CESifo Working Papers

www.cesifo.org/wp

A Model of Rush-Hour Traffic in an Isotropic Downtown Area

Richard Arnott
Anatolii Kokoza
Mehdi Naji

CESIFO WORKING PAPER NO. 5465
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
AUGUST 2015

ISSN 2364-1428

CESifo *Working Paper No. 5465*

A Model of Rush-Hour Traffic in an Isotropic Downtown Area

Abstract

For a quarter century, a top priority in transportation economic theory has been to develop models of rush-hour traffic dynamics that incorporate traffic jams (hypercongestion). The difficulty has been that "proper" models result in mathematical intractabilty, while none of the proposed approximating models has gained general acceptance. This paper takes a different tack, focusing on a particular proper model in which commuters decide when to travel so as to minimize a trip cost function that is linear in travel time and schedule delay (the so-called α-β-γ variant of the bottleneck model). Solutions of all the model variants entail departure/arrival masses.

JEL-Code: L910, R410.

Keywords: equilibrium, rush hour, traffic congestion.

Richard Arnott
Department of Economics
University of California, Riverside
4106 Sproul Hall
USA – Riverside, CA 92521-0427
richard.arnott@ucr.edu

Anatolii Kokoza
Department of Economics
University of Arizona
1130 E Helen St., 401 McClelland Hall
USA – Tucson, AZ 85721-0108
kokoza@email.arizona.edu

Mehdi Naji
Department of Economics
University of California, Riverside
4106 Sproul Hall
USA – Riverside, CA 92521-0427
mehdi.naji@email.ucr.edu

July 31, 2015

1 A Model of Rush-hour Traffic Dynamics in an Isotropic Downtown Area

William Vickrey's bottleneck model has been the workhorse of rush-hour traffic dynamic economic analysis for a quarter of a century. It is the kind of model every applied economic theorist dreams of formulating. Its simplicity leads to a degree of analytical tractability that admits numerous extensions (e.g. Arnott et al. (1993)); it is consistent with both the physics of traffic flow and individual rationality; it is rich in economic insight; and it has considerable predictive content. However, no model can capture all the important facets of an economic phenomenon, especially such a simple model, and an elegant model can seduce its users into overlooking its blemishes. Vickrey would have been the first to admit that there is much more to the economics of rush-hour traffic dynamics than is captured by his bottleneck model and its numerous extensions; indeed, even in the article that presented the bottleneck model (Vickrey, 1969), he outlined several alternative ways of modeling traffic congestion, each of which implies different rush-hour traffic dynamics than the bottleneck model.

Perhaps the most important unrealistic feature of the bottleneck model is its assumption that, in heavily congested conditions, downtown traffic flow is at its maximum sustainable level (which traffic engineers refer to as capacity flow). To someone stuck in a traffic jam, this assumption may seem absurd. It does jar with intuition, but it is not absurd and intuition can mislead. It is quite possible that we remember being stuck in traffic jams because it is so unpleasant, neglecting that for most of the journey traffic flows pretty smoothly. Detailed sensor data on freeway congestion have found that traffic flow is below capacity flow under heavily congested conditions, but not by much (Chung et al., 2007); the situation might be similar for city streets. Furthermore, there are downtown congestion technologies consistent with the bottleneck model's assumption.[1] In addition, the traditional static network equilibrium models that simulate the effects of changes to the road network assume that link travel time is increasing in link flow, implying that flow increases as congestion increases. For these reasons, until a few years ago, the realism of the bottleneck model's assumption that traffic flow is at its maximum under heavily congested conditions remained disputed. The issue could only be properly resolved through measurement, and there were no measurements of traffic flow at the level of a downtown neighborhood or of an entire downtown area.

That changed in 2008 with the publication of a landmark paper by Geroliminis and Daganzo (2008).

[1] Consider a symmetric grid network of downtown streets, in which each intersection is controlled by traffic signals and has a maximum capacity or throughput, which is achieved when, for one-way traffic there are queues in both entry directions to the intersection, and for two-way traffic there are queues in all four entry directions to the intersection – under heavily congested conditions. The network then operates at capacity where are there are queues at every intersection. If therefore the intersections do in fact operate at capacity when there are queues at every intersection, flow through the network is maximized under these heavily congested conditions. This seems paradoxical since a driver spends most of his travel time in queues at intersections, but the logic is sound.

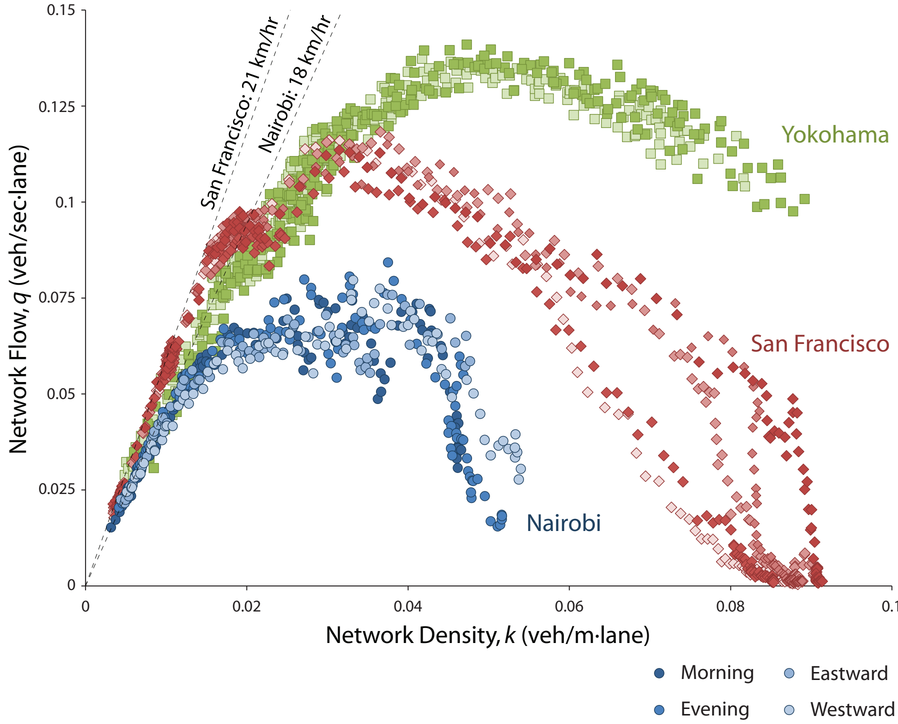


Figure 1: Macroscopic fundamental diagram for three cities

Note: Each dot corresponds to an observation. Observations were made at regular intervals throughout the business day and over days of the workweek

Using a combination of stationary and mobile (taxi) sensors, they measured traffic flow and density over a neighborhood of Yokohama, Japan essentially continuously over a period of weeks. They found an inverse U-shaped relationship between flow and density over the neighborhood that was stable over the course of the day, and also across days, which they termed the macroscopic fundamental diagram (MFD), as shown in Figure 1. They also found a stable relationship between outflow and density. Similar studies have since been done for other cities; traffic microsimulation models have been developed for these and other cities that obtain the same general results; and traffic flow models have been developed that are consistent with the empirical regularities. This subsequent work has uncovered the related empirical result that a neighborhood's MFD tends to be more stable, the more homogeneous the street system. Unlike a decade ago, the consensus today, among urban transportation economists, is that traffic flow downtown falls off sharply in heavily congested downtown traffic, which is characterized by frequent and often severe traffic jams.

Following Vickrey, transportation economists term traffic jam situations – situations in which traffic flow is negatively related to traffic density – hypercongestion or hypercongested travel. The debates have focused on whether hypercongestion is an equilibrium phenomenon and on how quantitatively important hypercongested travel is.

Several economists who have worked on downtown traffic congestion, including Vickrey (1991), have believed that hypercongestion is an equilibrium phenomenon in rush-hour traffic dynamics, and is quantitatively important. They have attempted to construct models with this feature that satisfy four criteria: i) consistency with the physics of traffic congestion; ii) individual rationality; iii) analytical tractability; and iv) robustness (in the sense that, at an aggregate level at least, the qualitative solution properties are preserved under realistic extensions). Numerous models have been put forward but none satisfies all four desiderata, and none has been widely accepted and adopted. The stumbling block has been mathematical.

The simplest context in which to illustrate the mathematical problem is the α-β form of the bottleneck model with identical individuals, no congestion tolling, and late arrival not permitted. The central equilibrium condition in the bottleneck model is that no commuter can lower her trip cost by altering her departure time from home, which is here referred to as the trip-timing (equilibrium) condition. Under the indicated assumptions, this condition reduces to the condition that trip cost, which comprises travel time cost and time early cost, be constant over the departure set (the set of times at which departures occur) and at least as high outside the departure set. Let t denote the departure time, t^* the common desired arrival time, $c(t)$ the trip cost of a commuter who departs home at time t, α the value of travel time and β the value of time late, and $T(t)$ the travel time of a commuter who departs at time t. Then the equilibrium trip cost condition is that, over the departure set,

$$\alpha(\text{travel time}) + \beta(\text{time early}) = \alpha T(t) + \beta(t^* - t - T(t)) = c(t) = \underline{c}, \tag{1}$$

and outside the departure time interval,

$$\alpha T(t) + \beta(t^* - t - T(t)) = c(t) \geq \underline{c}.$$

The simplest way to adapt the bottleneck model to account for reduced flow under heavily congested conditions is to assume that the capacity of the bottleneck,[2] s, falls off with the length of the queue, Q, waiting to get through the bottleneck: $s(t) = s(Q(t))$, $s' < 0$. Now consider the problem of determining the equilibrium departure rate from home, $r(t)$. First, we have that

$$Q(t) = \int_t^{t+T(t)} s(Q(u))du; \tag{2}$$

the time it takes for a commuter departing at time t to pass through the bottleneck is such that the integral of the bottleneck capacity over the time until she gets to the front of the queue, which equals her travel

[2]Capacity of the bottleneck is the rate at which cars pass through the bottleneck.

time, equals the length of the queue at the time she joined it. Second, we have that the rate of change of queue length equals the departure rate from home (and hence the arrival rate at the bottleneck) minus the discharge rate from the bottleneck, bottleneck capacity:

$$\dot{Q}(t) = r(t) - s(Q(t)). \tag{3}$$

We refer to models in this vein – that have sound physical and economic microfoundations – as "proper".

Now let us apply these equations to solve for the departure rate function consistent with the common trip price of $\underline{c}$. Differentiating (1) and (2) yields

$$\dot{T}(t) = \frac{\beta}{\alpha - \beta} \tag{4}$$

$$\dot{Q}(t) = s\Big(Q\big(t + T(t)\big)\Big)\big(1 + \dot{T}(t)\big) - s\big(Q(t)\big) \tag{5}$$

Substituting $T(t)$ from (1) and (4) into (5) gives

$$\dot{Q}(t) = s\left[Q\left(t + \frac{\underline{c} - \beta(t^* - t)}{(\alpha - \beta)}\right)\right]\frac{\alpha}{\alpha - \beta} - s\big(Q(t)\big). \tag{6}$$

This is a first-order, nonlinear delay differential equation in $Q(t)$ with an endogenous delay. If this were solvable, one could combine it with the boundary condition $Q(t^*) = 0$ to solve $Q(t)$, from which all other variables of interest could be determined. Unfortunately, even first-order linear delay differential equations with an exogenous delay are at the research frontier in applied mathematics, and (6) is even more difficult since it is nonlinear and its delay is endogenous. Even more fundamentally, we have not found existence and uniqueness theorems for this class of differential equations.

Standard differential equations describe how a system evolves as a function of a forcing function (such as the entry rate into a traffic stream) and a state variable that captures the history of the system, and is therefore backward looking. But a commuter's travel time on the journey to work depends not only on the state of traffic when she starts her journey, but also on how traffic evolves as she proceeds along her route, which depends on the entry of traffic onto the network after she has begun her journey.

Failure to crack this tough nut has delayed the development of a proper theory of downtown rush-hour traffic dynamics that admits hypercongestion by at least a quarter century. This situation is frustrating since it means that we lack a strong theoretical basis for designing policies to address rush-hour traffic congestion in the most congested cities.

Several authors have made simplifying assumptions that result in tractable models that can be regarded as approximations to a "proper" model. Agnew (1976) presented the first model of congestion that incorporates a state variable akin to density, but ignores the time spent in the system, hence converting a delay differential equation into an ordinary differential equation. Vickrey (1991) provides some preliminary notes, but not a complete model. Small and Chu (2003) circumvented the intractability by assuming that a commuter's travel time is a function of traffic density when she starts or ends her journey. While this simplification has been adopted in some other papers, including Geroliminis and Levinson (2009), it has not gained wide acceptance because it violates the laws of physics. Fosgerau and Small (2013) succeeded in circumventing the intractability but only by assuming a particular and unrealistic congestion technology. Without a solution to a proper model, it is rather fruitless to argue which of these models provides the best approximation to the corresponding proper model. Even if the proper models do not have analytical solutions, it should nonetheless be possible to obtain numerical solutions, against which the numerical solutions to the approximating models can be compared. Remarkably, no one has yet undertaken such an exercise.

Fosgerau (2015) presented a proper model, derived some analytical properties of the model under an assumed "regularity condition", and solved it numerically for some sets of parameter values. His paper made an important contribution to the literature, but the generality of its results is open to question. In particular, under the regularity condition, the equilibrium entails all cars being on the road at the peak of the rush hour. This is possible in cities in which road capacity is large relative to the demand. But in the world's heavily congested cities road capacity is insufficient to accommodate all cars on the road at the same time. This paper treats both cases.

In this paper we continue the search for the Holy Grail. We employ the same model of traffic congestion as that employed in Arnott (2013), which assumes an isotropic network of downtown streets with congestion described by a Greenshields' MFD,[3] which is consistent with the laws of physics. But, unlike Arnott (2013), the model's microeconomic foundations are solid. However, the model has two weaknesses. The first is specificity. Some of the qualitative properties of the equilibria and optima it describes are crucially dependent on the assumptions of an α-β-γ cost function and on a common desired arrival time. The basic bottleneck model makes these assumptions too, but most of its properties are robust to their relaxation, but the solution techniques for the model presented in this paper are not. Some mathematicians would argue that, as a result, this paper's model is, in fact, not a model but an extended example. The second weakness of this paper's model is that, in contrast to Vickrey's bottleneck model, the uniqueness of equilibrium remains unresolved.

