the two equilibria is stable. The stability of an equilibrium is defined with reference to the adjustment process, but whatever the adjustment process, equilibria alternate between stable and unstable. Here the adjustment process is the day-to-day departure time decision of the individual commuter. With any such process that is reasonable, the congested equilibrium is stable, which implies that the hypercongested equilibrium is unstable. However, stable, hyper congested equilibria can occur if the exogenous population constraint is replaced by a downward sloping demand function, as in Arnott and Inci (2010) but there in the context of equilibriums steady-state traffic flow. It remains to be seen whether rush-hour traffic equilibrium in any of the world’s most congested cities is hypercongested at the aggregate level and stable.

\textsuperscript{11} The derivative of the distance function for each entry cycle gives the velocity function over that entry cycle. Plotting cumulative distance against time would give the time-space diagram for vehicles over the rush hour.
5.3. Numerical results with the exponential utility function

The exponential utility function is

\[
U(t, T(t)) = \left( \frac{A_0}{\delta_t} \right) (1 - e^{-a_1 t}) + \left( \frac{B_0}{B_1} \right) (1 - e^{-b_1 (M - t - T(t))}).
\]

(5.2)

The properties of this utility function are recorded in the companion paper, Buli and Arnott (2017). Suffice it to say here that the exponential utility function satisfies all the conditions imposed on the utility function in the general analysis of the previous section, and has the same qualitative properties as those listed in Table 5.1 for the logarithmic utility function. The main difference between the exponential and logarithmic utility functions is that the logarithmic utility function effectively constrains the length of the rush hour, whereas the exponential utility function does not.
In the numerical example for the exponential utility function, the parameters of the utility function are\textsuperscript{12} $A_0 = 5$, $a_1 = 1/3$, $B_0 = 10$, and $b_1 = 1/2$. They were chosen to give sensible numerical results. The other parameters are the same as those for the logarithmic utility function example.

\textbf{Fig. 5.4a} displays the user cost curve with the logarithmic utility function, and \textbf{Fig. 5.4b} the user cost curve with the exponential utility function. The most interesting feature of the exponential utility function example is that, even though it does not impose a constraint on the length of the rush hour, aggregate hypercongestion nonetheless occurs.

\textsuperscript{12} Hjorth et al. (2015) estimate the parameters. Even though they impose the parameter restriction that $a_1 = b_1$, they find that the parameter estimates have large standard errors and are highly correlated.
Fig. 5.4. Fig. 5.4a displays the user cost curve for the logarithmic utility function. Fig. 5.4b displays the user cost curve for the exponential utility function.

5.4. Computational experience

Many methods were attempted to numerically solve the delay differential equation in (4.17) for the velocity function \( v(\cdot) \) directly, that satisfied the boundary conditions, the equilibrium timing condition, and the trip length condition. At first, we experimented with polynomial interpolation methods, which were applied between breakpoints.\(^\text{13}\) In one of the polynomial methods, we used successive derivatives on (4.3), to get a system of equations, whose number of equations was equal to the number of interpolation points. In a second method, we attempted to partition the time domain between breakpoints by using Chebyshev nodes for polynomial interpolation. Unfortunately, these methods did not provide convergence when the polynomial degree was increased, and the velocity function was discontinuous at the breakpoints.

A global polynomial method was attempted, to create one large polynomial that would interpolate over the whole rush hour domain. We utilized the derivatives of (4.3), which can be calculated, assuming enough regularity on \( v(\cdot) \). This method was also unsuccessful, as there was no apparent convergence, and depending on the number of Chebyshev nodes, large oscillations developed. Only later did we realize that the structure of the delay differential equation implies discontinuities in the entry rate function, hence only a \( C^0 \) velocity function. Polynomial approximation between the breakpoints is also unsatisfactory, since there are discontinuities in the entry rate function at the secondary breakpoints whose location is a priori unknown.

We next looked at a simple time-stepping method. The algorithm we developed solved the problem going forward, with \( y \) fixed and with an initial guess of \( t \), and the approximated velocity function was computed on the first cycle to create the velocity function over subsequent cycles. At that point, the trip length \( L \) is computed. From here, we can determine how to change \( t \) based on if the computed value of \( L \) is greater than or less than \( L_0 \). Then, the utility function \( U(t, T(t)) \) itself is used to compute the new level of utility. If the \( y \) is not within tolerance, the process is repeated. Once the tolerance is satisfied, the \( t \) can be used as the starting point, to run the problem backwards. Then the whole process is repeated in the backwards direction, and repeated forwards and backwards until convergence is achieved. The method employed above also failed to converge.