A differential equation is often solved by intuiting the solution and then confirming it, rather than

[3]Greenshields' Relation (Greenshields, 1935) assumes a negative, linear relationship between velocity and density. The Greenfields' MFD is therefore a parabola. Greenshields' Function is chosen only because the algebra it generates is simple.

determining the solution analytically. That was the approach taken in this paper. The paper evolved from the simple observation that, in the variant of the isotropic model with identical commuters in which no late arrivals are admitted, and when the number of commuting trips is low relative to street capacity, a mass of departures in which all commuters depart together and arrive together satisfies trip-timing equilibrium condition. Trivially, all commuters have the same trip cost. As well, if a commuter departs after the mass, she arrives late, which is inadmissible, while if she departs before the mass, her travel time savings are more than offset by an increase in time early costs. As it turns out, this line of argument can be extended. As the number of commuting trips increases, the departure mass becomes sufficiently congested that the travel time savings from departing before the mass rises to level of the increase in time early costs, which gives a deviating commuter the incentive to depart before the departure mass. As the number of commuting trips continues to increases, a second departure mass forms whose travel interval does not intersect with that of the first departure mass but is contiguous to it. As the number of commuting trips increases even further, a third departure mass forms whose travel interval does not intersect with that of the second departure mass but is contiguous to it, and so on.

One nice feature of this no-toll equilibrium with departure masses is that its analysis is tractable. Another nice feature is that its externalities are transparent, in contrast to other, non-steady-state flow congestion models. One may object to its equilibrium on the grounds that departure masses are not observed, or that the traffic flow theory employed in the paper is inappropriately applied to departure masses. But since the equilibrium is logically sound in the context of the abstract world of the model, one then needs to enquire into how the model needs to be modified to obtain a connected departure set. As well, even though the model may generate an unrealistic departure and arrival pattern, its aggregate properties – how, for example, aggregate trip costs respond to various policies – may nonetheless be realistic.

An obvious question is whether the equilibrium with departure masses that we analyze is the unique equilibrium in the isotropic model with identical commuters when no late arrivals are permitted. There may be applicable theorems in the theory of delay differential equations with endogenous delays that give an unambiguous answer, but, if there are, we are not aware of them. We report on the limited progress we have made in developing an *ad hoc* argument to address the uniqueness issue. The difficulty is to develop an exhaustive typology of departure patterns. We conjecture that the equilibrium with departure masses and non-overlapping trip intervals is unique, since we have not discovered any other type of departure pattern that satisfies the equilibrium conditions.

Section 2 presents the basic model with identical commuters. Section 3 provides a detailed analysis and discussion of equilibrium in the model with identical individuals and fixed demand. Section 4 undertakes a variety of extensions: price-sensitive demand, the optimum with identical individuals, late arrivals, and commuter heterogeneity with respect to both time values and visit lengths. The treatment of extensions is suggestive and illustrative rather than comprehensive. Section 5 concludes.

2 The Basic Model

Consider an isotropic downtown area – home locations, job locations, and road capacity are uniformly distributed over space. Per unit area, $\hat{N}$ identical commuters must travel from home to work over the morning rush hour. All commuters have the same commuting distance, L, and have the same desired arrival time, t^*.

The form of traffic congestion is flow[4] congestion. In particular, to simplify the algebra Greenshields' Relation is assumed, in which traffic velocity is negative linearly related to traffic density:

$$\hat{v} = v_f \left(1 - \frac{\hat{k}}{\Omega}\right) \tag{7}$$

where $\hat{v}$ is velocity, v_f is free-flow velocity, $\hat{k}$ is density, and Ω is jam density, is assumed. Density is measured in terms of cars per unit area. Jam density is the maximum number of cars per unit area, and measures road capacity per unit area. Thus, for each car travel time per mile is $\frac{1}{\hat{v}} = \frac{\Omega}{(\Omega - \hat{k})v_f}$. Greenshields' Relation has the properties that capacity flow occurs when traffic density equals one-half jam density. Traffic is congested at densities below this level and hypercongested at densities above this level.

The familiar α-β-γ trip cost function is employed; in particular,

$$c = \alpha(\text{travel time}) + \beta(\text{time early}) + \gamma(\text{time late}).$$

Letting t denote departure time from home and $T(t)$ denote travel time from home to work as a function of departure time, (1) can be written as

$$cc(t) = \alpha T(t) + \beta \max\left(0, t^* - t - T(t)\right) + \gamma \max\left(0, t + T(t) - t^*\right). \tag{8}$$

3 (No-toll) Equilibrium with Identical Individuals

A traffic equilibrium is a distribution of departure rates from home such that no commuter can reduce her trip price by altering her departure time. A no-toll traffic equilibrium is a traffic equilibrium in which no toll is applied.

[4]The Fundamental Identity of Traffic Flow is that flow is equal to density times velocity.

In this section, we assume that individuals are identical and that no late arrivals are admitted. These assumptions are relaxed in section 4. Because it is central to this paper we highlight the following assumption:

Assumption 1. *The departure pattern takes the form of non-overlapping and contiguous time intervals, in each of which a departure mass travels from home to work, and the latest of which arrives exactly on time.*

We shall argue that a pattern of departures satisfying Assumption 1 *plus* the condition that trip cost be the same for all departure masses *plus* another condition on the number of departure masses is *a* trip-timing equilibrium. While we do not prove uniqueness of equilibrium, in section 3.1.3 we present some results that exclude some other departure patterns not satisfying Assumption 1 as possible trip-timing equilibria.

3.1 Equilibrium with One or Two Departure Masses

Consider a city with a small population density relative to its road capacity, in fact sufficiently small that in equilibrium all commuters depart at the same time in a single departure mass and arrive at work exactly on time. No commuter has an incentive to depart earlier since the decrease in travel time cost from doing so is more than offset by the increase in schedule delay cost. As population density increases, there is a critical value at which a commuter has an incentive to depart earlier than the mass. At this population density, equilibrium switches from having one departure mass to having two departure masses, and at a higher critical population density equilibrium switches from having two departure masses to three, etc.

To avoid burdensome notation, throughout most of the paper we shall employ a set of normalizations. To start, however, in order to motivate the normalizations, we work with unnormalized notation, which is indicated by hats on variables.

Let m denote the number of departure masses, and i index a departure mass. Departure masses are indexed in reverse order of departure time; thus, the last mass to depart, which arrives on time, has the index $i = 1$. This may seem counterintuitive, but the indexation is chosen so that the index of the departure mass that arrives on time does not change as the number of departure masses changes. Let $\hat{c}_i^m(\hat{N})$ be the trip cost of each commuter in mass i when there are m departure masses and the population density is $\hat{N}$, $\hat{n}_i^m(\hat{N})$ be the number of commuters in the ith departure mass with population density $\hat{N}$, $\hat{c}^e(\hat{N})$ be the equilibrium travel cost with population density $\hat{N}$, and $\hat{N}^e_{m,m+1}$ be the critical population density at which equilibrium switches from having m to $m + 1$ departure masses.

3.1.1 One departure mass (m=1)

Since there is only the one departure mass, $\hat{n}_1^1 = \hat{N}$. Also, since this departure mass arrives on time, commuters experience no schedule delay cost. Travel time is trip distance, L, divided by velocity,

$$\hat{v} = v_f \left(1 - \frac{\hat{N}}{\Omega}\right),$$

and trip cost equals travel time times the value of travel time, α. Thus,

$$c_1^1(N) = \frac{\alpha L}{\hat{v}} = \frac{\alpha L}{v_f \left(1 - \frac{\hat{N}}{\Omega}\right)} \tag{9}$$

and the departure time is

$$t^* - \frac{L}{v_f(1 - \frac{\hat{N}}{\Omega})}.$$

To avoid notational clutter, for the rest of the paper we shall employ several normalizations, but shall record results both with and without the normalizations. There are four units of measurement employed in the paper, those with respect to distance, time, money, and population per unit area. The normalizations are $L = 1$, $v_f = 1$, $\alpha = 1$, and $\Omega = 1$. Thus, the normalized distance is trip length, the normalized time unit is the length of time it takes to travel the trip length at free-flow velocity, the normalized money unit is the cost of travel per normalized time unit, and normalized population density is jam density. For most of the paper, unnormalized variables are indicated with a ˆ and normalized variables without the ˆ. With these normalizations, (9) reduces to

$$c_1^1(N) = \frac{1}{1 - N}, \tag{10}$$

and the travel time of the mass is $\frac{1}{1-N}$. With this normalization, travel in a departure mass is congested if the size of the departure mass is less than 0.5 and hypercongested if the size of the departure mass is greater than 0.5. To convert from normalized units to unnormalized units, employ the following rules

- 1 normalized distance unit equals L unnormalized distance units.
- 1 normalized time unit equals $\frac{L}{v_f}$ unnormalized time units.
- 1 normalized money unit equals $\frac{\alpha L}{v_f}$ unnormalized monetary units.
- 1 normalized population density unit equals Ω unnormalized population density units.[5]

[5] It will be useful to provide some intuition for the magnitude of N. Let q denote flow, $q = kv$. Applying Greenshields' Relation, the relationship between flow and density is $q = kv(k) = k(1 - k)$. Maximum or capacity flow is $1/4$. Thus, with $N = 1$, the duration of the rush hour at capacity flow would be four normalized time units. In the extended example that we

To further simplify notation: t^* is set equal to zero, so that time is measured relative to the desired arrival time; $\theta \equiv \frac{\beta}{\alpha}$ equals the ratio of the value of time early to the value of travel time and is assumed to be less than one; and $\rho \equiv \frac{\gamma}{\alpha}$ equals the ratio of the value of time late to the value of travel time and is assumed to be greater than one.

We now proceed with the analysis in normalized units. Consider an infinitesimal commuter who is considering deviating – departing earlier than the departure mass. Suppose that she considers departing a period Δt earlier than the departure mass. Since normalized free-flow velocity equals 1, she travels a distance Δt before encountering the departure mass. She then travels the remaining distance $1 - \Delta t$ with the departure mass at the speed $1 - N$, arriving at work at

$$-\frac{1}{1-N} - \Delta t + \Delta t + \frac{1-\Delta t}{1-N} = \frac{-\Delta t}{1-N}.$$

Thus, her travel time is

$$\frac{-\Delta t}{1-N} + \Delta t + \frac{1}{1-N} = \frac{1}{1-N} - \frac{N\Delta t}{1-N}.$$

Her trip cost is therefore

$$c_1^1(N) - \frac{N\Delta t}{1-N} + \frac{\theta\Delta t}{1-N}.$$

Her trip cost is therefore lower when she departs earlier than the departure mass if $N > \theta$, and higher otherwise. Thus, the critical population density at which equilibrium switches from having one to two departure masses is $N_{1,2}^e = \theta$. Note that this condition applies for any size of Δt less than or equal to 1. Consistent with Assumption A-1, we assume that the deviating commuter travels by herself in a separate departure mass that arrives at the departure time of the departure mass that arrives on time, and that as population density increases, successive departure masses form, each departing such that the mass arrives at work when the next (lower index) departure mass departs for work. We shall provide a justification for this assumption in section 3.1.3.

Let $TC_{(m)}(N)$ denote total trip cost with population density N conditional on there being m departure masses, and $TC^e(N)$ denote trip cost with the equilibrium number of departure masses for population density N. From (10), when there is a single departure mass in equilibrium, thus when $c^e(N) = c_1^1(N)$, total cost is

$$TC^e(N) = TC_{(1)}(N) = Nc^e(N) = \frac{N}{1-N}. \tag{11}$$

shall employ, we shall assume that $v_f = 15$ mph and $L = 5$ mls, so that the duration of a trip at free-flow speed, which is the normalized time unit, is 20 minutes. With these parameters and $N = 1$, the duration of the rush hour at capacity flow would be 80 minutes.

The corresponding marginal social cost and marginal congestion externality cost are therefore

$$\begin{aligned} MSC^e(N) &= \frac{dTC^e_{(1)}}{dN} = \frac{1}{(1-N)^2} \\ MCE^e(N) = MSC^e(N) - c^e(N) &= \frac{N}{(1-N)^2}. \end{aligned} \tag{12}$$

Total trip cost may be decomposed into total travel time cost, TTC, and total schedule delay cost, SDC. With only one departure mass, since all commuters arrive exactly on time and therefore experience no schedule delay, all of the total trip cost is total travel time cost. In this case, the marginal congestion externality cost has a simple interpretation. It is simply the cost imposed on other users from increasing traffic density in the single departure mass by one unit. Define the severity of congestion, s, to be the ratio of the marginal congestion externality cost to the private trip cost. Then in equilibrium with one departure mass

$$s^e(N) = \frac{MCE^e(N)}{c^e(N)} = \frac{N}{1-N}. \tag{13}$$

Earlier it was asserted that, when equilibrium entails a single departure mass, the departure mass arrives on time. Suppose not. Then the departure mass arrives early. But then a deviating commuter may reduce both her travel time cost and her schedule delay cost by departing after the departure mass and arriving exactly on time. The same line of argument applies to all subsequent cases.

We bring together the above results in

Proposition 1. *An equilibrium with a single departure mass occurs when $N \leq \theta$. Over this interval of N, $c^e(N) = \frac{1}{1-N}$, $MSC^e(N) = \frac{1}{(1-N)^2}$, $MCE^e(N) = \frac{N}{(1-N)^2}$, and $s^e(N) = \frac{N}{1-N}$.*

3.1.2 Two departure masses (m=2)

There are now two departure masses. To satisfy the trip-timing equilibrium condition, trip cost must be the same for each departure mass. Letting n_i^m denote the normalized number of commuters in departure mass i when there are m departure masses, equilibrium with two departure masses solves the following pair of equations:

$$n_1^2 + n_2^2 = N \tag{14}$$

$$\frac{1}{1-n_1^2} = c_1^2 = c_2^2 = \frac{1}{1-n_2^2} + \frac{\theta}{1-n_1^2} \tag{15}$$

Departure mass 1 arrives on time, so that $c_1^2 = \frac{1}{1-n_1^2}$. Departure mass 2 arrives immediately before departure mass 1 departs, so that a commuter in departure mass 2 experiences schedule delay of $\frac{1}{1-n_1^2}$ and travel time of $\frac{1}{1-n_2^2}$. Solving (14) and (15) gives

$$\begin{aligned} {}^e n_1^2 &= \frac{N+\theta-N\theta}{2-\theta} \\ {}^e n_2^2 &= \frac{N-\theta}{2-\theta}. \end{aligned} \tag{16}$$

Two additional conditions are required for (16) to describe an equilibrium with two departure masses. The first is that each departure mass have a strictly positive density, which requires that $N > \theta$. The second is that a deviating commuter does not have an incentive to form a third departure mass. It is shown below that this condition is that $N \leq \theta(3-\theta)$. Thus, equilibrium entails two departure masses for $N \in (\theta, \theta(3-\theta))$. For N in this interval

$$c^e(N) = c_1^2 = \frac{2-\theta}{(2-N)(1-\theta)} \tag{17}$$

$$TC^e(N) = \frac{(2-\theta)N}{(2-N)(1-\theta)} \tag{18}$$

$$MSC^e(N) = \frac{2(2-\theta)}{(2-N)^2(1-\theta)} \tag{19}$$

$$MCE^e(N) = \frac{(2-\theta)N}{(2-N)^2(1-\theta)} \tag{20}$$

$$SDC^e(N) = \frac{{}^e n_2^2\theta}{1-{}^e n_1^2} = \frac{\theta(N-\theta)}{(2-N)(1-\theta)} \tag{21}$$

$$TTC^e(N) = TC^e(N) - SDC^e(N) = \frac{2N(1-\theta)+\theta^2}{(2-N)(1-\theta)} \tag{22}$$

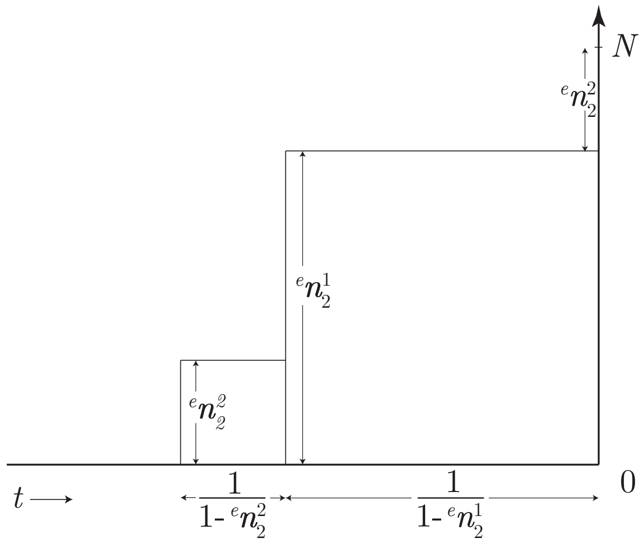
$N_{2,3}^e$ is that N for which a commuter is indifferent between departing in the second departure mass and departing in a third departure mass by herself. If she departs in a third departure mass by herself, her travel time cost decreases by $\frac{1}{1-n_2^2} - 1$ and her schedule delay cost increases by $\frac{\theta}{1-n_2^2}$. The decrease in travel time cost equals the increase in schedule cost when $\frac{1-\theta}{1-n_2^2} = 1 \Rightarrow n_2^2 = \theta \Rightarrow N_{2,3}^e = \theta(3-\theta)$.

Comparing (17) and (20) gives the severity of congestion

$$s^e(N) = \frac{N}{2-N}. \tag{23}$$

We bring together the results for $m = 2$ in

Proposition 2. *An equilibrium with two departure masses occurs when $N \in (\theta, \theta(3-\theta))$. Over this interval of population density:* ${}^e n_1^2 = \frac{N+\theta-N\theta}{2-\theta}$, ${}^e n_2^2 = \frac{N-\theta}{2-\theta}$, $c^e(N) = \frac{2-\theta}{(2-N)(1-\theta)}$, $MSC^e(N) = \frac{2(2-\theta)}{(2-N)^2(1-\theta)}$, $MCE^e(N) = \frac{(2-\theta)N}{(2-N)^2(1-\theta)}$, *and* $s^e(N) = \frac{N}{2-N}$.



(a)

$$N = {}^e n_1^2 + {}^e n_2^2$$
$${}^e c_1^2 = \frac{1}{1 - {}^e n_1^2}$$
$${}^e c_2^2 = \frac{1}{1 - {}^e n_2^2} + \frac{\theta}{1 - {}^e n_1^2}$$
$${}^e c_1^2 = {}^e c_2^2 \Rightarrow \frac{1-\theta}{1 - {}^e n_1^2} = \frac{1}{1 - {}^e n_2^2}$$

(b)

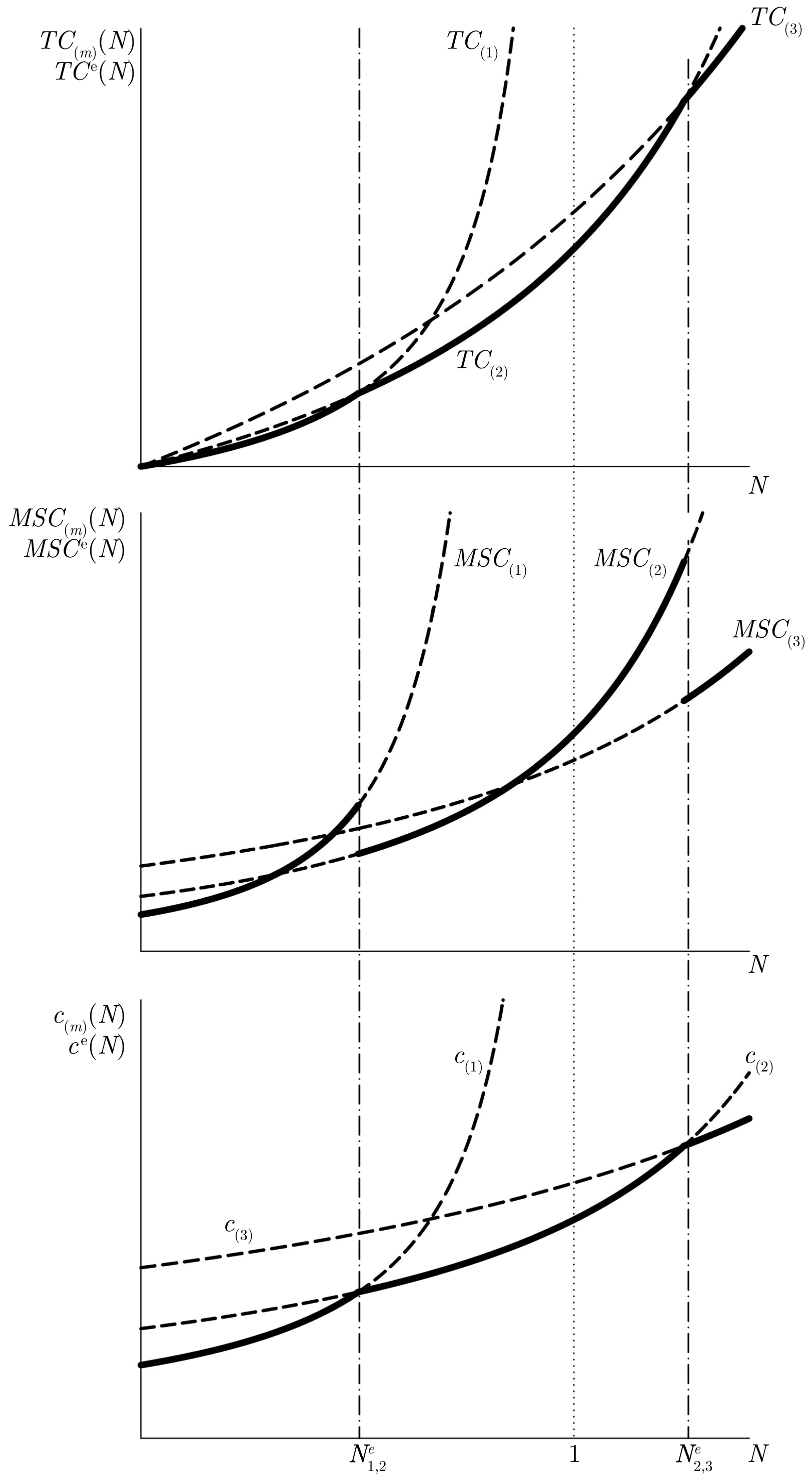
Figure 2: No-toll equilibrium with two departure masses

Figure 2 displays the equilibrium with two departure masses graphically. The abscissa is the normalized time axis and the ordinate is normalized population density. Departure masses are numbered so that departure mass 1 arrives on time, and departure mass 2 arrives immediately before departure mass 1 departs. Since in equilibrium commuters in departure mass 1 have the same trip cost as commuters in departure mass 2, and since commuters in departure mass 1 arrive on time, experiencing no schedule delay cost, while those in departure mass 2 arrive early, experiencing schedule delay cost, travel time cost must be higher for commuters in departure mass 1 than those in departure mass 2. Thus, the size of the departure mass, and hence traffic density, must be higher in departure mass 1 than in departure mass 2. Travel speed is therefore lower for commuters in departure mass 1, resulting in a longer trip duration. The sum of the normalized population densities over the two departure masses gives the exogenous normalized population density, N. The duration of the rush hour equals the sum of the trip durations of the two departure masses.

To illustrate the results thus far, consider a numerical example in which $\theta = 1/2$, so that $N^e_{1,2} = 1/2$ and $N^e_{2,3} = 5/4$. The values of N considered are 0, 1/2, 1, and 5/4. To convert costs from normalized to unnormalized units, the following parameter values are assumed: $\alpha = \$20/\text{hr}$, $L = 5$ mls, and $v = 15$ mph. $L/v_f = 0.333$ hrs is the assumed trip time at free-flow speed. The numerical results are recorded in Table 1. Complementing Table 1 is Figure 3, which plots $TC^e(N)$, $c^e(N)$, and $MSC^e(N)$, for $N \in (0, 6/4)$.

Turn first to the three panels of Figure 3. The top one plots total cost against N, the middle one marginal social cost against N, and the bottom one trip cost against N. The three panels are aligned vertically.

Start with the bottom panel. $c_{(1)}(N) = \dfrac{1}{1-N}$ gives trip cost as a function of N when the entire



$TC_{(m)}(N)$
$TC^{e}(N)$
$TC_{(1)}$
$TC_{(2)}$
$TC_{(3)}$
N
$MSC_{(m)}(N)$
$MSC^{e}(N)$
$MSC_{(1)}$
$MSC_{(2)}$
$MSC_{(3)}$
N
$c_{(m)}(N)$
$c^{e}(N)$
$c_{(1)}$
$c_{(2)}$
$c_{(3)}$
$N^{e}_{1,2}$
1
$N^{e}_{2,3}$
N

Figure 3

population travels in a single departure mass. Since the mass arrives on time, the entire trip cost is travel time cost. Trip cost is a convex function having the properties that $c_{(1)}(0) = 1$, $c_{(1)}(1/2) = 2$, and $c_{(1)}(1) = \infty$. Normalized trip cost is 1.0 when population is zero since trip cost equals travel time cost at free-flow speed, which has been normalized to 1, is equal to 2.0 when normalized population is 0.5 since density equals one-half jam density, and is equal to ∞ when normalized population is $\Omega = 1.0$ since density equals jam density. The curve is drawn as a solid, bold line for $N \in [0, 1/2]$, the interval over which the equilibrium number of departure masses is 1, and as a dashed line outside this interval.

$c_{(2)}(N) = \dfrac{2-\theta}{(2-N)(1-\theta)}$ is a convex function. $N = \theta$ is the lowest population density at which the equal trip cost condition for each departure mass is consistent with both departure masses having positive population density, while N = 2 corresponds to jam density. $c_{(2)}(N)$ has the properties that for $N = \theta$ $c_{(2)}(\theta) = c_{(1)}(\theta) = \dfrac{1}{1-\theta}$, $c_{(2)}(1) = \dfrac{2-\theta}{1-\theta}$, and $c_{(2)}(2) = \infty$. The curve is drawn as a solid line for $N \in (1/2, 5/4]$, the interval over which the equilibrium number of departure masses is 2, and as a dashed line outside this interval. $c_{(3)}(N)$ is displayed as well, though its functional form has not yet been derived. Since a switch occurs from one to two, from two to three, etc. departure masses when a commuter faces the same trip cost whether she departs in the existing departure masses, or deviates and departs in her own departure mass, the equilibrium trip cost function, $c^e(N)$, which is drawn as a solid, bold line, is the lower envelope of the trip cost functions for specific numbers of departure masses, and hence has an escalloped shape.

The middle panel displays the marginal social cost functions with one, two, and three departure masses. If the bottom and the middle panels were combined, it would be seen that each marginal social cost function lies above the corresponding trip cost functions, with the vertical distance between two functions measuring the congestion externality cost. The equilibrium marginal social cost function is not the lower envelope of the departure-mass specific marginal social cost functions. Instead, $MSC^e(N)$, which is drawn as the solid bold line, jumps downward at each critical population density at which there is a switch from one to two, from two to three, etc. departure masses. The reason is that at any of these critical population densities, a commuter imposes a lower congestion externality cost if she departs in her own departure "mass", than if she departs in the existing departure masses. The top panel displays total cost as a function of population density with one, two, and three departure masses, as well as the equilibrium total cost function, which is the lower envelope of the total cost functions for specific numbers of departure masses.