With the failure of these methods, we gradually developed the solution procedure described in Section 4 that takes full account of the mathematical structure of the problem. Our motive for recounting our earlier failures is to warn that standard numerical methods used to solve smooth ordinary or partial differential equations may not work well when applied to the bathtub model or to other models whose dynamics are described by delay differential equations. The main reason seems to be that standard methods assume smoothness of the solution, but DDE’s generate discontinuities in the solution even when the underlying functions are smooth. Another motive for recounting our earlier failures is positive. Even though the

\(^\text{13}\) The values of the breakpoints are exactly known from the exogenous nature of the \( T(t; y) \) function for a fixed equilibrium value of utility, so the value of the velocity function is known exactly.
bathtub model is analytically intractable, it can be solved numerically without encountering computational difficulties by using customized algorithms that account for its particular mathematical structure.

For all of the computational results, the algorithm was run on a DELL® XPS 13 laptop with an Intel® Core™ i5-5200U CPU and 8 GB memory using MATLAB®. All results were computed in double precision, with loop tolerances set to at least $10^{-14}$ for the loop where $L$ is computed as given in the flowchart, Fig. 4.3. Depending on the number of entry cycles, the increments of utility are taken to be in the range of $\Delta u = 0.00001$ to $\Delta u = 0.01$. Finer scales were used in the congested region of the user-cost curve, and the coarser scales were used in the hypercongested region. For the time increments, we only divide the first cycle between the first two breakpoints, as the points in cycle 1 are mapped to subsequent cycles via the $t_i(t;\ u)$ mechanism. In the numerical work, the number of time increments range from 10,000 to 400,000, where more points are required in the hypercongested region. Special care is taken to choose the number of time increments, and how much the value of $t_i$ is changed in the inner loop as $N$ gets closer to $N_{\text{max}}$. The point at which this occurs is where the $H(t;\ u,\ L)$ function becomes steep, and the number of entry cycles begins to decrease. The steep slope of $H(t;\ u,\ L)$ affects the algorithm, as a small decrease in $t_i$ can cause a large change in $u$. This is the reason we take $\Delta u$ to be coarser, and the time increments to be smaller, to keep the same tolerance of $10^{-4}$.

6. Concluding remarks

6.1. Synopsis

The primary aim of this paper has been to develop a method to solve numerically for equilibrium in the basic (identical individuals) bathtub model when the underlying utility and congestion functions are smooth. The difficulty in doing so results from the unfamiliar mathematics of the bathtub model, which requires the solution of a delay differential equation. Except for special cases, which were reviewed earlier, analytical solution is not possible, so that computation is necessary. When we started our work on the paper, we first went to literature on the computational solution of delay differential equations with an exogenous initial history. In contrast, in the bathtub model the initial history – either the entry function or the density function over the first cycle – has to be solved for as part of the overall equilibrium. Our broad conception was that we would add an outer loop that would iteratively converge to the equilibrium initial history. But we encountered numerical and conceptual difficulties. It was not until we developed a customized algorithm that builds on the mathematical properties of the bathtub model that we had success.

The method that we developed has three elements. The first is to invert the trip-timing equilibrium condition that $U(t,\ T(t)) = u$ over the departure interval to give the function $T(t;\ u)$, which gives trip duration as a function of departure time consistent with the exogenous utility level. The solution entails entry cycles. The first commuter departs at $t_i$, which is the beginning of the first entry cycle, and arises at $t_i + T(t_i;\ u)$, which is the end of the first entry cycle and the beginning of the second, etc. The second element is to derive computationally the entry rate function consistent with the equilibrium trip duration function for a particular $u$ and with a given time of first entry, $t_i$. The entry rate function is solved by: i) twice differentiating the distance condition that, over the departure set, the integral of velocity over the equilibrium trip duration equal the exogenous trip distance, which we termed the acceleration condition; ii) substituting (4.4), (2.3), and (2.5) into the acceleration condition to obtain a system of linear equations relating the entry rate at a point in time to velocity at that point in time and the entry rate at the corresponding points in time in the cycle ahead and the cycle behind; and iii) applying the boundary conditions that traffic density is zero at the times of the first entry and of the last arrival. The solution has the property that all trip distances are the same, but not generally equal to the exogenous trip distance. The third element is to solve for the $t_i$ consistent with the exogenous trip distance. The solution obtains a unique solution for each value of $u$.

While not difficult to understand or to apply, our solution method is complex. We conjecture that it is nonetheless the simplest solution method that avoids computational problems. On one hand, the complexity is discouraging. On the other hand, from the user’s point of view, the computational method is a reliable black box.

We started with the maximum utility level (which is achieved when population is zero) and gradually lowered the utility level, obtaining the equilibrium solution for each utility level. Since there is a unique population of commuters for each utility level, and since user cost can be straightforwardly related to utility, our procedure generates a user cost curve, relating user cost to the commuter population.