Table 1 displays the quantitative properties of equilibrium for $N = 0$, $1/4$, $1/2$, 1, $5/4$, as well as higher N which are discussed later in section 3.5. The new data presented in the Table are the severity of congestion, the ratio of total schedule delay cost to total travel time cost, trip cost in dollars, and the length of the rush hour in hours. Observe that: i) the ratio of total schedule delay cost to total travel time cost appears

Table 1: Numerical example for certain values of N and m

N normalized population density	m number of departure masses	c^e normalized trip cost	MSC^e normalized marginal social cost	s^e severity of congestion	SDC^e/TTC^e total schedule delay cost /total travel time cost	$\hat{D}$ rush hour length in hrs	$\hat{c}^e$ trip cost in $
0	1	1	1	0	0	1/3	20/3
1/4	1	4/3	16/9	7/3	0	4/9	80/9
1/2	1	2	4	3	0	2/3	40/3
1/2	2	2	8/3	5/3	0	1	40/3
1	2	3	6	5/2	1/5	3/2	20
5/4	2	4	32/3	29/9	1/4	2	80/3
5/4	3	4	48/7	205/105	1/4	7/3	80/3
2	3	7	21	2	11/17	49/12	140/3
17/8	3	8	192/7	17/7	7/10	14/3	160/3
17/8	4	8	256/15	17/15	7/10	5	160/3
3	4	15	60	3	49/41	225/24	100
49/16	4	16	1024/15	49/15	27/22	10	320/3
49/16	5	16	1280/31	49/31	27/22	31/3	320/3
4	5	31	165	4	159/89	961/48	620/3
129/32	5	32	5120/31	129/31	83/46	62/3	640/3
129/32	6	32	2048/21	43/21	83/46	21	640/3

Notes: 1. The money normalization is that a trip at free-flow travel speed that arrives at the common work start time costs 1 unit. Since a trip has a length of $L = 5$ miles, since free-flow speed is 15 mph, since a trip that arrives at the common work start time entails no schedule delay cost, and since the value of travel time is $20/hr, the (unnormalized) dollar cost of a trip at free-flow speed that arrives at the common work start time is $6.66.
2. D is the length of the rush hour in normalized time, measured from the time of the first departure to the time of the last arrival, and $\hat{D}$ is the unnormalized length. The time normalization is that a trip at free-flow travel speed takes 1 time unit. Since a trip has a length of 5 miles and since the free-flow speed is 15 mph, the unnormalized time unit is 20 minutes.

to increase monotonically with population density; ii) the severity of congestion increases with population density over each population density interval for which the number of departure masses is constant, and decreases discontinuously at each critical population density where a departure mass is added; and iii) the length of the rush hour increases with population density over population density intervals where the number of departure masses remains constant, and increases discontinuously at each critical population density where a departure mass is added.

Hypercongestion occurs when the normalized density of cars exceeds 1/2. For the population density interval over which there is one departure mass in equilibrium, hypercongestion occurs when $\theta > 1/2 > N$, and does not occur when $\theta \leq 1/2$; with the assumed parameter value of $\theta = 1/2$, hypercongestion never occurs. For the population density interval over which there are two departure masses in equilibrium, hypercongestion occurs in departure mass 1 when $n_1^2 = \dfrac{N + \theta - N\theta}{2 - \theta} > 1/2$ and in departure mass 2 when $n_2^2 = \dfrac{N - \theta}{2 - \theta} > 1/2$; with the assumed parameter value of $\theta = 1/2$, hypercongestion occurs in the first departure mass for $N \in (1/2, 5/4)$, but for no values of N in the second departure mass.

The stage is now set to work out equilibrium with three or more departure masses in equilibrium. But before we do this, we present the limited results that we have obtained related to qualitative properties of the departure distribution consistent with trip-timing equilibrium.

3.1.3 Qualitative properties of the departure time distribution consistent with trip-timing equilibrium

In this section, we shall continue to employ normalized notation. $T(t)$ is the trip duration of a departure at time t.

Proposition 3. *When late arrival is not permitted, in the interior of any interval of the rush hour over which departures occur continuously,* $v(k(t + T(t))) = (1 - \theta)v(k(t))$.

Proof. The following two conditions must hold in the interior any interval of the rush hour over which departures occur continuously.

$$1 = \int_t^{t+T(t)} v(k(u))du \tag{i}$$

$$T(t) + \theta(t^* - (t + T(t))) = c^e \tag{ii}$$

The first condition is that the integral of velocity over the duration of a trip equals trip length, which is normalized to 1. The second is that, for consistency with equilibrium, over any interval of the rush hour over which departures occur continuously, trip cost must equal equilibrium trip cost.

Since these conditions must hold over the interior of any interval of the rush hour over which departures occur continuously, the time derivatives of these conditions apply. Time differentiation of both conditions, and substitution of one into the other yields the stated result. ■

Proposition 3 states that, for any commuter who departs in the interior of any interval of the rush over which departures occur continuously, travel speed at the end of her journey must be 1 - θ times that at the beginning of her journey.

Corollary. *If a commuter who arrives on time departs on the "late" boundary of an interval of the rush hour over which departures occur continuously, there must be an arrival mass at work start time and hence a departure mass on the late boundary.*

Proof. From Proposition 3, the limit of velocity as $t \uparrow 0$ must be strictly less than free-flow velocity, which implies that the limit of traffic density as $t \uparrow 0$ is strictly positive. But since no late arrivals are permitted, traffic density at work start time must be zero. These two properties can be consistent only if there is a mass of arrivals at work start time, causing traffic density to fall discontinuously from a strictly positive level to zero. ■

By a similar line of argument, if a commuter who arrives on time does not depart on the interval of the rush hour over which departures occur continuously, then he must arrive with an arrival mass and hence depart with the same mass.

Combining results, a commuter who arrives on time must do so as one of a mass.

Proposition 4. *Two departure masses with overlapping travel intervals is inconsistent with equilibrium.*

Proof. With two departure masses with overlapping travel intervals, the rush hour is described by three intervals. Only commuters who depart in departure mass 2 travel in the first interval, all commuters travel in the middle interval, and only commuters who depart in departure mass 1 travel in the last interval. Let δ be the normalized distance traveled by all commuters in the middle interval. The condition that commuters in the two departure masses incur the same cost is

$$c = \frac{1-\delta}{1-n_2^2} + \frac{\delta}{1-N} + \frac{\theta(1-\delta)}{1-n_1^2} = \frac{\delta}{1-N} + \frac{1-\delta}{1-n_1^2}, \tag{iii}$$

which reduces to

$$\frac{1}{1-n_2^2} = \frac{1-\theta}{1-n_1^2},$$

which is the same as (15) for the situation where time intervals for the two departure masses are disjoint. Thus, from (15),

$$n_1^2 = \frac{N+\theta-N\theta}{2-\theta}.$$

Inserting this into (iii) yields

$$c = \frac{\delta}{1-N} + \frac{(1-\delta)(2-\theta)}{(2-N)(1-\theta)}.$$

Now consider a deviating commuter who travels a distance δ solo before joining the commuters in departure mass 2, with whom she travels the remaining distance 1 - δ. By doing so, she avoids traveling over the congested time interval during which commuters in departure masses 1 and 2 travel together. She travels a distance δ solo at a speed of 1, and a distance 1 - δ at a speed of

$$1 - n_2^2 = \frac{2-N}{2-\theta},$$

and experiences schedule delay of

$$\frac{1-\delta}{1-n_1^2} = \frac{(1-\delta)(2-\theta)}{(2-N)(1-\theta)}$$

for a total cost of

$$c' = \delta + \frac{(1-\delta)(2-\theta)}{2-N} + \frac{\theta(1-\delta)(2-\theta)}{(2-N)(1-\theta)}$$

$$\begin{aligned}
c - c' &= \frac{\delta}{1-N} + \frac{(1-\delta)(2-\theta)}{(2-N)(1-\theta)} - \delta - \frac{(1-\delta)(2-\theta)}{2-N} - \frac{\theta(1-\delta)(2-\theta)}{(2-N)(1-\theta)} \\
&= \frac{\delta N}{1-N} + \frac{(1-\theta)(1-\delta)(2-\theta)}{(2-N)(1-\theta)} - \frac{(1-\delta)(2-\theta)}{2-N} \\
&= \frac{\delta N}{1-N} > 0
\end{aligned}$$

Thus, the deviating commuter experiences lower trip cost than those traveling in the two departure masses with overlapping travel intervals. ■

While Propositions 3 and 4 are helpful, together they fall a long way short of establishing that, with no late arrivals and identical individuals, the equilibrium with the departure pattern taking the form of non-overlapping and contiguous time intervals, in each of which a departure mass travels from home to work, and the latest of which arrives exactly on time, is unique.

3.2 General Solution of Equilibrium

Fortunately, a recursive structure in the equilibrium size of adjacent departure masses permits neat, closed-form solution for the equilibrium in cities of all sizes and with any number of departure masses. The analysis below first solves for total trip cost, marginal social cost, and marginal congestion externality cost as functions of m and N, such that trip cost is the same in each departure mass (even though this can entail negative departure masses) and then determines the equilibrium m as a function of N.

With m departure masses,

$$c_j^m = \frac{1}{1-n_j^m} + \theta \sum_{i=1}^{j-1} \frac{1}{1-n_i^m}. \tag{24}$$

The trip-timing equilibrium condition implies that

$$1 - n_{j+1}^m = \frac{1-n_j^m}{1-\theta}. \tag{25}$$

Combining (25) with the condition that $\sum_{j=1}^{m} n_j^m = N$ yields a finite series expression for n_1^m. Rewriting the finite series expression as the difference between two infinite series, and then applying standard results on the sum of infinite series and solving for n_1^m gives

$$n_1^m = 1 - (m - N)\frac{1-\theta}{\theta}\frac{1-(1-\theta)^m}{(1-\theta)^m}. \tag{26}$$

Combining (26) and (24) for $j = 1$, and noting that in equilibrium the trip cost is the same for all departure masses, yields the equilibrium trip cost

$$c^e(N) = \frac{1}{m-N}\frac{1-\theta}{\theta}\frac{1-(1-\theta)^m}{(1-\theta)^m}. \tag{27}$$

The total trip cost can then be calculated as $TC^e_{(m)}(N) = Nc^e_{(m)}(N)$:

$$TC^e_{(m)} = \frac{N}{m-N}\frac{1-\theta}{\theta}\left[\frac{1-(1-\theta)^m}{(1-\theta)^m}\right]. \tag{28}$$

Differentiation of $TC^e_{(m)}(N)$ with respect to N yields marginal social cost:

$$MSC^e_{(m)} = \frac{m}{N(m-N)}TC^e_{(m)}. \tag{29}$$

Marginal congestion externality cost can be calculated either as $MCE^e_{(m)}(N) = MSC^e_{(m)}(N) - c^e_{(m)}(N)$ or as $MCE^e_{(m)}(N) = N\left(\frac{\partial c^e_{(m)}(N)}{\partial N}\right)$:

$$MCE^e_{(m)} = \frac{1}{m-N}TC^e_{(m)}. \tag{30}$$

The equilibrium number of departure masses is now calculated as a function of N. The switch from m to $m+1$ departure masses occurs for that N for which the trip cost with $m+1$ departure masses equals the trip cost with m departure masses: $TC^e_{(m+1)}(N^e_{m,m+1}) = TC^e_{(m)}(N^e_{m,m+1})$. Using (28), this reduces to

$$N^e_{m,m+1} = m - \frac{1-\theta}{\theta}\Big(1-(1-\theta)^m\Big). \tag{31}$$

This can be rewritten as a recursive relationship:

$$N^e_{m+1,m+2} = \theta(m+1) + (1-\theta)N^e_{m,m+1}. \tag{32}$$

Using (25) and (26), the duration of the rush hour with m departure masses and population density N is

$$D^e_{(m)} = (N;\theta) = \sum_{i=1}^{m}\frac{1}{1-\ {}^e n_i^m} = \left[\frac{1-(1-\theta)^m}{\theta}\right]^2\frac{(1-\theta)^{1-m}}{m-N} \tag{33}$$

We also have that

$$
\begin{aligned}
TTC^e_{(m)}(N;\theta) &= \sum_{i=1}^{m} \frac{{}^e n_i^m}{1-{}^e n_i^m} = \sum_{i=1}^{m}\left(\frac{1}{1-{}^e n_i^m}-1\right) \\
&= \left[\frac{1-(1-\theta)^m}{\theta}\right]^2 \frac{(1-\theta)^{1-m}}{m-N} - m
\end{aligned} \tag{34}
$$

and

$$
\begin{aligned}
SDC^e_{(m)}(N;\theta) &= TC^e_{(m)}(N;\theta) - TTC^e_{(m)}(N;\theta) \\
&= \frac{1-(1-\theta)^m}{\theta^2}\frac{(1-\theta)^{1-m}}{m-N}\left[N\theta - 1 + (1-\theta)^{1-m}\right] + m
\end{aligned} \tag{35}
$$

Table 2 brings together results in normalized form. Table 3 gives the corresponding results in unnormalized form.

Table 2: Algebraic results in normalized form: equilibrium with no late arrivals

population of ith departure mass per unit area	$n_i^m(N;\theta) = 1 - \frac{m-N}{(1-\theta)^{i-1}A(\theta)}$
trip cost per commuter	$c_{(m)}(N;\theta) = \frac{1}{m-N}A(\theta)$
total trip cost per unit area	$TC_{(m)}(N;\theta) = \frac{N}{m-N}A(\theta)$
marginal social cost per unit area	$MSC_{(m)}(N;\theta) = \frac{m}{(m-N)^2}A(\theta)$
marginal congestion externality cost per unit area	$MCE_{(m)}(N;\theta) = \frac{N}{(m-N)^2}A(\theta)$
ratio of marginal social cost to trip cost	$\frac{MSC_{(m)}(N;\theta)}{c_{(m)}(N;\theta)} = \frac{m}{m-N}$
severity of congestion	$s^e_{(m)}(N;\theta) \equiv \frac{MCE_{(m)}(N;\theta)}{c^e(N;\theta)} = \frac{N}{m-N}$
total travel time cost per unit area	$TTC_{(m)}(N;\theta) = \frac{1-(1-\theta)^m}{\theta}\frac{A(\theta)}{m-N} - m$
total schedule delay cost per unit area	$SDC_{(m)}(N;\theta) = \frac{A(\theta)}{m-N}\left[N - \frac{1+(1-\theta)^m}{\theta}\right] + m$
mass switching population densities	$N^e_{m,m+1} = m - (1-\theta)^m A(\theta)$
duration of rush hour	$D_{(m)}(N;\theta) = \frac{1-(1-\theta)^m}{\theta}\frac{A(\theta)}{m-N}$

Note: Let $A(\theta) = \frac{1-\theta}{\theta}\left[\frac{1-(1-\theta)^m}{(1-\theta)^m}\right]$.

3.3 Comparative Static and Dynamics Properties of Equilibrium

The comparative static properties of the no-toll equilibrium are given in Table 4. Comparing Tables 2 and 3, it can be seen that some of the comparative static effects operate through the normalizations, and might therefore be called scale effects, while the others operate via θ and N. The discreteness of departure masses

Table 3: Algebraic results in unnormalized form: equilibrium with no late arrivals

population of ith departure mass per unit area	${}^{e}\hat{n}_i^m(\hat{N};\theta) = \Omega\Big[1 - \dfrac{m-\frac{\hat{N}}{\Omega}}{(1-\theta)^{i-1}A(\theta)}\Big]$
trip cost per person	$\hat{c}^e_{(m)}(\hat{N};\theta) = \dfrac{\alpha L}{v_f\left(m-\frac{\hat{N}}{\Omega}\right)}A(\theta)$
total trip cost per unit area	$\hat{TC}^e_{(m)}(\hat{N};\theta) = \dfrac{\alpha L\hat{N}}{v_f\left(m-\frac{\hat{N}}{\Omega}\right)}A(\theta)$
marginal social cost	$\hat{MSC}^e_{(m)}(\hat{N};\theta) = \dfrac{\alpha L m}{v_f(m-\frac{\hat{N}}{\Omega})^2}A(\theta)$
marginal congestion externality cost	$\hat{MCE}^e_{(m)}(\hat{N};\theta) = \dfrac{\alpha L\frac{\hat{N}}{\Omega}}{v_f(m-\frac{\hat{N}}{\Omega})^2}A(\theta)$
ratio of marginal social cost to trip cost	$\dfrac{\hat{MSC}^e_{(m)}(\hat{N};\theta)}{\hat{c}^e_{(m)}(\hat{N};\theta)} = \dfrac{m}{m-\frac{\hat{N}}{\Omega}}$
severity of congestion	$\hat{s}^e_{(m)}(\hat{N};\theta) \equiv \dfrac{\hat{MCE}^e_{(m)}(\hat{N};\theta)}{\hat{c}^e_{(m)}(\hat{N};\theta)} = \dfrac{\hat{N}}{m\Omega-\hat{N}}$
total travel time cost per unit area	$\hat{TTC}^e_{(m)}(\hat{N};\theta) = \frac{\alpha L\Omega}{v_f}\Big(\frac{1}{m-\frac{\hat{N}}{\Omega}}\frac{1-(1-\theta)^m}{\theta}A(\theta) - m\Big)$
total schedule delay cost per unit area	$\hat{SDC}^e_{(m)}(\hat{N};\theta) = \frac{\alpha L\Omega}{v_f}\Big[\frac{A(\theta)}{m-\frac{\hat{N}}{\Omega}}\big[\frac{\hat{N}}{\Omega} - \frac{1-(1-\theta)^m}{\theta}\big] + m\Big]$
mass switching population densities	$\hat{N}^e_{m,m+1} = \Omega\Big[m-(1-\theta)^m A(\theta)\Big]$
duration of rush hour	$\hat{D}^e_{(m)}(\hat{N};\theta) = \dfrac{L}{v_f}\dfrac{1-(1-\theta)^m}{\theta}\dfrac{A(\theta)}{m-\frac{\hat{N}}{\Omega}}$

Notes: Let $A(\theta) = \dfrac{1-\theta}{\theta}\left[\dfrac{1-(1-\theta)^m}{(1-\theta)^m}\right]$. The normalized monetary unit is the cost of the time it takes to travel trip distance at free-flow velocity; thus, to convert to unnormalized units, multiply by $\dfrac{\alpha L}{v_f}$. The normalized time unit is the time it takes to travel the trip length at free-flow velocity; thus, to convert to unnormalized units, multiply by $\dfrac{L}{v_f}$, trip length divided by free-flow velocity. The normalized population density is relative to the jam density; thus, to convert to unnormalized units, multiply by jam density, Ω.

raises difficulties for comparative static analysis, since an infinitesimal change in an exogenous variable can cause a change in the equilibrium number of departure masses, and when this occurs some endogenous variables change discontinuously. We have chosen to present the comparative static results, holding constant number of departure masses, and then in the last row of Table 4 to indicate whether an increase in the exogenous parameter might cause either an increase or a decrease in the equilibrium number of departure masses.

The signs of the comparative static derivatives are all the same with three or more departure masses as they are with two departure masses. To more easily convey the intuition, we shall therefore focus on the case when there are two departure masses in equilibrium, which was presented in detail in Section 3.1.2. The discussion that follows provides intuition for a sample of the comparative static results.

Table 4: Some Comparative Static Properties of Equilibrium (Unnormalized)

	$\hat{N}$	Ω	prop $\uparrow$ in $\hat{N}$ and Ω	v_f	L	prop $\uparrow$ in L and v_f	θ with α fixed	α with β fixed	prop $\uparrow$ in α and β
${}^{e}\hat{n}_i^m$	+	+	+	0[2]	0[2]	0	+[3]	-	0
$\hat{c}^e_{(m)}$	+	-	0	-	+	0	+	?[9]	+
$M\hat{S}C^e_{(m)}$	+	-	0	-	+	0	+	?[9]	+
$M\hat{C}E^e_{(m)}$	+	-	0	-	+	0	+	?[9]	+
$\frac{M\hat{S}C^e_{(m)}}{\hat{c}^e_{(m)}}$	+	-	0	0	0	0	0	0	0
$s^e_{(m)}$	+	-	0	0	0	0	0	0	0
$T\hat{T}C^e_{(m)}$	+	-	+	-	+	0	+[4]	?[10]	+
$S\hat{D}C^e_{(m)}$	+	-	+	-	+	0	?[5]	?[11]	+
$\frac{T\hat{T}C^e_{(m)}}{S\hat{D}C^e_{(m)}}$	-[1]	+	0	0	0	0	+[6]	-[6]	0
$\hat{N}^e_{m,m+1}$	0	+	+	0	0	0	+[7]	-[7]	0
$\hat{D}^e_{(m)}$	+	-	0	-	+	0	+[8]	-[8]	0
m^e	≥ 0	≤ 0	0	0	0	0	≤ 0	≥ 0	0

Notes: 1. In deriving the results, we drew heavily on Hardy et al. (1952), HLP hereafter. This particular result uses the inequality $-\left[\frac{1-(1-\theta)^m}{\theta}\right]^2(1-\theta)^{1-m}+m^2<0$ for m strictly positive integer and for $\theta\in(0,1)$. 2. An increase in v_f or a decrease in L has no effect on the density of the departure masses, but travel times associated with each of the departure masses fall in the same proportion, which causes TTC and SDC to fall in the same proportion. 3. Via an elementary inequality, $A'(\theta)>0$. 4. From (2.15.6) of HLP, $\frac{d[\frac{1-(1-\theta)^m}{\theta}]}{d\theta}$ is positive. 5. The ambiguity in sign is confirmed by (21). Suppose that $\theta=1/2$. Recall that with this value of θ, the number of departure masses is two in equilibrium when $N\in(0.5,1.25)$. It follows straightforwardly from (21) that SDC is decreasing in N for $N\in(0.5,0.75)$ and is increasing in N for $N\in(0.75,1.25)$. 6. Where E denotes an elasticity: $E_{(\frac{TTC}{SDC}):\theta}=E_{TTC:\theta}-E_{SDC:\theta}$. $E_{TTC:\theta}>E_{(TTC+m):\theta}<E_{A:\theta}$. $E_{SDC:\theta}<E_{(SDC-m):\theta}<E_{A:\theta}$. Thus, $E_{(\frac{TTC}{SDC}):\theta}>0$. 7. This result can be obtained by recursion on (32). 8. The effect operates through θ. For a given number of departure masses and a given population density, the duration of the rush hour is minimized when the departure masses have the same size. An increase in θ causes the departure masses to become more unequal in size, which increases the duration of the rush hour. 9. The ambiguity in sign is confirmed by (17). Writing (17) in unnormalized form and then substituting β/α for θ gives $\hat{c}^e=\dfrac{\alpha(2\alpha-\beta)}{(2-N)(\alpha-\beta)}$. The partial derivative of $\hat{c}^e$ with respect to α has the sign of $2\alpha^2-4\alpha\beta+\beta^2$, and is negative for smaller values of α relative to β, and positive for larger values. The derivatives for MSC and MCE have the same sign as for c. 10. From (22), the partial derivative of $T\hat{T}C^e$ has the sign of $2N-[\frac{\beta}{\alpha-\beta}]^2$. With $\theta=0.5$, and $N=1$, the equilibrium number of departure masses is two and the derivative is positive. With $\theta=0.6$, and $N=1$, the equilibrium number of departure masses is two and the derivative is negative. 11. Take the case of two departure masses and $\theta=1/2$. With this value of θ, equilibrium entails two departure masses when $N\in(1/2,5/4)$. From (21), $dS\hat{D}C/d\alpha$ has the sign of $(1-N)\beta$, which is positive for $N\in(1/2,1)$ and negative for $N\in(1,5/4)$.

The only comparative static derivative with respect to population density that is worthy of remark is that the ratio of $\frac{T\hat{T}C}{S\hat{D}C}$ unambiguously decreases with $\hat{N}$. The intuitive reason is that schedule delay is experienced only by those in the second departure mass, and as $\hat{N}$ increases the proportion of the population in the second departure mass increases (see (16)). The only comparative static derivative with respect to jam density that is worthy of remark is that the population in the first departure mass increases with jam density. The reason is that, as jam density increases, the level of traffic congestion falls, so that trip costs are equalized for those traveling in the first and second departure mass when the first departure mass receives a larger proportion of the population. When there is a proportional increase in $\hat{N}$ and Ω, the equilibrium is unchanged except for a scaling up; all per capita magnitudes remain unchanged.

The comparative static properties with respect to free-flow velocity derive from Greenshield's Relation. Speed increases proportionally for all densities. The size of each departure mass remains unchanged, but travel time and schedule delay shrink in the same proportion as free-flow travel time. A proportional increase in free-flow velocity and trip length has no effect on the listed endogenous variables. In each departure mass, commuters travel double the distance at double the speed, resulting in no change in trip cost.

The comparative static properties with respect to α and β are a quantum level more complex. Consider the effects of an increase in θ, holding α constant, which is the same as an increase in β holding α constant. This change causes commuters to attach more weight to reducing schedule delay, which tilts the distribution of commuters over departure masses towards masses that arrive less early. This increases the severity of congestion and hypercongestion in the masses that arrive less early, leading to some counterintuitive and anomalous comparative static results. One striking result is that an increase in β can lead to a decrease in schedule delay costs, which implies that schedule delay falls more than in proportion to the rise in β. Recall that, in the extended example presented earlier in this section, a second departure mass starts to form when $\mathrm{N} = \theta$. Now consider an initial situation when $N_0 = \theta_0 + \Delta$, where Δ is a small, positive number. From (16), the number of commuters in the second departure mass is

$$\frac{N - \theta_0}{2 - \theta_0} = \frac{\Delta}{2 - \theta_0},$$

which is the number that experience schedule delay. An increase in β, holding α constant, causes an increase in θ from θ_0 to θ_1. This causes the number of commuters in the second departure mass to shrink from

$$\frac{\Delta}{2 - \theta_0}$$

to

$$\frac{\Delta - (\theta_1 - \theta_0)}{2 - \theta_1}.$$

As θ increases, a point is reached where the number of commuters in the second departure mass, and hence schedule delay costs, shrinks to zero, decreasing at an infinite rate, far exceeding the proportional increase in β. Another striking result is that an increase in β causes as *unambiguous increase* in the duration of the rush hour. It is paradoxical that a stronger desire to arrive closer to the desired arrival time results in a lengthening of the rush hour. Reconciling intuition with the result combines three observations. The first is that, holding the number of departure masses fixed, as is done in the table, an increase in β causes population to be redistributed from earlier to later departure masses, in this sense concentrating the arrival distribution, which is consistent with intuition. The second observation is that, because travel time is convex in congestion, again holding fixed the number of departure masses, an increasing concentration of commuters in later departure masses increases the total length of the rush hour. The third observation is that the result becomes ambiguous once the number of departure masses is allowed to vary.

Another result worthy of note is that an increase in α may cause equilibrium trip cost to *fall.* This is paradoxical since in other contexts an increase in the price of a factor of production (here the factor of production is travel time) increases costs. Here inputs are not combined in a cost-minimizing way because of externalities within the process of production. More specifically, by deconcentrating departures across departure masses, the rise in α causes travel time in the first departure mass to fall. If travel is severely hypercongested in the first departure mass, the proportional decrease in travel time in the mass may exceed the proportional increase in α. The result then follows from noting that, since travelers in the first departure mass experience no schedule delay, their trip cost, which is the equilibrium trip cost, equals their travel time cost.

3.4 Other Properties of Equilibrium: Existence, Uniqueness, and Stability

As observed in the introduction, solving for equilibrium rush-hour traffic in models of an isotropic downtown area entails the solution of a delay differential equation with an endogenous delay. Such differential equations are at the research frontier in applied mathematics. We do not know whether there are existence or uniqueness proofs that can be applied to models of rush-hour traffic dynamics in an isotropic downtown area.

We have solved for an equilibrium for a particular model with identical individuals, flow congestion described by Greenshields' Relation, no late arrivals, and an α-β cost function. We intuited the solution for $m = 1$ and $m = 2$, and then generalized the solution to an arbitrary number of departure masses. In so doing, we have provided a constructive proof of the existence of an equilibrium.

To what extent can our model be generalized while still permitting the construction of an equilibrium? In the next section, we shall illustrate the construction of equilibrium with $m = 2$ with $\alpha-\beta-\gamma$ preferences, and with commuter heterogeneity in $\frac{\alpha}{\beta}$, $\frac{\alpha}{\gamma}$, $\frac{\beta}{\gamma}$, or trip length. On this basis, we conjecture that an equilibrium can be constructed with $\alpha-\beta-\gamma$ preferences, and commuter heterogeneity in all these dimensions simultaneously. We also conjecture that our method of construction of equilibrium can be generalized to other well-behaved congestion functions, though what constitutes "well-behaved" remains to be seen. Finally, we conjecture that our method of construction of equilibrium, which entails mass points, requires that there be a slope discontinuity in the trip cost function at the common desired arrival time.[6] Put alternatively, it is the slope discontinuity that causes equilibrium to be characterized by mass points. If this conjecture is correct, then the form of equilibrium that we have constructed applies only to special cases with the slope discontinuity. Nevertheless, since these special cases can be obtained as limits of more general cases, we conjecture that many of the aggregative properties of the equilibrium we have constructed hold for trip cost functions without the slope discontinuity.