6.2. Directions for computational research on the (proper) bathtub model

Our solution method derives the user cost curve computationally. If the aim is to solve for equilibrium for a particular population level, the method is obviously inefficient. The difficulty in solving for equilibrium with an exogenous population level is that the number of equilibria is unknown a priori. Depending on the form of the utility and congestion functions, there are three qualitative possibilities: (i) There is no upper bound on the population of commuters that the street system

---

14 Since the distance between breakpoints is constantly changing, one can hold the number of time increment points to be held fixed, while the $\Delta t$ changes; or one can hold $\Delta t$ fixed and let the number of time increments change. Numerical experimentation has shown both methods to yield similar results.
can accommodate; (ii) there is an upper bound, which equilibrium approaches asymptotically as user cost increases; (iii) there is an upper bound that is achieved at a finite user cost, and higher user costs correspond to populations below this upper bound. The two examples that we presented illustrate possibility (iii). Arnott et al. (2016) illustrates possibility (i) with the $\alpha - \beta - \gamma$ user cost function, and it is natural to conjecture that this asymptotic property applies with smooth approximations of that cost function. Through analysis of the asymptotic properties of the bathtub model as user cost approaches infinity, it should be possible to determine a priori which of the three possibilities arise.

The next item on the research agenda is to solve for the social optimum in the basic bathtub model. Formally, the optimum problem can be stated in a form that differs from the equilibrium only in that (2.1), the trip-timing equilibrium condition, is replaced by the condition that the marginal social cost of a trip is the same throughout the departure interval and weakly higher outside the departure interval. Via the Envelope Theorem, marginal social cost at a particular departure time can be calculated as the increase in total cost from inserting a commuter at that time, holding fixed the entry rate function. The difficulty lies in obtaining a manageable expression for the marginal social cost function. Since there will be entry cycles in the social optimum, discontinuities in the entry rate function should be expected, which will likely create difficulties for the application of generic numerical optimization procedures that assume smoothness.

Earlier, in the literature review, we briefly discussed the progress that Fosgerau (2015) and Lamotte and Geroliminis (2017) have made in solving for equilibria with a continuum of commuters, who differ in trip length or preferences. For the situation where the departure and arrival order of commuters are both monotonic with respect to the heterogeneity parameter (trip length or a taste parameter), Fosgerau (2015) has developed an algorithm that obtains the equilibrium by numerically solving a pair of ordinary differential equations. However, little progress has been made in developing an algorithm to solve for equilibrium when the monotonicity condition is not satisfied.

6.3. Whither the (proper) bathtub model?

There is increasing evidence that in the downtown areas of almost all major metropolitan areas, hypercongestion (traffic jam situations in which exit flow decreases as traffic density increases) is pervasive during the peak of the rush hour, and that a large proportion of the delay due to traffic congestion occurs in hypercongested traffic. Yet the workhorse model of rush-hour traffic dynamics of both transportation scientists and urban transportation economists has been Vickrey’s bottleneck model, which excludes hypercongestion by assumption (assuming instead that the exit flow is constant over the rush hour). There are now several papers that develop alternative models of equilibrium and optimal rush-hour traffic dynamics that incorporate hypercongestion. All these models have encountered mathematical difficulties that preclude straightforward solution. This paper has considered the most popular of these models, the basic bathtub model, which incorporates MFD flow congestion over an isotropic downtown area with identical commuters. To circumvent the mathematical difficulties, most of the papers in this branch of the literature have made approximating assumptions. The difficulty of such assumptions is that, without solution of the “proper” bathtub model, it is not possible to gauge their accuracy or the biases they introduce.

This paper has taken the bull by the horns, developing a procedure to solve numerically for equilibrium in the proper basic bathtub model. The procedure is not simple, but, by building on the mathematic structure of the model, it is reliable and accurate. Now that the equilibrium rush-hour traffic dynamics of the proper, basic bathtub model can be solved for, it will be possible to evaluate the accuracy and the biases of the bathtub models that make approximating assumptions.

The way forward is unclear. On one hand, progress working with the proper basic bathtub model has been so slow and difficult that it is hard to see it ever being widely used or applied in practice. On the other hand, the transportation community needs models that take hypercongestion into account. The efficiency costs of hypercongested traffic are likely already substantial in most of the world’s major metropolitan areas, and in developing and middle income countries, due to rapid urbanization, sharp increases in auto ownership, and inadequate fiscal capacity to fund the much-needed transportation infrastructure, these costs are likely to grow substantially.

To incorporate hypercongestion into transportation policy analysis will require models that can be readily applied in practice and that account for the richness and complexity of real-world traffic. Bathtub models that make approximating assumptions are so much more tractable than the proper bathtub model that they should be able to accommodate the many extensions in the direction of realism that will be needed.

A generic problem with models that make approximating assumptions is that there are many possible combinations of approximating assumptions – models of bounded rationality being a case in point. This problem has been particularly acute in the bathtub literature, since there has been no reference point against which to compare the accuracy of competing models. But now that at least equilibrium in the proper basic bathtub model can be solved numerically, more information will be available to decide which of the bathtub models that make approximating assumptions is the most promising to develop further.

Funding

This material is based upon work supported by the National Science Foundation under grant no. CMMI-1629195.
Acknowledgments

The authors would like to thank the National Science Foundation for its financial support of this research. Buli would also like to thank the National Center for Sustainable Transportation (NCST) for providing fellowship support during the 2016–2017 academic year to support this research.

References