In section 3.1.3, we investigated whether the equilibrium we constructed for our basic model is unique. We presented two Propositions, each of which ruled out classes of departure patterns. But together these Propositions fall a long way short of establishing uniqueness of equilibrium. Thus, even though we have not constructed other types of equilibria, uniqueness must be regarded as an open issue.

Stability is defined with respect to an adjustment mechanism. A reasonable adjustment mechanism is as follows: Out of equilibrium, between days a small proportion of commuters experiment with a new departure time different from their regular departure time. If a commuter's trip cost is lower with the new departure time, she makes it her regular departure time; otherwise, she returns to her previous regular departure time. Ben-Akiva et al. (1984) apply such an adjustment mechanism to examine the stability of equilibrium in the bottleneck model. Application of that mechanism to our model should produce similar results.

3.5 Numerical Example Extended to Higher N

N measures the population density of commuters relative to the capacity of the road network per unit area. In most US cities, N is modest. Even though hypercongestion may occur, the rush hour is relatively short. In the world's mega-cities, however, most of which are in developing countries, N can be much larger. In 1990 (before the subway system there was built), one of the authors was told that in Bangkok, the rush hour starts at 3:30 in the morning. Earlier in the paper, in section 3.1.2, we presented a numerical example, but discussed only those cases where equilibrium entailed two mass points. Now that we have generalized our analysis, we discuss the remaining cases, with larger N, which provides insight into the behavior of our

[6]We thank Mogens Fosgerau for posing this conjecture.

model under heavily congested conditions. In the numerical example, we assumed that $\theta = 1/2$. From (26), we have that $N^e_{1,2} = 1/2$, $N^e_{2,3} = 5/4$, $N^e_{3,4} = 17/8$, $N^e_{4,5} = 49/16$, and $N^e_{5,6} = 129/32$.

Particularly striking is how rapidly trip cost increases with population density. At integer values of population density, $c^e(N) = 2^{N+1} - 1$. Since trip cost in the first departure mass is entirely travel time cost, and since the cost of travel time is normalized at 1, travel time in the first departure mass too is related to N according to $2^{N+1} - 1$, while travel speed as function of N is related to N according to $\dfrac{1}{2^{N+1}-1}$. In particular, speed in the first departure mass, and therefore at the peak of the rush hour, is 15 mph with $N = 0$, 15/3 mph with $N = 1$, 15/7 mph with $N = 2$, 15/15 mph with $N = 3$, and so on. The relationship between peak speed and N, which relates demand to capacity, is specific to Greenshields' Relation, but employing an empirically estimated relationship between velocity and density would give a qualitatively similar result.

With $\theta = 1/2$, as assumed in the example, for integer N, $m = N + 1$. The severity of congestion,

$$s^e(N) = \frac{MCE^e(N)}{c^e(N)} = \frac{N}{m - N},$$

then reduces to $s(N) = N$. But a conceptually superior measure of the severity of congestion is the ratio of the congestion externality cost imposed by a commuter divided by the congestion cost experienced by the commuter, which we term the private congestion cost, $PCC(N)$, and is defined as trip cost minus trip cost with no congestion, which has been normalized to one: $PCC(N) = c(N)$ - 1. Defining this alternative measure of the severity of congestion as

$$S(N) \equiv \frac{MCE(N)}{PCC(N)},$$
$$S(N) = \frac{MCE(N)}{c(N) - 1} = \frac{N}{m - N}\frac{c(N)}{c(N) - 1},$$

which with integer N and $\theta = 1/2$ reduces to

$$S(N) = \frac{N(c(N) - 1)}{c(N)}.$$

Thus, $S(1) = 3/2$, $S(2) = 7/3$, $S(3) = 45/14$. As demand relative to capacity rises, the congestion cost experienced by a commuter is a decreasingly small fraction of the congestion cost she imposes on others. In this sense, as N rises not only does the absolute distortion due to unpriced congestion increase but so too does the relative distortion.

A further point to note is how rapidly the length of the rush hour increases with N. For integer N and $\theta = \frac{1}{2}$, $\hat{D}(N) = \dfrac{(2^{N+1} - 1)(2 - (\frac{1}{2})^N)}{3}$.

4 Extensions

The basic model of the previous section can be extended in many of the ways that the bottleneck model has been extended. While the extensions are conceptually straightforward, they are algebraically more cumbersome than extensions in the bottleneck model. For this reason we have chosen to illustrate these extensions by considering only optima and equilibria that entail two departure masses rather than to undertake comprehensive analyses. If the model is well received, we hope that other researchers will choose to further develop these extensions.

4.1 Price-sensitive Demand

As in the bottleneck model, the function relating trip cost to the number of trips can be regarded as a reduced-form supply curve. Adding to this a demand curve relating the number of trips to trip cost permits solution of equilibrium with price-sensitive demand. When hypercongestion occurs in equilibrium, the reduced-form supply curve is backward bending, which may result in multiple equilibria, which alternate between stable and unstable (see Arnott and Inci (2010)).

4.2 Social Optimum

For the rest of the paper, we shall return to the assumption that demand is completely inelastic and shall employ the same normalizations as applied in the previous section. Here we present a local social optimum entailing departure masses. It is a local optimum in the sense that the social costs of travel by the N commuters cannot be reduced, conditional on departures occurring in masses.[7] Whether it is the global social optimum we do not know.

When a single departure mass is optimal, the social optimum coincides with the corresponding equilibrium. As population density increases, a critical population density is reached at which it becomes optimal for there to be two departure masses. Determination of the social optimum with two departure masses is analogous to that for the no-toll equilibrium except that the marginal social cost of trips in each of the two

[7]Consider the social optimum problem, conditional on the solution entailing two departure masses. Let δ be the proportion of the trip distance that the two masses travel together. The social optimum problem is to minimize social costs with respect to n_1^2 and δ, and subject to the constraint that $n_1^2 \in [0, N]$. Social costs are $\dfrac{\delta N}{1-N} + \dfrac{(1-\delta)n_1^2}{1-n_1^2} + \dfrac{(1-\delta)(N-n_1^2)}{1-N+n_1^2} + \dfrac{(1-\delta)\theta(N-n_1^2)}{1-n_1^2}$. The first-order condition with respect to n_1^2 is independent of δ. Substituting the optimized value of n_1^2 into the expression of social costs and maximizing with respect to δ shows that a corner solution is optimal. $\delta = 1$ corresponds to a single departure mass, which is inconsistent with the assumption that the optimum entails two departure masses. $\delta = 0$ corresponds to non-overlapping and contiguous time intervals for the two departure masses. The argument appears to be general.

departure masses is equalized rather than the trip cost. The optimality conditions are

$$n_1^2 + n_2^2 = N$$
$$MSC_1^2 = MSC_2^2.$$

Total social costs are

$$TC_{(2)} = \frac{n_1^2}{1-n_1^2} + \frac{n_2^2}{1-n_2^2} + \frac{\theta n_2^2}{1-n_1^2}. \tag{36}$$

The first term is the travel time cost of departure mass 1 commuters; the second is the travel time cost of departure mass 2 commuters; and the third is the schedule delay cost of mass 2 commuters. Thus:

$$MSC_1^2 = \frac{1}{(1-n_1^2)^2} + \frac{\theta n_2^2}{(1-n_1^2)^2} \tag{37}$$
$$MSC_2^2 = \frac{1}{(1-n_2^2)^2} + \frac{\theta}{1-n_1^2} \tag{38}$$

The social cost of inserting an extra commuter in departure mass 1 equals the direct cost associated with the added commuter, $\frac{1}{1-n_1^2}$, plus the travel time externality cost imposed on other commuters in departure mass 1, $\frac{n_1^2}{(1-n_1^2)^2}$, plus the schedule delay externality cost imposed on commuters in departure mass 2, $\frac{\theta n_2^2}{(1-n_1^2)^2}$. The social cost of inserting an extra commuter in departure mass 2 equals the direct cost associated with the added commuter, $\frac{1}{1-n_2^2} + \frac{\theta}{1-n_1^2}$, plus the travel time externality cost imposed on other commuters in departure mass 2, $\frac{n_2^2}{(1-n_2^2)^2}$. Equating the marginal social costs for the two departure masses and substituting in the population condition yields

$$\frac{(1-\theta(1-N))}{(1-n_1^2)^2} = \frac{1}{(1-N+n_1^2)^2},$$

which reduces to

$${}^*n_1^2 = \frac{1-(1-N)A}{1+A}, \text{ where } A = \sqrt{1-\theta(1-N)}. \tag{39}$$

Thus

$$
{}^*c_1^2 = \frac{1}{1 -^* n_1^2} = \frac{1+A}{A(2-N)} \tag{40}
$$

$$
{}^*n_2^2 = N -^* n_1^2 = \frac{A-(1-N)}{1+A} \tag{41}
$$

$$
{}^*c_2^2 = \frac{1}{1 -^* n_2^2} + \frac{\theta}{1 -^* n_1^2} = \frac{(A+\theta)(1+A)}{A(2-N)} \tag{42}
$$

$$
\begin{aligned}
{}^*MSC_1^2 &= \frac{1+\theta n_2^2}{(1-n_1^2)^2} = \frac{(A+1+\theta)(1+A)}{A(2-N)^2} \\
{}^*MSC_2^2 &= \frac{1}{(1-n_2^2)^2} + \frac{\theta}{1-n_1^2} = \frac{(A+1+\theta)(1+A)}{A(2-N)^2}
\end{aligned} \tag{43}
$$

Note that when $A > 1$, which corresponds to $N > 1$, departure mass 2 is larger than departure mass 1, so that the peak of rush-hour congestion does not occur at the peak of the rush hour. The reason is that, as population density increases, the marginal schedule delay externality cost becomes relatively more important because the marginal commuter in departure mass 1 increases the schedule delay of more and more commuters.

By setting ${}^*n_1^2 = N$, we obtain the critical population density at which there is a switch from one to two departure masses in the social optimum:[8]

$$
N_{1,2}^* = \frac{(2+\theta)-(\theta^2+4)^{1/2}}{2}. \tag{44}
$$

Result 1. ${}^eN_{1,2} = \theta >\ {}^*N_{1,2}$ *for* $\theta \in (0,1)$

The critical population density at which there is a switch from one to two departure masses is lower in the social optimum than in the no-toll equilibrium, which is consistent with the downward shift in MSC at ${}^eN_{1,2}$ noted in the previous section. In the no-toll equilibrium, there are externalities associated with adding a commuter to either departure mass 1 or departure mass 2, but the externality cost of adding a commuter to departure mass 1 is higher than adding a commuter to departure mass 2 since first the travel time externality cost is higher and second a commuter added to departure mass 1 generates a schedule delay externality cost whereas a commuter added to departure mass 2 does not.

Result 2. $N_{1,2}^* < 1/2$ *for* $\theta \in (0,1)$

In the social optimum, as population density increases, departure mass 2 is created before hypercongestion arises in departure mass 1. In contrast, in the no-toll equilibrium, when N is greater than $1/2$ but less than θ, there is a single, hypercongested departure mass. This is an example of a general result that is worthwhile

[8] There is a negative root and a positive root. The positive root has ${}^*n_1^2 > 1$, which does not make economic sense.

recording as a

Proposition 5. *Hypercongestion does not occur in the social optimum.*

Proof: TO BE COMPLETED.

We have already identified the externalities associated with adding a commuter to departure mass 1 and then to departure mass 2. As is standard, the social optimum can be decentralized by imposing a congestion toll equal to the trip externality cost, evaluated at the social optimum. Thus,

$$ {}^{*}\tau_i^2 = {}^{*}MSC_i^2 - {}^{*}c_i^2, \; i = 1, 2, \tag{45} $$

which can be calculated from (40), (42), and (43):

$$ {}^{*}\tau_1^2 = \frac{(1+A)(A+N-1+\theta)}{A(2-N)^2} $$
$$ {}^{*}\tau_2^2 = \frac{(1+A)(1+(N-1)(A+\theta))}{A(2-N)^2} $$

We now construct a numerical example with two departure masses in the social optimum, and compare the social optimum and the no-toll equilibrium. The first step is to obtain ${}^{*}N_{2,3}$. To calculate this, we solve for the N for which ${}^{*}n_3^3 = 0$ when ${}^{*}MSC^3(N) = {}^{*}MSC^2(N)$. Now, when ${}^{*}n_3^3 = 0$,

$$ {}^{*}MSC_3^3(N_{2,3}^{*}) = 1 + \frac{\theta}{1 - {}^{*}n_2^2(N_{2,3}^{*})} + \frac{\theta}{1 - {}^{*}n_1^2(N_{2,3}^{*})}, \tag{46} $$

where the 1 is the marginal social travel time cost of a commuter in the third departure mass. Setting ${}^{*}MSC_2^2(N_{2,3}^{*})$, which is given by (43) when ${}^{*}n_1^3 = 0$, equal to ${}^{*}MSC_3^3(N_{2,3}^{*})$, which is given by (46):

$$ 1 + \frac{\theta(1+A)^2}{(2-N_{2,3}^{*})A} = \frac{(A+1+\theta)(1+A)}{A(2-N_{2,3}^{*})^2} \tag{47} $$

yields $N_{2,3}^{*} = 0.776$. We choose our parameter values to obtain an example with instructive properties: $N = 0.75$ and $\theta = 0.9$. In general, the number of departure masses in the social optimum is greater than or equal to the number in the no-toll equilibrium. With the parameter values chosen, there are two departure masses in the social optimum but only one in the no-toll equilibrium. Table 5 compares numerical properties of the social optimum and no-toll equilibrium with these parameter values.

To put the example into perspective, recall that a normalized time unit is 20 minutes, and that, if traffic flow were at capacity throughout the rush hour, the duration of the rush hour would be $4N = 3$ normalized time units or one hour. Thus, we are considering a small city, not a mega-city. The unit schedule delay

Table 5: Comparison of the no-toll equilibrium and social optimum with $N = 0.75$ and $\theta = 0.9$ (normalized units)

	m	n_1^2	n_2^2	c_1	c_2	MSC	TC	TTC	SDC	D
no-toll eq.	1	0.75	0	4	n.a.	16	3	3	0	4
social opt.	2	0.4326	0.3174	1.796	3.042	3.801	1.742	1.238	0.5035	3.227

Notes: Recall that a normalized time unit equals 20 minutes, that dollar trip cost equals \$6.66 times normalized trip cost, and that the number of commuters is measured relative to jam density, so that $N = 1$ corresponds to a rush hour lasting 80 minutes at capacity flow.

cost was chosen to be high so that rush hour in the no-toll equilibrium would be so concentrated that hypercongestion would develop, resulting in substantial efficiency gains from congestion tolling. Because commuters attach a high value to arriving at work close to on time, the no-toll equilibrium is highly congested. There is only a single departure mass, which travels at only 3.75 mph – severe hypercongestion – resulting in 80 minutes of travel time. Each commuter imposes 4 hours of delay on other commuters, resulting in a marginal social cost of a trip, in terms of travel time of 5 hours 20 mins. In the no-toll equilibrium, in contrast, commuters distribute themselves between two departures masses. Travel speed in the more congested departure mass 1 is 8.51 mph, compared to a free-flow speed of 15 mph, while that in departure mass 2 is at 10.23 mph. Commuters in departure mass 1 experience a travel time of 35.92 minutes, for a normalized cost of 1.796, while commuters in mass 2 have a 29.29 minute commute, which, along with a 26.94 minute schedule delay, results in a normalized cost of 3.042. The marginal social cost of a trip is 3.801 normalized time units, which is less than one-quarter that in the no-toll equilibrium. Even though the optimum has two departure masses, its rush hour is 65 minutes, significantly shorter than that in the no-toll equilibrium. This example illustrates well a paradox of hypercongestion – even though commuters in the no-toll equilibrium ignore the high cost that their traveling at the peak of the rush hour imposes on others, which intuitively should result in concentration of the rush hour, the length of the rush hour is in fact longer than in the social optimum. The resolution of the paradox is that ignoring the external cost they impose on others causes commuters to concentrate their departure times (in fact, in the model they all depart at the same time), but the concentration of departure times creates such severe hypercongestion that ignoring the external cost increases the length of the rush hour. Congestion tolling, by causing commuters to face the external cost, results in commuters deconcentrating their departure times, eliminating traffic jams and shortening the rush hour.

Table 5 also illustrates another important point. Under circumstances where the no-toll equilibrium is highly congested, the efficiency gains from imposing the optimal congestion toll exceed the toll revenue raised! In the example, the optimal toll is 2.005 time units (\$13.36) for commuters traveling in departure mass 1 and 0.759 time units ($5.06) for those traveling in departure mass 2. In normalized units, the toll revenue raised is 1.108 normalized units, while the efficiency gain from congestion tolling is 1.258 normalized units. Thus, the example illustrates the very considerable efficiency gains achievable through congestion pricing even in a small city, albeit one highly prone to congestion.

A word of caution is in order. The parameters were chosen to keep the calculations simple (only two departure masses at the optimum) while at the same time illustrating the very substantial efficiency gains achievable under congestion tolling when the no-toll equilibrium is highly congested. With a more realistic choice of θ, commuter efficiency gains from congestion tolling of the magnitude in the example would occur only for considerably "larger" cities – cities with considerably longer rush hours.

In all the subsequent extensions, we shall treat only the no-toll equilibrium and not the social optimum.

4.3 Late Arrivals

The analysis can quite straightforwardly be extended to treat late arrivals. We shall consider only no-toll equilibria in which there is a departure mass that arrives exactly on time. To keep the indexation consistent, we will denote by $i = -1$ the first late departure mass, $i = -2$ as the second late departure mass, and so on, and the departure masses arriving early or on time are indexed as before.[9]

We shall again restrict our analysis to two departure masses. With a low population density, there is only a single departure mass, which departs early and arrives on time. In due course, as population density grows, a critical population density is reached at which a deviating commuter will choose to depart either in a second early departure mass or the first late departure mass. Under what conditions will each occur? With departure in a second early departure mass, the deviating commuter's trip cost is ${}^{e}c_{2}^{2} = 1 + \dfrac{\theta}{1-N}$. With late departure, her trip cost is ${}^{e}c_{-1}^{2} = 1+\rho$ (recall that $\rho \equiv \dfrac{\gamma}{\alpha}$); she travels at free-flow speed, incurring one unit of travel time cost and one unit of time late cost. We have seen that if the second mass to form is a second early departure mass, it starts to form when N = θ. Thus, the second mass to form is a second early departure mass if $\dfrac{\theta}{1-\theta} < \rho$, and a first late departure mass if the inequality is reversed. Empirical work suggests that θ is around 0.5, while ρ is around 2.0, in which case the second departure mass to form would be a second early departure mass. On the assumption that this is the case, we can determine the third departure mass to form. With late departure, the trip cost of a deviating commuter remains ${}^{e}c_{-1}^{2} = 1 + \rho$. With departure in the third early departure mass and $\theta = 0.5$, the trip cost of a deviating commuter is 4 (see Table 1). Thus, with $\theta = 0.5$ and $\rho \in (1, 3)$, the third departure mass to form is the first late departure mass. The analysis to determine the departure masses that form as population density increases is straightforward,

[9] An alternative, more intuitive, indexation indexes the departure mass that arrives at $t = 0^-$ by $i = -1$, and the departure mass that departs at $t = 0^+$ by $i = 1$, etc. Thus, departure masses which arrive early and hitherto have been indexed by positive integers, would be indexed by negative integers, while positive integers denote late arrivals.

albeit tedious. We conjecture that the equilibrium that we have identified is unique, but have not proved this. In what follows, we shall revert to the assumption that late arrivals are not permitted.

4.4 Commuters Who Differ in θ

It is well understood that in the no-toll equilibrium with no late arrivals, commuters with a higher θ depart (weakly) later. The analysis is easier if we assume that there is a continuous cumulative distribution of commuters according to θ, $F(\theta)$. In this case, as population density increases, a second departure mass will form when the *marginal* commuter finds it desirable to deviate. Let ${}^{e}c_i^m(\theta')$ denote the equilibrium trip cost of a commuter with $\theta = \theta'$ who travels in departure mass i when there are m departure masses, and $\theta_{i,i+1}$ denote the θ of the marginal commuter who is indifferent between departing in mass i and mass $i+1$. We now derive the conditions for the existence of an equilibrium with two departure masses. The first condition is that everyone with $\theta > \theta_{1,2}$ travel in the first departure mass, and everyone with $\theta < \theta_{1,2}$ travel in the second. The second condition is that the commuter with $\theta = \theta_{1,2}$ be indifferent between traveling in departure masses 1 and 2. And the third is that the commuter with the lowest value of θ, $\underline{\theta}$, weakly prefer being in the second departure mass to deviating and forming her own, third departure mass.

$$\frac{1}{1 - N(1 - F(\theta_{1,2}))} = {}^{e}c_1^2(\theta_{1,2}) = {}^{e}c_2^2(\theta_{1,2}) = \frac{1}{1 - NF(\theta_{1,2})} + \frac{\theta_{1,2}}{1 - N(1 - F(\theta_{1,2}))} \quad (48)$$

$$1 + \frac{\underline{\theta}}{1 - NF(\theta_{1,2})} + \frac{\underline{\theta}}{1 - N(1 - F(\theta_{1,2}))} = {}^{e}c_3^3(\underline{\theta}) \geq {}^{e}c_2^2(\underline{\theta}) = \frac{1}{1 - NF(\theta_{1,2})} + \frac{\underline{\theta}}{1 - N(1 - F(\theta_{1,2}))} \quad (49)$$

To construct an example, we assume that the distribution of θ in the population is uniformly distributed on the interval (0.1, 0.9), so that $F(\theta) = 1.25(\theta - 0.1)$, and $1 - F(\theta) = 1.125 - 1.25\theta$. Substituting these formulae into (48) yields

$$\frac{1 - \theta_{1,2}}{1 - N(1.125 - 1.25\theta_{1,2})} = \frac{1}{1 - 1.25N(\theta_{1,2} - 0.1)}.$$

and into (49) yields

$$1 \geq \frac{0.9}{1 + 0.125N - 1.25N\theta_{1,2}}.$$

With $N = 0.3$, in equilibrium there are two departure masses and $\theta_{1,2} = 0.2198$ (with this information, all other variables of interest may be calculated). In order that commuters with the lowest unit schedule delay cost not form a third departure mass, departure mass 2 must exhibit little congestion, which requires that the bulk of commuters travel in departure mass 1.

4.5 Commuters Who Differ in L

Compared to the bottleneck model, this paper's model has the advantage that it can accommodate variable trip length. Not only does this add descriptive realism, but it also permits the analysis of policies that contain a trip distance element, including a gasoline tax and distance-dependent congestion tolling (Daganzo and Lehe, 2015).

In the no-toll equilibrium with no late arrivals, congestion increases over the early morning rush hour. Commuters order themselves by trip length. Commuters with longer trips choose to travel in the left tail of the rush hour, before congestion becomes heavy, in order to reduce travel time. Those with shorter trips prefer to suffer heavy congestion for the relatively short duration of their trips in order to arrive close to the work start time.

Thus far, we have used the term "departure masses". Since everyone had the same trip distance, we could have just as well used the term "arrival masses". In this subsection, however, the distinction matters. Since schedule delay depends on arrival time and not departure time, in this subsection the masses that constitute equilibrium are arrival masses. Arrival mass 1, which is assumed to arrive on time, is composed of the commuters with the shortest trip lengths. Among commuters in this arrival mass, the commuter with the longest trip is the first to depart, and when she departs she faces empty streets. As time proceeds on her commute, commuters with increasingly short trip distances enter and traffic becomes increasingly dense. The result is a saw tooth pattern of traffic density over the rush hour. This is unintuitive, but the trip timing condition is satisfied.

Thus far cumbersome equations have been avoided through use of a particular set of normalizations. Since, however, one of the normalizations was based on a common trip length, and since it would be confusing to introduce a new set of normalizations just for this one subsection, we shall proceed without normalization but without using $\hat{}$s to indicate normalized variables. Let $G(L)$ denote the cumulative distribution function of trip length, ${}^{e}L^{m}_{i,i+1}$ denote the equilibrium trip length of the marginal commuter who is indifferent between departing in mass i and mass $i+1$ when there are m departure masses, and ${}^{e}c^{m}_{i}({}^{e}L^{m}_{i,i+1})$ denote the equilibrium trip cost of this commuter. After considering conditions for equilibrium with two arrival masses, we shall proceed to an example with two arrival masses. The first condition is that the commuter with $L = {}^{e}L^{2}_{1,2}$ be indifferent between traveling in arrival masses 1 and 2. The second is that the commuter with the longest trip, $\bar{L}$, weakly prefer being in the second arrival mass to deviating and forming her own, third arrival mass. All other commuters are inframarginal in that they strictly prefer their chosen arrival mass to the alternatives.

As Fosgerau (2015) observed, the isotropic model with heterogeneous trip lengths has the interesting and curious property that *over any interval of time during which there are continuous departures that are naturally ordered by trip length, the departure rate is fully determined by the congestion technology and the distribution of trip lengths.* Let $L(t)$ be the trip length of the commuter who departs at time t, with $t = 0$ denoting the common arrival time. For this commuter

$$L(t) = \int_t^0 v\big(k(u)\big)du, \tag{50}$$

which states the physical identity that a commuter's trip length equals the integral of her velocity over her trip duration.

Also, the traffic density at time u equals the density cars on the road at time u, which equals the density of cars that have departed over that arrival interval according to the natural order. In our model, the natural order of departures over an arrival interval entails commuters with shorter trips departing closer to the peak of the rush hour, when travel is more congested. For arrival interval 1,

$$k(t) = N\Big(G({}^eL^2_{1,2}) - G\big(L(t)\big)\Big). \tag{51}$$

Combining (50) and (51) yields

$$L(t) = \int_t^0 v(N(G({}^eL^2_{1,2}) - G(L(u)))du. \tag{52}$$

The corresponding differential equation is

$$\dot{L}(t) = -v(N(G(L_{1,2}) - G(L(t)))). \tag{53}$$

There is also a boundary condition. If the shortest trip is zero, this condition is $L(0) = 0$.

This condition is curious since it implies that, in general, no congestion pricing scheme that preserves the natural order of departures has any effect on the departure rate over any interval of continuous departures. If therefore the departure interval is connected, no congestion pricing scheme that preserves the natural order of departures has any effect on the evolution of congestion over the rush hour. This result seems to defy the logic of congestion pricing under which the social optimum is decentralized through a congestion pricing scheme that gets commuters to face the full social cost of their travel. One way for a congestion pricing scheme to be effective would be for it to alter the natural departure order, but this defies the logic that the social optimum should entail commuters departing in the natural order. The only way to reconcile the result with the logic of congestion pricing would appear to be that the departure set must not be connected. We have chosen to leave this important issue unresolved. We may do so since, in the model of this subsection, the departure interval is not connected over the rush hour.

If the marginal commuter who is indifferent between traveling in arrival masses 1 and 2 travels in arrival mass 1, she is the first commuter in that arrival mass to depart, and hence experiences increasing congestion over her journey. If she travels in arrival mass 2, she is the last commuter in that arrival mass to depart, and hence travels with all other commuters in that mass throughout her journey. Let ${}^eT_i^2(L_{1,2})$ denote the travel time of the marginal commuter in arrival mass i when there are two arrival masses. ${}^eT_1^2(L_{1,2})$ can be solved from (53). ${}^eT_2^2(L_{1,2})$ can be solved for more easily since the marginal commuter's velocity is constant throughout her trip. The identity of the marginal commuter is then determined by the first condition noted above, that the trip cost for the marginal commuter be the same in the two arrival intervals, ${}^ec_1^2(L_{1,2}) = {}^ec_2^2(L_{1,2})$; more specifically,

$$\alpha^e T_1^2(L_{1,2}) = \alpha^e T_2^2(L_{1,2}) + \beta^e T_1^2(L_{1,2}). \tag{54}$$

The second condition that needs to be satisfied for two departure masses to constitute an equilibrium is that the commuter with the longest trip prefer to travel in arrival mass 2, rather than earlier. If these conditions are satisfied, the trip-timing condition is satisfied for all commuters.

In this model congestion pricing is effective in altering the identity of the marginal commuter and hence the composition of the two arrival masses. We conjecture that, in general, the social optimum can be decentralized through optimal tolling, and that the socially optimal departure set (the set of times when departures occur) is not in general connected.

We now construct an example in which the no-toll equilibrium entails two departure masses. We assume that the congestion technology is described by Greenshields' Relation, that the distribution of trip lengths is uniformly distributed between zero and ten miles so that the average trip length is the same as in the previous examples, and that the other parameters are the same as in the paper's central example.

We first solve (53), which characterizes the time path of congestion for arrival mass 1, as a function of $L_{1,2}$. We then solve the analog to (53) for arrival mass 2, again as a function of $L_{1,2}$. We then use the first condition to identify the marginal commuter, and the second condition to verify that the equilibrium number of arrival masses is indeed two. Fortunately, (53) is a first-order linear ordinary differential equation, which can be solved straightforwardly, yielding the following function $L(t)$ for arrival mass 1:

$$L(t) = (\frac{1}{B} - L_{1,2})(e^{-Bv_f t} - 1),$$

where $B \equiv \dfrac{\hat{N}}{\Omega\bar{L}}$. Defining $t^1_{1,2}$ to be the departure time of the marginal commuter when traveling in arrival mass 1 yields $L_{1,2} = (\dfrac{1}{B} - L_{1,2})(e^{-Bv_f t^1_{12}} - 1)$, the solution of which is

$$t^1_{1,2} = \frac{\ln(1 - L_{1,2}B)}{Bv_f}.$$

Since commuters in arrival mass 1 arrive on time, $-t^1_{1,2}$ is the travel time of the marginal commuter when traveling in mass 1. Thus,

$${}^e c^2_1(L_{1,2}) = \alpha {}^e T^2_1(L_{1,2}) = -\alpha t^1_{1,2} = -\frac{\alpha}{Bv_f}\ln(1 - L_{1,2}B),$$

If instead the marginal commuter is the last person to depart in arrival mass 2, her trip cost is

$${}^e c^2_2(L_{1,2}) = \alpha {}^e T^2_2(L_{1,2}) + \beta {}^e T^2_1(L_{1,2}) = \alpha {}^e T^2_2(L_{1,2}) - \frac{\beta}{Bv_f}\ln(1 - L_{1,2}B).$$

The travel time for the marginal commuter in the second departure mass is the travel distance of the marginal commuter, $L_{1,2}$, divided her velocity. Since all commuters in the second departure mass are on the road when she travels, her velocity is $v_f\big(1 - B(\bar{L} - L_{1,2})\big)$, so that

$${}^e T^2_{1,2} = \frac{L_{1,2}}{v_f\Big(1 - B\big(\bar{L} - L_{1,2}\big)\Big)}.$$

Thus, the equilibrium condition for the marginal commuter is

$$\frac{-(\alpha - \beta)}{Bv_f}\ln(1 - L_{1,2}B) = \frac{\alpha L_{1,2}}{v_f\Big(1 - B\big(\bar{L} - L_{1,2}\big)\Big)}. \tag{55}$$

We take as our parameter values $\alpha = 20$, $\beta = 10$, $v_f = 15$, $\Omega = 1$, and $\bar{L} = 10$. In the basic example, the parameter values were the same, except that trip distance was 5 mls for everyone rather than being uniformly distributed between 0 and 10. Recall that in the basic example ${}^e N_{1,2}$ was equal to θ, which with these parameter values equals 0.5. We assume this value for N. Thus, if there is a second arrival mass in equilibrium, it is due to heterogeneity in departure length. We find that all commuters choose the first arrival mass. Raising N to 0.90, we find ${}^e L_{1,2} = 9.248$. There is so little congestion in the second arrival mass that the commuter with the longest trip distance chooses to travel in that mass rather than to deviate, satisfying the second condition for equilibrium to entail two arrival masses. We note as well that hypercongestion occurs at the peak of the rush hour since the density of cars on the road exceeds one-half jam density.

We conjecture, but have not proved, that the equilibrium we have constructed is unique.

4.6 Directions for Future Research

In this section, we have illustrated extensions by treating only the cases of one or two masses cursorily. All of the extensions we have investigated merit a thorough treatment with an arbitrary number of masses. We were able to derive neat expressions for the no-toll equilibrium values of variables in the case of identical individuals by exploiting a simple recursive relationship between successive departure masses. Whether or not this can be done for the various extensions treated in this section remains to be seen. We have tried but so far without success to obtain a recursive relationship between successive departure masses for the social optimum with identical individuals. There remains one obvious basic extension of the basic model that we have not investigated – heterogeneity in desired arrival time. The hope is that this extension is at the same time analytically tractable and leads to a solution that smoothes out the masses.

The model presented in this paper can also be extended in the many ways that the bottleneck model has been extended, but which of these extensions can be undertaken without undermining analytical tractability remains to be seen.

5 Conclusions

The bottleneck model has been the workhorse for the economic analysis of rush-hour traffic dynamics for a quarter century. It has served our community well, having proved amenable to a rich set of extensions and having provided a bounty of insights. However, like every model, it has deficiencies. Perhaps its biggest weakness is that it assumes away hypercongestion – situations of heavy traffic in which traffic flow falls as traffic density increases, which may be thought intuitively as traffic jams. This is a serious deficiency because when and where downtown traffic congestion is really bad, it is due to hypercongestion.

Urban transportation economists have long recognized this deficiency and understood how the bottleneck model can be extended to treat hypercongestion. The problem has been that all "proper" extensions – those that respect individual choice and the physics of traffic congestion – have proven to be analytically intractable. All result in a delay differential equation with an endogenous delay, whose study is at the research frontier in applied mathematics. Intuitively, the mathematics of dynamical systems in which decisions are based on the current state of the system, which captures the system's past, is well developed. But a commuter makes her trip decisions based not only on the current state of the system but how it will evolve as she travels from her origin to her destination.

The literature has dealt with this problem by attempting to circumvent it through the construction of improper models, interpreted as approximations to the proper model. None of these models has gained general currency because their approximations have been challenged on *a priori grounds*. It is strange that no one has developed computational solutions to the proper models against which the accuracy of the various approximations can be gauged.

This paper has taken a different approach. It has explored the solution properties of a proper model of rush-hour congestion dynamics in an isotropic downtown area for the familiar α-β-γ case, which has been widely employed in the context of the bottleneck model. In this special case, commuters make their trip-timing decisions so as to minimize a trip cost function that is linear in travel time and schedule delay. For this special case, it turns out that closed-form solutions exist not only for the equilibrium with identical individuals, but also for the limited extensions considered in the paper.

It is premature however to make grand claims for the model. Even though proper, the model has three serious weaknesses. The first is that the solutions given in the paper entail masses. When commuters have the same trip length, this entails a subset of the population departing, traveling, and arriving together, followed by successive subsets of population also departing, traveling, and arriving together, with the number of masses being determined endogenously. When commuters do not have the same trip length, the solution entails each subset of the population arriving together but departing at different times. The difficulty, of course, is that this is not what we observe. How consequential this difficulty is remains to be seen. It may be that we have identified only one of several, or even a continuum, of equilibria and local optima. It may be that, even though the model's solutions have this unrealistic property, its aggregate properties conform well to observation. While the model is consistent with the laws of physics, some have challenged the realism of the physics, in particular its assumption that masses of cars either enter or exit the street system without creating any turbulence, which requires infinite acceleration and deceleration over infinitesimal periods of time.

The second weakness is not a weakness of the model per se but instead reflects the very preliminary state of the model's development. We do not know the existence and uniqueness properties of the model. We conjecture that equilibria and local optima always exist simply because, for all the cases we examined we were able to construct them. We conjecture as well that the equilibrium or optimum identified for each of the model variants considered is unique, but we do so with less confidence since we base the conjecture only on having been unable to intuit other equilibria and optima.

The third weakness is analytical tractability. We were able to get closed-form solutions for the no-toll equilibrium with identical individuals and for the limited extensions that we considered. It remains to be seen what extensions and generalizations permit closed-form solutions. Even in those cases where closed-form solutions exist, it was necessary to derive them conditional on the number of masses, and then to determine the equilibrium or optimum number of departure masses, which is cumbersome. If any of the approximating models generates solutions that have the same qualitative properties as a proper model and produce results that are sufficiently accurate quantitatively, it may become acceptable to work with the approximating model.

These weaknesses notwithstanding, this paper's model has many virtues. First, it is a proper model; the user does not have to worry about the importance of any source of approximation error. Second, the physics of the model is simple and easy to understand – traffic velocity is decreasing in traffic density. Third, the economics of the model is simple and easy to understand. In many models of traffic dynamics, including the corridor model of Arnott and DePalma (2011), tracing the external effects of a car requires a complicated perturbation analysis. In contrast, in the paper's model a car generates two readily identifiable and quantifiable external effects – it slows down all cars that are traveling in the same mass but in no others, and it generates schedule delay for all commuters departing in earlier masses – which permits straightforward calculation of optimal toll functions. Fourth, the model accommodates hypercongestion in a way that is simple and easy to understand.

References

Agnew, C. E. (1976). Dynamic modeling and control of congestion-prone systems. *Operations Research*, 24(3):400–419.

Arnott, R. (2013). A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, 76:110 – 121.

Arnott, R., de Palma, A., and Lindsey, R. (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 83(1):pp. 161–179.

Arnott, R. and DePalma, E. (2011). The corridor problem: Preliminary results on the no-toll equilibrium. *Transportation Research Part B: Methodological*, 45(5):743 – 768.

Arnott, R. and Inci, E. (2010). The stability of downtown parking and traffic congestion. *Journal of Urban Economics*, 68(3):260 – 276.

Ben-Akiva, M., Cyna, M., and de Palma, A. (1984). Dynamic model of peak period congestion. *Transportation Research Part B: Methodological*, 18(45):339 – 355.

Chung, K., Rudjanakanoknad, J., and Cassidy, M. J. (2007). Relation between traffic density and capacity drop at three freeway bottlenecks. *Transportation Research Part B: Methodological*, 41(1):82 – 95.

Daganzo, C. F. and Lehe, L. J. (2015). Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B: Methodological*, 75:89 – 99.

Fosgerau, M. (2015). Congestion in the bathtub. MPRA Paper 63029, University Library of Munich, Germany.

Fosgerau, M. and Small, K. A. (2013). Hypercongestion in downtown metropolis. *Journal of Urban Economics*, 76:122 – 134.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759 – 770.

Geroliminis, N. and Levinson, D. (2009). Cordon pricing consistent with the physics of overcrowding. In Lam, W. H. K., Wong, S. C., and Lo, H. K., editors, *Transportation and Traffic Theory 2009: Golden Jubilee*, pages 219–240. Springer US.

Greenshields, B. D. (1935). A study of traffic capacity. *Highway Research Board Proceedings*, 14:448–477.

Hardy, G., Littlewood, J., and Polya, G. (1952). *Inequalities*. Cambridge University Press, Cambridge, UK.

Small, K. A. and Chu, X. (2003). Hypercongestion. *Journal of Transport Economics and Policy*, 37(3):319–352.

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

Vickrey, W. S. (1991). Congestion in Manhattan in relation to marginal cost pricing. Memo, Columbia University.

Notational Glossary

c, $\hat{c}$	normalized, unnormalized trip cost
$c(t)$	trip cost as a functional of departure time
$\underline{c}$	equilibrium trip cost
c_i^m, $\hat{c}_i^m$	normalized, unnormalized trip in mass i conditional on m masses
$c_{(m)}$	common unit cost with m masses
e	equilibrium
i	index of departure or arrival mass
k, $\hat{k}$	normalized, unnormalized density (per unit area)
m	number of masses
n_i^m, $\hat{n}_i^m$	normalized, unnormalized population density in mass i conditional on m masses
s	bottleneck capacity
s^e	severity of congestion ($\equiv MCE^e/c^e$)
t, $\hat{t}$	normalized, unnormalized time
t^*	desired arrival time (set to zero in much of the paper)
v, $\hat{v}$	normalized and unnormalized velocity
v_f	unnormalized free-flow velocty (normalized free-flow velocity equals 1)
A	intermediate variable ($\equiv (1-\theta(1-N))^{1/2}$)
$A(\theta)$	constant term ($\equiv \frac{1-\theta}{\theta}\left[\frac{1-(1-\theta)^m}{(1-\theta)^m}\right]$)
B	intermediate variable ($\equiv \frac{\hat{N}}{\Omega L}$)
$D_{(m)}$, $\hat{D}_{(m)}$	normalized, unnormalized duration of rush hour with m masses
E	elasticity
$F(\theta)$	cumulative distribution function of θ
$G(L)$	cumulative distribution function of $\bar{L}$
L	unnormalized trip distance (normalized trip distance equals 1)
$L_{m,m+1}$	trip length of marginal commuter, who is indifferent between traveling in masses m and $m+1$
$\bar{L}$	maximum trip length
MCE, $M\hat{C}E$	normalized, unnormalized marginal congestion externality cost
MSC, $M\hat{S}C$	normalized, unnormalized marginal social cost
N, $\hat{N}$	normalized and unnormalized population of commuters per unit area
$N_{m,m+1}$, $\hat{N}_{m,m+1}$	normalized, unnormalized population density at which switch occurs from m to $m+1$ masses
$Q(t)$	queue length at time t behind the bottleneck
SDC, $S\hat{D}C$	normalized, unnormalized total schedule delay cost
$T(t)$	normalized travel time as a function of normalized departure time
$\hat{T}(\hat{t})$	unnormalized travel time as a function of departure time
TC, $T\hat{C}$	normalized, unnormalized total trip cost
TTC, $T\hat{T}C$	normalized, unnormalized total travel time cost

α	unnormalized unit value of travel time (normalized unit value of travel time is 1)
β	unnormalized unit value of time early (normalized unit value of time early is θ)
γ	unnormalized unit value of time late (normalized unit value of time late is ρ)
θ	$\equiv \frac{\beta}{\alpha}$
ρ	$\equiv \frac{\gamma}{\alpha}$
τ_1^m	congestion toll applied to each commuter in mass i, conditional on there being m departure masses
Δ	finite increment
Ω	unnormalized jam density (normalized jam density is 1)
*	social optimum